

Table des matières

Préface	7
Natalie Kübler	
Introduction à la place des corpus et des outils numériques pour l'étude des langues de spécialité	11
Cécile Frérot – Mojca Pecman	
<i>Partie 1 – Constitution, annotation et exploitation de corpus en vue de la création de ressources exploitables</i>	35
Chapitre 1 – Méthodologie pour la préparation d'une campagne d'annotation manuelle d'expressions référentielles	37
Frédéric Landragin	
1 Introduction	37
2 La référence et les expressions référentielles	40
2.1 Les phénomènes linguistiques	40
2.2 Leur annotation	40
2.3 L'outil d'annotation et le choix du schéma d'annotation	41
2.4 L'accord interannotateurs	42
3 Identifier des stratégies d'annotation	45
3.1 Annotation systématique des expressions référentielles	45
3.2 Annotation de certaines expressions référentielles	46
3.3 Annotations réalisées	47
4 Pré-annotation automatique : aide ou gêne ?	48
4.1 Quelles pré-annotations ?	48
4.2 Repérages automatiques de haut niveau	49
4.3 Repérage automatique des groupes nominaux	50
4.4 Expérimentations chronométrées	52
5 Exploitation des annotations	54
6 Conclusion et perspectives	55
Remerciements	57
Références bibliographiques	57

Chapitre 2 – Agrégation automatisée de corpus de français parlé	61
Christophe Parisse – Christophe Benzitoun – Carole Étienne – Loïc Liégeois	
1 Introduction	62
2 État des lieux	63
2.1 Revue des principales ressources disponibles	63
2.2 Les agrégations de corpus déjà réalisées	65
2.3 Agrégation de sources existantes : difficultés méthodologiques et enjeux pour l'étude de la langue de spécialité	68
3 TEI-CORPO : un format pivot et une bibliothèque d'outils	70
3.1 La TEI comme format pivot	70
3.2 TEI-ISO pour l'oral	71
3.3 TEI-CORPO : format et outil	71
3.4 Implémentation	73
4 Construction du corpus et utilisation de TEI-CORPO	74
4.1 Les corpus	74
4.2 Traitements avec TEI-CORPO	76
4.2.1 <i>Conversion du format original en TEI</i>	76
4.2.2 <i>Application de l'analyse syntaxique en parties du discours</i>	76
4.2.3 <i>Export vers un format d'import de TXM</i>	77
5 Deux exemples d'analyse	78
5.1 Futur flexionnel et futur périphrastique	78
5.1.1 <i>Distribution FS/FP et acquisition du langage</i>	78
5.1.2 <i>Distribution FS/FP en (micro-)diachronie</i>	80
5.2 Apport de la méthodologie d'agrégation de corpus pour l'étude de la langue de spécialité	81
6 Conclusion	84
Références bibliographiques	84

Chapitre 3 – Étudier les interactions verbales en milieu professionnel : aspects méthodologiques du recueil de données en réunion de travail	87
---	-----------

Anouchka Divoux

1 Introduction	87
2 Principes méthodologiques d'une enquête de terrain en milieu professionnel	89
2.1 Principes éthiques de la recherche	92
2.2 De la familiarisation avec le terrain au recueil de données	94

3 De la démarche d'enquête à l'analyse – de l'analyse à la démarche d'enquête : allers et retours	96
3.1 Technolecte de l'entreprise	96
3.2 Connaissances partagées des locuteurs	98
3.3 Dimensions identitaires des locuteurs	99
3.4 Des données de terrain à l'analyse sociolinguistique des interactions	100
4 Conclusion	103
Références bibliographiques	104
Annexe	106
Conventions de transcription	106
Chapitre 4 – Le français oral quotidien, un objectif spécifique en FLE? Retour sur les défis de la création d'un corpus de français parlé annoté à visée pédagogique	107
Christian Surcouf	
1 Introduction : les enjeux de la compréhension orale	108
2 Une illustration des difficultés de compréhension de l'oral spontané en FLE	110
3 Les défis de la construction d'un corpus de français parlé à visée pédagogique pour le FLE	115
4 Les traits langagiers, leur étiquetage et leur organisation dans l'interface-administrateur	117
4.1 La mise en place de la structure d'accueil des données dans les interfaces administrateur et usager	118
4.2 Le traitement des données audio dans Elan	121
5 Les défis de la construction d'un corpus de français parlé à visée pédagogique pour le Français sur Objectifs Spécifiques (FOS)	124
6 Conclusion	129
Références bibliographiques	130
Partie 2 – Analyses lexicales, lexicologiques et phraséologiques des discours spécialisés	135
Chapitre 5 – La « langue-culture » de la COP21 : autour d'un corpus de spécialité en espagnol	137
Thierry Nallet – Sandrine Rol-Arandjelovic	
1 Introduction	138

2	Présentation du corpus et méthodologie	140
2.1	Limites temporelles et génériques du corpus	141
2.2	Méthode de constitution du corpus	143
3	Analyse du corpus	144
3.1	Termes les plus représentatifs	144
3.2	Variations observées dans les sous-corpus sur l'usage des concepts de <i>justice</i> et d' <i>équité</i>	147
3.3	Variations observées sur l'usage de termes <i>forts</i> tels que <i>sanctions</i> ou <i>responsabilité historique</i>	149
4	Approche qualitative :	
	la variation, la détermination et la recontextualisation	150
4.1	Analyse du concept de <i>résilience</i>	151
4.2	Analyse du concept de <i>matrice énergétique</i>	152
4.3	Analyse du concept de <i>pauvreté</i>	152
4.4	Analyse du concept de <i>vulnérabilité</i>	153
4.5	Analyse des segments répétés tels que <i>changement climatique</i> ou <i>émissions de gaz à effet de serre</i>	153
4.6	Analyse des concepts à forte connotation militante tels que <i>Fondo Verde del Clima (Green Climate Fund)</i> ou <i>empoderamiento (empowerment)</i>	155
4.7	Analyse de concepts nouveaux tels que <i>neutralité</i> ou <i>intersection</i>	156
4.8	De l'analyse lexicale à l'analyse culturelle	157
5	Conclusion et perspectives	158
	Références bibliographiques	159
	Annexe	161
	Annexe 1 : Liste des 23 journaux d'où sont tirés les articles de presse (classés par nombre d'articles)	161
	Annexe 2 : Mots-clés utilisés pour identifier les articles	162

Chapitre 6 – Constitution d'un lexique terminologique trans-biomédical : expérimentations à partir de corpus, listes de vocabulaire et ressources spécialisées 163

Izabella Thomas – Anastasia Galmiche

1	Introduction et contexte	164
2	État de l'art : études sur la constitution de listes de vocabulaire en anglais	166
2.1	Études sur les listes de vocabulaire	166
2.2	Analyse des listes existantes	169
2.3	Problématique de recherche	170
3	Expériences sur les listes	171

3.1	Expérience préliminaire : vérification de la liste de Wang, Liang et Ge (2008)	171
3.2	Ressources : listes et corpus	172
3.2.1	<i>Création des listes à partir d'un dictionnaire</i>	172
3.2.2	<i>Corpus</i>	173
3.3	Projection des listes	174
4	Établissement et validation de la liste finale	178
4.1	Établissement de la liste finale	178
4.2	Vérification du sens spécialisé	178
4.2.1	<i>Validation par des experts</i>	179
4.2.2	<i>Validation par le contexte</i>	179
4.2.3	<i>Validation par une ontologie médicale</i>	180
5	Conclusion	181
	Remerciements	182
	Références bibliographiques	182
Chapitre 7 – Les adjectifs axiologiques dans les guides touristiques : une étude basée sur corpus		185
Jarukan Jitwongnan		
1	Introduction	185
2	Les adjectifs axiologiques dans les guides touristiques	186
2.1	Les guides touristiques : un lieu privilégié pour l'étude des jugements de valeur	186
2.2	Choix du corpus	187
2.3	Problématique	188
3	Cadre théorique	189
4	Méthode de repérage et d'identification des adjectifs axiologiques	190
4.1	Pré-identification	190
4.2	Extraction et identification des adjectifs axiologiques	191
5	Résultats et analyse	193
5.1	Vue globale des adjectifs axiologiques dans le <i>Guide du Routard</i> et le <i>Guide Gallimard Thaïlande</i>	194
5.1.1	<i>Répartition des adjectifs axiologiques</i>	194
5.1.2	<i>Catégories d'adjectifs axiologiques</i>	195
5.1.3	<i>Fonction grammaticale des adjectifs</i>	198
5.2	Les cibles des adjectifs axiologiques	199
6	Discussion	200
7	Conclusion	202
	Remerciements	204
	Références bibliographiques	204

Chapitre 8 – Vers un modèle d'analyse des verbes semi-spécialisés. Le cas de *traiter/treat* dans les domaines médical et environnemental 207

Beatriz Sánchez-Cárdenas – Cécile Frérot

1 Introduction	207
2 Le verbe en terminologie et son fonctionnement linguistique en corpus	210
3 Contexte et objectifs de l'étude	211
4 Cadre théorique et dispositif méthodologique	213
4.1 La terminologie basée sur cadres	213
4.2 Proposition de modélisation des données	214
4.3 Corpus spécialisés et outils d'analyse de corpus	215
5 Informations verbales basées sur corpus	217
5.1 Sélection de verbes semi-spécialisés	217
5.2 Première extraction des données	219
5.3 Extraction avancée des données	220
5.4 Analyse des arguments des verbes <i>treat/traiter</i>	223
6 Conclusions et perspectives	230
Références bibliographiques	232

Partie 3 – Analyses syntaxiques et sémantiques des textes et des discours spécialisés 239

Chapitre 9 – Différenciation des noms d'action dérivés : le facteur de technicité étudié en corpus 241

Marine Wauquier – Cécile Fabre – Nabil Hathout

1 Introduction	241
2 Distinction distributionnelle des noms d'action	243
2.1 Lexeur	243
2.2 Modèle distributionnel Word2Vec	244
2.3 Méthode	245
2.4 Analyse des résultats	246
3 La notion de technicité	247
3.1 Définition générale	247
3.2 Perspective linguistique	248
3.3 Critères d'approximation de la technicité	249
3.3.1 Définition de la technicité des noms d'action	249
3.3.2 Caractéristiques linguistiques de la technicité	250
3.3.3 Ressources	250
3.3.4 Critères	251

4 Modélisation statistique de la technicité	253
4.1 Annotation des traits de technicité	253
4.2 Pouvoir prédictif des traits	256
5 Conclusion	257
Remerciements	258
Références bibliographiques	258
Chapitre 10 – Utilisation d’un corpus spécialisé pour étudier les prépositions dans une perspective didactique	263
Thi Thu Hoai Tran – Rui Yan	
1 Introduction	263
2 Contexte de l’étude	265
2.1 Apprentissage sur corpus	265
2.2 Public d’apprenants ciblé	265
2.3 Aborder les prépositions en cours de FLE	266
3 Méthodologie de travail	267
3.1 Linguistique de corpus	267
3.2 Méthode d’analyse	268
4 Première analyse des résultats	270
5 Expérimentations et réflexions didactiques	272
5.1 Expérimentation avec un public de B1-B2	272
5.2 Expérimentation avec un public de B2-C1	275
6 Conclusion	277
Références bibliographiques	279
Annexe	280
Chapitre 11 – Contextes et connaissances dans les discours de vulgarisation scientifique : dynamiques définitoires et problématiques cognitives	285
Hélène Ledouble	
1 Introduction	286
2 Cadre théorique et méthodologique	287
2.1 Travaux antérieurs et problématiques terminologiques	287
2.2 Corpus et méthodologie	289
2.2.1 <i>Discours de vulgarisation scientifique</i>	289
2.2.2 <i>Méthodologie d’analyse des données textuelles (ADT)</i>	290
3 Définitions et cadrage	291
3.1 Définitions de référence de lutte intégrée	291
3.2 Cadrage définitoire	293
3.2.1 <i>Relation générique</i>	293

3.2.2	<i>Relation partitive</i>	294
3.2.3	<i>Relation de fonction/télique</i>	294
4	Contextes définitoires : relations et connaissances	295
4.1	Relation générique et connaissances associées	295
4.1.1	<i>Revue Phytoma</i>	295
4.1.2	<i>Discours de presse</i>	296
4.2	Relation partitive et connaissances associées	298
4.2.1	<i>Revue Phytoma</i>	298
4.2.2	<i>Discours de presse</i>	299
4.3	Relation télique et connaissances associées	301
4.3.1	<i>Revue Phytoma</i>	301
4.3.2	<i>Discours de presse</i>	301
5	Discussion	302
5.1	Relations et hiérarchie des connaissances	302
5.1.1	<i>Différenciation des niveaux taxonomiques</i>	302
5.1.2	<i>Ellipse d'un niveau taxonomique</i>	303
5.2	Stratégies de vulgarisation scientifique	305
5.2.1	<i>Lexique et connaissances</i>	305
5.2.2	<i>L'approche analogique et le « recadrage »</i>	306
6	Perspectives et conclusions	308
	Références bibliographiques	309

Chapitre 12 – Apport de la traduction dans l'analyse des marqueurs de relations conceptuelles. Une étude en corpus aligné français-italien 313

Anne Condamines – Marie-Pierre Escoubas-Benveniste –
Silvia Federzoni

1	Introduction	314
2	Contexte de l'étude, travaux existants	314
2.1	Travaux en terminologie textuelle	314
2.2	Travaux en traductologie	316
2.2.1	<i>Le texte traduit : un genre communicatif à part entière</i>	316
2.2.2	<i>Corpus, traductologie et analyse contrastive</i>	317
2.2.3	<i>Les dimensions de la variation du texte traduit</i>	319
3	Mise en place de l'étude pour l'hyponymie	321
3.1	Le Corpus	321
3.2	Le choix d'un corpus de presse	323
3.3	Les marqueurs de relation conceptuelle	324
3.4	Mise en œuvre des marqueurs avec TXM	325
4	Étude des résultats obtenus	326
4.1	Ajout d'un hyponyme au moment de la traduction	327

4.2	Suppression du marqueur dans la traduction	328
4.3	Ajout d'un marqueur dans la traduction	329
4.4	Utilisation d'un autre type de marqueur dans la traduction	330
4.5	Métalangage et métadiscours dans la traduction en italien	330
6	Conclusion	332
	Références bibliographiques	333
Chapitre 13	– Rôles et valeurs des emplois en « c'est »	
dans le corpus Philosophèmes	337
Philippe Roiné – Mylène Blasco – Emmanuèle Auriac-Slusarczyk		
1	Introduction	338
2	Le corpus Philosophèmes	339
2.1	Présentation du corpus	339
2.2	Le corpus <i>Philosophèmes</i> : orientation vers une exploration écologique	340
3	Les principes méthodologiques de l'exploitation du corpus	342
3.1	Fouille et mise en grille syntaxique des propos philosophiques ..	342
3.2	Regard pragmatique sur les propos présentés en grille autour des emplois en « c'est »	343
4	Résultats : tendances et tri des emplois en « c'est »	344
4.1	Exploration outillée du corpus <i>Auriac-Slusarczyk 2009-2013</i>	344
4.2	Le regard pragmatique : première considération sur la langue de spécialité	344
4.3	Fouille complémentaire croisée dans le corpus <i>Roiné 2012-2016</i>	345
4.3.1	<i>Cas 1</i> : le <i>N</i> c'est quelque chose qui/que	346
4.3.2	<i>Cas 2</i> : le <i>N</i> c'est quand, le <i>N</i> c'est comme	347
4.3.3	<i>Cas 3</i> : <i>N</i> c'est + <i>construction infinitive</i>	348
4.4	Le regard pragmatique : vers la confirmation d'une langue de spécialité exercée en atelier de philosophie	349
4.4.1	<i>L'inversion de la polarité assertive en philosophie</i>	350
4.4.2	<i>La désignation en philosophie</i>	354
5	Conclusion et discussion	355
	Références bibliographiques	356
Index	361