

UNIVERSIDAD DE SALAMANCA
FACULTAD DE FILOLOGÍA
DEPARTAMENTO DE FILOLOGÍA FRANCESA



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

DOCTORADO EN LENGUAS MODERNAS

TESIS DOCTORAL

**DICTIONNAIRE BILINGUE FRANÇAIS-ESPAGNOL DES NOMS
PRÉDICATIFS DES <AIDES FINANCIÈRES>/<AYUDAS
FINANCIERAS>. ÉTUDE SYNTACTICOSÉMANTIQUE**

Autora: ELENA DIEGO HERNANDEZ

Directora de Tesis: ISABEL UZCANGA VIVAR

Codirectora de tesis: ARACELI GÓMEZ FERNÁNDEZ

Tutora: ELENA LLAMAS POMBO



RÉSUMÉ

Con esta tesis hemos pretendido estudiar las propiedades sintáctico-semánticas de los sustantivos predicativos de las <ayudas económicas>. Nos hemos interesado especialmente en la inserción del sustantivo en la frase, desde la perspectiva del tratamiento automático de la lengua. Para ello, hemos considerado conveniente adoptar la metodología desarrollada en la teoría de las clases de objeto.

La tesis está organizada en dos grandes partes. En la primera, hacemos una incursión teórica a nociones que están en la base de nuestro estudio. Comenzamos profundizando nuestro objeto de estudio: el sustantivo, la predicación nominal, la diferencia entre el discurso de especialidad y el discurso o la lengua general. Seguidamente, hacemos una incursión en la fraseología: su historia, su evolución como disciplina y su extensión como concepto. Una vez presentadas estas nociones básicas, nos adentramos en el modelo de la clase de objetos para comprender esta aproximación a la descripción de la lengua. Finalmente, terminamos esta parte teórica con una presentación de las especificidades del tratamiento automático de la lengua. Las dos grandes dificultades a las que se enfrenta el tratamiento automático de las lenguas naturales es precisamente aquello que las caracteriza: la polisemia y la fijación. Precisamente consideramos que una descripción de las unidades léxicas en empleos, permite solventar estos problemas. De hecho, desde hace años se lleva trabajando desde esta perspectiva y por ello presentamos las herramientas léxicas e informáticas que se han desarrollado desde esta aproximación lingüística.

En una segunda parte, presentamos nuestra metodología de trabajo, que está esencialmente adoptada del marco teórico en el que se inscribe nuestra investigación. Para llevar a cabo nuestro estudio sobre los predicados nominales de <ayudas financieras> hemos optado por utilizar un corpus web. En último lugar, presentamos en análisis en clases sintáctico-semánticas que hemos podido llevar a cabo siguiendo el método de las clases de objetos.

Finalmente, consideramos que hemos conseguido un recurso útil para el tratamiento automático, pero quizá notan exhaustivo como lo hubiéramos deseado.

TABLE DE MATIÈRES

LISTE DE FIGURES ET DES TABLEAUX	11
LISTE DES ABRÉVIATIONS, SYMBOLES ET CONVENTIONS GRAPHIQUES .	14
CHAPITRE I : INTRODUCTION GÉNÉRALE.....	15
1. Objet d'étude	15
1.1. Présentation succincte de la notion de dictionnaire électronique.....	17
2. Objectifs	20
3. Méthodologie.....	20
4. Plan de la thèse	21
CHAPITRE II : LES PRÉDICATS NOMINAUX DES <AIDES FINANCIÈRES>....	24
1. La notion de prédicat nominal.....	24
1.1. Mise en cause de la notion de <i>partie du discours</i>	25
1.2. Le prédicat nominal : différentes positions théoriques	28
1.2.1. Nom prédicatif : un nom morphologiquement dérivé	29
1.2.2. Nom prédicatif : un nom abstrait	31
1.2.3. Nom prédicatif : un nom à structure argumentale	34
1.2.3.1. Le prédicat nominal dans la théorie Sens-Texte.....	35
1.2.3.2. Le prédicat nominal dans le modèle des classes d'objets.....	37
2. Le discours de spécialité.....	43
2.1. Langue de spécialité ou discours de spécialité ?.....	44
2.2. Caractéristiques des discours de spécialité de notre corpus.....	49
2.2.1. Caractéristiques lexicales des discours de spécialité.....	49
2.2.1.1. Présence de noms composés.....	51
2.2.1.2. Présence de néologismes.....	52
2.2.1.3. Présence de sigles	53

2.2.1.4.	Présence de langage figuré	53
2.2.2.	Caractéristiques syntaxiques des discours de spécialité	53
3.	Terminologie et dictionnaires	54
3.1.	La terminologie	55
3.1.1.	L'optique conceptuelle	55
3.1.1.1.	Mise en cause de l'optique conceptuelle	57
3.1.2.	L'optique lexicosémantique.....	59
3.2.	La terminographie	60
3.2.1.	Les dictionnaires spécialisés	64
3.2.1.1.	Dictionnaires monolingues, dictionnaires bilingues et dictionnaires multilingues.....	65
3.2.1.2.	La nomenclature des dictionnaires spécialisés.....	67
3.2.1.3.	La syntaxe dans les dictionnaires spécialisés.....	68
4.	Conclusions	68
CHAPITRE III : LA PHRASÉOLOGIE		70
1.	Bref aperçu historique de la phraséologie	71
1.1.	La phraséologie française.....	90
1.2.	La phraséologie espagnole	92
2.	Le statut de la phraséologie	95
2.1.	Discipline autonome ou indépendante	95
2.2.	Objet d'étude et limites de la phraséologie	99
2.3.	Phénomène marginal ou central de la langue.....	102
3.	Qu'est-ce qu'une unité phraséologique ?	104
3.1.	La terminologie phraséologique.....	104
3.2.	Caractéristiques des unités phraséologiques	106
3.2.1.	La polylexicalité.....	109
3.2.2.	La fréquence	109
3.2.3.	L'institutionnalisation	110

3.2.4.	Le figement	110
3.2.5.	L'idiomaticité ou non-compositionnalité.....	117
3.3.	Classification des unités phraséologiques	121
3.3.1.	Typologie de Gläser	122
3.3.2.	Typologie de Mel'čuk.....	123
3.3.3.	Typologie de G. Gross	125
3.3.5.	Typologie de Corpas Pastor	126
4.	Les collocations	128
4.1.	Qu'est-ce qu'une collocation ?.....	129
4.2.	Propriétés des collocations	133
4.2.1.	Le caractère arbitraire de la collocation.....	134
4.2.2.	Le sens compositionnel ou la transparence de la collocation.....	136
4.2.3.	Le caractère binaire de la collocation.....	139
4.2.4.	La dissymétrie des composants de la collocation.....	140
4.2.5.	La cooccurrence lexicale restreinte de la collocation.....	141
4.3.	Le statut phraséologique des collocations	143
5.	Conclusions	145
CHAPITRE IV : LE MODÈLE DES CLASSES D'OBJETS.....		147
1.	La notion de prédicat dans la théorie des classes d'objets.....	149
1.1.	Les origines harrissiennes de la notion de prédicat.....	149
1.2.	Le prédicat comme relation.....	152
1.3.	Les formes morphologiques du prédicat	153
1.4.	Les prédicats polymorphiques.....	154
1.5.	Les racines prédicatives	155
1.6.	Le schéma d'arguments vs la polysémie	156
2.	La phrase élémentaire	157
2.1.	La notion de phrase élémentaire.....	157
2.2.	L'organisation du lexique en trois fonctions primaires.....	159

2.3.	La phrase élémentaire vs la polysémie	160
2.4.	La notion de contexte	162
2.4.1.	Les contextes aléatoires.....	162
2.4.2.	Les contextes contraints	163
2.4.3.	Les contextes significatifs	163
3.	La description des prédicats : la notion d'emploi.....	164
3.1.	Un emploi est associé à un schéma d'arguments	165
3.2.	Un emploi est associé à un sens	166
3.3.	Un emploi est associé à une forme morphologique	167
3.4.	Un emploi est associé à une actualisation	168
3.5.	Un emploi est associé à un système aspectuel	168
3.6.	Un emploi est associé à des modifications.....	170
3.6.1.	Les transformations	170
3.6.2.	Les mises en évidence.....	171
3.6.3.	Les effacements	171
3.6.4.	Les restructurations	172
3.7.	Un emploi est associé à un domaine spécifique.....	173
3.8.	Un emploi est associé à un registre de langue.....	173
4.	La description des arguments : les classes d'objets.....	174
4.1.	La notion de classe d'objets	174
4.2.	L'élaboration des classes d'arguments : les opérateurs appropriés.....	177
4.3.	La hiérarchisation des classes d'objets.....	179
4.4.	La notion d'héritage	181
5.	La description des actualisateurs	182
5.1.	L'actualisation prédicative vs l'actualisation argumentale	183
5.2.	L'actualisation des prédicats nominaux	184
5.2.1.	L'actualisation moyennant le verbe support.....	184

5.2.1.1.	Propriétés syntaxiques du verbe support.....	187
5.2.1.1.1.	Effacement du verbe support	187
5.2.1.1.2.	Double analyse du complément <i>N Prep N</i>	188
5.2.1.1.3.	Contraintes sur la détermination du nom supporté	189
5.2.1.1.4.	Pas de nominalisation du verbe support	190
5.2.1.2.	Types et fonctionnalités des verbes supports	192
5.2.1.2.1.	Verbes supports généraux	194
5.2.1.2.2.	Verbes supports appropriés	194
5.2.1.2.3.	Verbes supports aspectuels	196
5.2.1.2.4.	Variantes diastratiques	199
5.2.1.2.5.	Verbes supports et métaphore	200
5.2.1.2.6.	Verbes supports passifs.....	200
5.2.2.	L'actualisation moyennant la détermination	204
5.2.2.1.	L'hétérogénéité de la détermination.....	204
5.2.2.2.	La détermination et les classes d'objets.....	205
5.2.2.2.	La détermination des prédicats	207
6.	La phrase complexe	209
6.1.	Les complétives constituent des arguments propositionnels.....	209
6.2.	La subordonnée circonstancielle	211
6.2.1.	La description des connecteurs	213
7.	Conclusions	215
CHAPITRE V : LE TRAITEMENT AUTOMATIQUE DES LANGUES		218
1.	Les difficultés du TAL	220
1.1.	Le traitement de la polysémie	220
1.1.1.	La notion de polysémie en TAL.....	221
1.1.2.	L'analyse automatique de la polysémie	222
1.1.3.	Le traitement lexicographique de la polysémie	223
1.2.	LE TRAITEMENT DU FIGEMENT	228
1.2.1.	La notion de figement dans le modèle des classes d'objets	229

1.2.2.	Reconnaissance et génération automatique des unités polylexicales	232
2.	Méthodes d'analyse automatique : méthode statistique vs méthode mixte	233
2.1.	La méthode statistique	233
2.2.	La méthode mixte	234
3.	Ressources lexicales pour l'étiquetage	236
3.1.	Ressources pour l'étiquetage morphosyntaxique	238
3.1.1.	DicoValence	238
3.1.2.	Lefff (Lexique des Formes Fléchies du Français) – Leffe (Léxico de Formas Flexionadas del Español)	239
3.2.	Ressources pour l'étiquetage sémantique	240
3.2.1.	WordNet	240
3.2.2.	DiCo (Dictionnaire de Combinatoire)	245
3.2.2.1.	DiCouèbe	248
3.2.2.2.	DiCoPop	250
3.2.2.3.	Lexique actif du français (LAF)	252
3.2.3.	Réseau Lexical du Français (RL-fr)	254
3.2.4.	Definiens	258
3.3.	Ressources pour l'étiquetage en rôles sémantiques	260
3.3.1.	FrameNet	260
3.3.2.	VerbNet	265
3.3.3.	PropBank (Propositional Bank)	267
3.4.	Ressources pour l'étiquetage de l'aspect et de la temporalité	270
3.4.1.	TimeBank	270
3.4.2.	Nomage	274
3.5.	Informatisation des dictionnaires papier	278
4.	Le traitement automatique et la méthode des classes d'objets	280
4.1.	Ressources linguistiques pour le système de TAL du modèle des classes d'objets	282
4.1.1.	Dictionnaires électroniques	282

4.1.1.1.	Dictionnaires morphosyntaxiques	283
4.1.1.2.	Dictionnaires syntacticosémantiques	286
4.1.2.	Grammaires locales	289
4.1.3.	Corpus	290
4.2.	Étapes tu traitement automatique	291
4.2.1.	Le prétraitement : normalisation des documents.....	292
4.2.2.	Segmentation en phrases et en mots.....	292
4.2.3.	Annotation morphosyntaxique.....	293
4.2.4.	Annotation syntacticosémantique.....	294
5.	Conclusions	296
CHAPITRE VI : MÉTHODOLOGIE		298
1.	Principes d'analyse	298
1.1.	La langue comme métalangue.....	299
1.2.	La phrase élémentaire comme cadre d'analyse.....	299
1.3.	L'approche syntacticosémantique du lexique	299
1.4.	Le souci d'exhaustivité et de systémativité.....	300
1.5.	La nécessité d'homogénéité	300
1.6.	L'analyse ascendante.....	300
2.	Étapes de travail et outils d'analyse	301
2.1.	Établissement de la nomenclature	301
2.2.	Constitution du corpus	303
2.3.	Élaboration d'une grille d'analyse	309
3.	Conclusions	311
CHAPITRE VII : DESCRIPTION DES PRÉDICATS NOMINAUX D'<AIDE FINANCIÈRE>		313
1.	<aide financière : allocation>	315
2.	<aide financière : donation>.....	340

3. <aide financière : dégrèvement>	350
CHAPITRE VIII : CONCLUSIONS GÉNÉRALES	370
BIBLIOGRAPHIE	373
OUTILS LEXICOGRAPHIQUES CONSULTÉS.....	450
ANNEXE I	451

LISTE DE FIGURES ET DES TABLEAUX

Figure 1 : Relation (type i) entre la langue générale (LG) et la langue de spécialité (LE)	46
Figure 2 : Relation (type ii) entre la langue générale (LG) et la langue de spécialité (LE).....	46
Figure 3 : Relation (type iii) entre la langue générale (LG) et la langue de spécialité (LE).....	47
Figure 4 : Polysémie et synonymie en terminologie	58
Figure 5 : Structure du système phraséologique selon Gläser	122
Figure 6 : Classification des phrasèmes.....	125
Figure 7 : Trois entrées du verbe apprendre dans le Lefff	240
Figure 8 : Entrées du nom help dans WordNet	242
Figure 9 : Hyponymes de la première entrée du nom help dans WordNet	243
Figure 10 : Formulaire de requête du DiCouèbe	249
Figure 11 : Exemple de navigation hiérarchique dans le DiCoPop.....	251
Figure 12 : Entrée meurtre du DiCPop	252
Figure 13 : Entrée meurtre du LAF	254
Figure 14 : Graphe de l'unité lexicale hold-up I du RL-fr	256
Figure 15 : Échantillon de l'article lexicographique de l'unité lexicale hold-up I du RL-fr	257
Figure 16 : Étiquetage sémantique de brouette dans Definiens.....	259
Figure 17 : Définition et phrases du frame assistance	262
Figure 18 : Échantillon des frame elements du frame assistance.....	262
Figure 19 : Échantillon de la description de l'unité lexicale aid	263
Figure 20 : Exemple de notation d'une phrase dans PropBank.....	269
Figure 21 : Exemple d'annotation ISO-TimeML	272

Figure 22 : Exemple n'annotation ISO-TimeML de French TimeBank	274
Figure 23 : Hiérarchie des classes aspectuelles simples des noms.....	277
Figure 24 : Architecture du système TextBox.....	281
Figure 25 : Architecture du système Morfetik.....	284
Figure 26 : Graphe de groupe nominal	289
Figure 27 : Graphe de phrases avec le prédicat de < sentiment> joie	290
Figure 28 : Annotation morphosyntaxique de porte	294
Figure 29 : Représentation XML de l'analyse prédicative.....	296
Figure 30 : Définition d'allègement	371
Tableau 1 : La terminographie selon l'optique conceptuelle ou l'optique lexicosémantique	63
Tableau 2 : Trois classes majeures de phrasèmes	124
Tableau 3 : Classification des unités phraséologiques de Corpas Pastor	128
Tableau 4 : Entrée supprimer du DicoValence.....	238
Tableau 5 : Description partielle de l'unité lexicale meurtre dans le DiCo	247
Tableau 6 : Entrée meurtre du DiCouèbe	250
Tableau 7 : Entrée de la classe hit dans VerbNet	266
Tableau 8 : Description du nom aménagement#1 dans Nomage	275
Tableau 9 : Description du verbe aménager#1 dans Nomage	276
Tableau 10 : Une occurrence de aménagement#1 dans Nomage.....	276
Tableau 11 : Structure de la table de flexion du verbe acquérir.....	285
Tableau 12 : Descripteurs de définition de dévorer.....	288
Tableau 13 : Descripteurs de définition de dévorer.....	288
Tableau 14 : Évolution du nombre de sous-classes d'<aides financières>	314
Tableau 15 : Schéma d'arguments d'<aide financière : allocation> en français	315

Tableau 16 : Schéma d'arguments d'<aide financière : allocation> en espagnol	316
Tableau 17 : Schéma d'arguments d'<aide financière : allocation> avec 5 arguments en français.....	316
Tableau 18 : Schéma d'arguments d'<aide financière : allocation> avec 5 arguments en espagnol	316
Tableau 19 : Fiche descriptive de la classe <aide financière : allocation>	340
Tableau 20 : Schéma d'arguments d'<aide financière : donation> en français	340
Tableau 21 : Schéma d'arguments d'<aide financière : donation> en espagnol.....	340
Tableau 22 : Fiche descriptive de la classe <aide financière : donation>	350
Tableau 23 : Schéma d'arguments d'<aide financière : allègement> en français.....	351
Tableau 24 : Schéma d'arguments d'<aide financière : allègement> en français.....	351
Tableau 25 : Fiche descriptive de la classe <aide financière : dégrèvement>	369

LISTE DES ABRÉVIATIONS, SYMBOLES ET CONVENTIONS GRAPHIQUES

< >	classe synacticosémantique
Adj.	adjectif
Adv.	Adverbe
Ncoll	nom collectif
Dét.	déterminant
E	vide
Fam.	familier
f	féminin
Fig.	figuré
inf.	infinitif
N1N0	construction converse
Nhum	Nom humain
Prép.	préposition
TAL	Traitement automatique des langues naturelles
tps	temps
V	Verbe

CHAPITRE I : INTRODUCTION GÉNÉRALE

À l'aube du traitement automatique des langues, l'optimisme des spécialistes en linguistique computationnelle les mena à penser que l'analyse linguistique pouvait se faire à partir de l'extraction de règles générales de grammaire et son application sur de simples mots. Pourtant, le discours ne se construit pas seulement à partir de l'application d'un certain nombre de règles de grammaire. Ces analyses se heurtèrent rapidement contre ce qui constitue l'essence même des langues naturelles : la polysémie et le figement.

En effet, la présumée liberté combinatoire, c'est-à-dire la capacité du locuteur de construire un discours à partir de la libre sélection des unités lexicales et de l'application des règles de grammaire, n'est qu'une utopie linguistique. La linguistique computationnelle et la linguistique de corpus ont rapidement démontré que notre discours est majoritairement contraint. Il est soumis à de nombreuses restrictions de sélection.

D'autre part, la polysémie constitue l'une des caractéristiques fondamentales des langues naturelles. L'être humain est capable de résoudre les problèmes d'interprétation liés à la polysémie en s'appuyant sur le contexte linguistique dans lequel un mot donné apparaît. Autrement dit, une unité lexicale acquiert un sens donné dans un contexte linguistique donné. De même, l'outil informatique sera capable de faire une bonne interprétation du discours s'il peut se servir de ce contexte linguistique.

1. Objet d'étude

L'objet de notre étude est le nom prédicatif désignant des aides financières. Ces noms appartiennent tantôt à la langue générale (*pension, réduction, don, soutien*), tantôt au discours spécialisé (*allocation, aide à la recherche d'emploi, abattement, prestation d'accueil du jeune enfant*). La classe sémantique des <aides financières> étant d'une grande ampleur, la nomenclature de notre dictionnaire comprend des termes de différents domaines (le droit, les finances, l'enseignement, la protection sociale, etc.). Ainsi, l'originalité de cette étude est axée sur deux versants. D'une part, nous réunissons dans un même travail des noms de la langue générale et des noms de ce qu'on appelle des *langues de spécialité*, et nous les décrivons au moyen des mêmes

outils linguistiques. En effet, nous considérons que la séparation entre langue de spécialité et langue générale n'est pas fondée linguistiquement. D'autre part, en ce qui concerne les termes, nous ne limitons pas notre recherche à un seul domaine de spécialité, comme c'est la pratique courante dans les travaux en terminologie ou en langues de spécialité.

Nous portons notre intérêt sur la prédication, car elle constitue l'élément central de l'analyse syntacticosémantique. La description de la prédication nous permet de décrire le contenu propositionnel. Dans notre cas de figure, la description de la prédication nominale nous permet de décrire la phrase dans laquelle ces noms s'inscrivent et prennent leur sens.

La grammaire traditionnelle, fortement influencée par la conception binaire de la phrase issue de la philosophie antique (la phrase est découpée en *thème* et *rhème*) et reprise par la grammaire générative (la phrase est découpée en *groupe nominal* et *groupe verbal*), a toujours porté son attention sur le verbe. Pourtant, la prédication n'est pas une fonction exclusive du verbe. D'autres parties du discours telles que le nom, l'adjectif, voire la préposition, peuvent être prédicatives, mais elles ont souvent été délaissées dans l'analyse syntacticosémantique de la langue.

Par ailleurs, comme l'affirme Béjoint (2007), « les lexicographes ont toujours eu pour objet de description l'élément lexical isolé. Dans les dictionnaires, cet élément est débarrassé de son contexte pour que les informations données à son sujet puissent mieux s'adapter à tous les contextes possibles » (p. 17). Les seules descriptions de l'environnement syntagmatique du mot-vedette contenues dans les dictionnaires sont beaucoup plus nombreuses pour les verbes (et pour certains adjectifs) que pour les noms. Tel est clairement le cas des descriptions des langues de spécialité, habituellement limitées à la terminologie. L'étude du nom a souvent été reléguée à de simples listes contenant des définitions ou des traductions en d'autres langues. En effet, selon Català (2003) :

Le terminologue-terminographe s'attache principalement à délimiter et à définir des concepts représentés par des termes. Ses fiches terminologiques donnent priorité aux données de type encyclopédique comme le domaine et la définition, de type signalétique (les sources documentaires) et de type traductologique (équivalents en une ou plusieurs langues cibles). Mais le terme, pour sa part, demeure une réalité linguistique dont le fonctionnement syntacticosémantique est rarement abordé ou même simplement ignoré. (p. 324)

Quelques études ont été faites sur la classe des <aides> (cf. G. Gross, 2012, 2017 ; Massoussi et Sfar, 2009 ; Mel'čuk et al., 1984). Massoussi et Sfar (2009) et G. Gross (2017) ont mené des études préliminaires sur cette classe sémantique dans le cadre du modèle des classes d'objets. Ces études ont permis de faire une première classification des différents types de verbes supports sélectionnés par ces prédicats nominaux. Pourtant, il ne s'agit que des études à l'état embryonnaire, car elles ne visent ni l'exhaustivité ni la systémativité. En effet, certains verbes supports repérés ne peuvent pas se combiner avec tous les noms signalés (*verser une subvention* ; **verser un concours*) (cf. G. Gross, 2017). D'autre part, l'analyse exhaustive de la combinatoire verbale des noms prédictifs d'<aide financière> a révélé qu'elle est bien plus riche que ce que ces études montrent.

Nous avons puisé dans le lexique du français et de l'espagnol pour constituer la nomenclature de notre dictionnaire électronique. Il est capital de remarquer que le résultat n'est pas un dictionnaire bilingue, mais deux dictionnaires monolingues coordonnés.

1.1. Présentation succincte de la notion de dictionnaire électronique

Le terme *dictionnaire électronique* fait allusion à deux objets différents : des dictionnaires en support électronique destinés à l'usage humain, et des dictionnaires conçus pour être utilisés par des logiciels de traitement automatique. Parmi les dictionnaires électroniques à usage humain, on peut distinguer ceux qui sont une version électronique d'un dictionnaire papier (souvent appelés *dictionnaires informatisés*), et ceux qui sont élaborés d'emblée sur un support électronique.

Veronis (2004) signale que, traditionnellement, les dictionnaires à usage humain ont essayé de décrire le sens des mots au lieu de leurs emplois (à l'exception de quelques indications sur le domaine ou le registre). Vers la moitié du XX^e siècle, à la suite des travaux lexicographiques de Hornby (1942, 1954), certains dictionnaires ont commencé à incorporer de manière systématique des informations syntaxiques, distributionnelles ou paradigmatiques. Le *Dictionnaire du français contemporain* (1966) de Dubois intègre les schémas de valence (nombre et nature des compléments) dans le dégroupement des entrées. Pourtant, ce dictionnaire manque de systémativité. Ultérieurement, le *Trésor de la Langue Française* (1971-1994), fondé sur l'étude d'un

grand corpus, contient aussi quelques indications sur la combinatoire, mais elles sont loin d'un relevé systématique. En espagnol, le *Diccionario de uso del español* (1966-1967) de María Moliner¹ est très intéressant du point de vue de la combinatoire et laisse une forte empreinte dans la lexicologie et la lexicographie espagnoles, mais il manque encore de systématisme.

Béjoint (2007) soutient que « l'utilisation de l'informatique est l'évènement le plus important de l'histoire de la lexicographie des cinquante dernières années » (p. 7). En effet, l'avènement de l'informatique a modifié la représentation des dictionnaires et le travail même du lexicographe. Ainsi, le développement de grands corpus et des outils informatiques permettant leur exploitation a bouleversé le travail lexicographique. D'une part, la lexicographie s'est tournée vers la description des emplois des mots, c'est-à-dire vers la combinatoire. D'autre part, les lexicographes ont compris la nécessité de faire des descriptions exhaustives et systématiques. Depuis les années 1980, la combinatoire des mots retient l'attention des lexicographes² (cf. Verlinde et al., 2005). C'est dans ce contexte que le *Dictionnaire explicatif et combinatoire du français contemporain* de Mel'čuk et al. (1984, 1988, 1992, 1999) a vu le jour³. En espagnol, le plus important dictionnaire de combinaisons n'a été publié qu'en 2004, sous la direction de Bosque : *REDES, Diccionario combinatorio del español contemporáneo*⁴.

Ces ouvrages lexicographiques plus récents sont très intéressants du point de vue lexicologique et lexicographique, mais ils sont toujours insuffisants pour le traitement automatique. Le dictionnaire électronique destiné aux logiciels de traitement automatique n'est pas un dictionnaire de consultation. Il s'agit plutôt d'un outil

¹ Quand María Moliner feuillette le *Learner's Dictionary of Current English* (1948) de Hornby, elle sème le germe de son *Diccionario de uso del español*.

² Des outils statistiques ont souvent été utilisés pour relever les cooccurrences pertinentes : l'*information mutuelle* (*mutual information*), le *Z score* ou le *T score*.

³ On notera que dans ce dictionnaire les définitions lexicographiques partent de l'introspection, le corpus ne servant qu'à posteriori comme vérification. Il s'agit d'un dictionnaire papier, mais des interfaces informatiques contenant une bonne partie de ce dictionnaire ont été développées.

⁴ Il s'agit d'un dictionnaire papier. Postérieurement, toujours sous la direction de Bosque, le *Diccionario combinatorio práctico del español contemporáneo* (2006) a été publié (encore en papier).

contenant des informations linguistiques formalisées, et qui sera utilisé par des systèmes informatiques dans le but de reconnaître ou de générer des textes. Ce type de dictionnaires est donc au cœur des systèmes de traitement automatique des langues naturelles. Comme l'automate n'a pas de connaissances préalables de la langue ni du monde, pour que le dictionnaire électronique soit performant, il doit contenir des informations bien plus explicites et précises que celles présentées dans un dictionnaire informatisé ou dans un dictionnaire papier (cf. Pruvost, 2000).

Pour résumer, Courtois et Silberztein (1990) passent en revue les différences les plus importantes entre les dictionnaires électroniques destinés aux automates et ceux qu'ils appellent les « dictionnaires du commerce », c'est-à-dire des dictionnaires à usage humain :

- a) les dictionnaires électroniques doivent être aussi complets que possible ; par contre, le dictionnaire du commerce n'est jamais complet ;
- b) les informations du dictionnaire électronique doivent être explicites, tandis que les dictionnaires du commerce laissent implicites des informations supposées connues ; ainsi, souvent ceux derniers font appel à des connaissances préalables du monde ;
- c) les dictionnaires électroniques sont intrinsèquement liés à des logiciels de traitement automatique. Toutes les informations de ces dictionnaires sont précises et cohérentes parce qu'elles doivent être formalisées pour être utilisables par des programmes informatiques.

Néanmoins, l'opposition entre les dictionnaires papier et les dictionnaires électroniques n'est pas justifiée, car ceux derniers se nourrissent des données des premiers. En effet, comme le soutient Cartier (2009) : « les dictionnaires papier sont sans aucun doute les ressources qui contiennent le plus d'informations sémantiques exploitables car ils sont le fruit de longues années de travail » (p. 19). Ainsi, les dictionnaires papier contiennent des informations sémantiques exploitables pour le développement des dictionnaires électroniques telles que les définitions, les marques liées au domaine technique d'utilisation d'un sens donné, certaines constructions ou restrictions de combinatoire, des liens sémantiques avec d'autres unités lexicales (des synonymes, des antonymes, des hyperonymes ou des hyponymes) et même des listes de mots composés construits à partir du mot vedette.

2. Objectifs

La présente étude vise la description et la classification syntacticosémantique des noms prédicatifs d'<aide financière> en français et en espagnol dans le but d'élaborer un module d'un dictionnaire électronique. Ce module ne prend pas la forme d'un dictionnaire habituel, mais il est construit sous la forme d'une base de données. Nous nous intéressons tout particulièrement à l'intégration de ces noms prédicatifs dans le cadre de la phrase. Nous nous proposons donc de décrire la structure argumentale de ces noms prédicatifs, ainsi que d'analyser leur combinatoire.

Notre hypothèse de travail est que la description exhaustive et systématique du fonctionnement syntacticosémantique des noms prédicatifs étudiés devrait nous permettre de les regrouper en classes sémantiques ayant le même comportement dans la phrase. De cette manière, l'outil de traitement automatique qui utiliserait cette ressource pourrait prédire de manière automatique la combinatoire verbale de ces noms.

Nous tenons à signaler que notre approche est strictement linguistique. Cette étude vise à étudier le fonctionnement dans la phrase d'un lexique spécifique et non à faire un recensement exhaustif des différents types d'aides financières disponibles en France et en Espagne ni à les classer en fonction du type d'aide ou du bénéficiaire ou d'autres sortes de critères extralinguistiques.

3. Méthodologie

Le présent travail s'inscrit dans le cadre d'une théorie lexicaliste, la théorie des classes d'objets. Cette théorie continue la lignée de la grammaire transformationnelle de Harris (1976) et des travaux sur le lexique-grammaire (M. Gross, 1975 ; Leclère, 1998). Elle part d'une classification sémantique, complétée par des règles syntaxiques.

Les unités lexicales sont conçues comme des parties primordiales de la langue. Cette méthodologie permet une description multilatérale du lexique, c'est-à-dire les unités lexicales décrites sont considérées simultanément sous leur facette sémantique (l'établissement de classes sémantiques), sous leur facette syntaxique (description de leur schéma d'arguments) et sous leur facette lexicocombinatoire (les collocations).

Selon G. Gross (2012),

Un des objectifs que nous fixons à la recherche linguistique [...] est de permettre une analyse automatique de textes. Cela implique un degré de couverture des faits linguistiques suffisant, que seul peut apporter un dictionnaire électronique construit à cet effet. Tous les mots du texte doivent bien entendu être reconnus morphologiquement, mais aussi décrits quant à leurs propriétés combinatoires, c'est-à-dire syntaxiques, et leur signification en contexte. (p. 11)

Ainsi, dans notre étude nous proposons une description des noms prédicatifs d'<aide financière> qui recouvre plusieurs niveaux d'analyse. Du point de vue sémantique, les noms prédicatifs sont regroupés en classes sémantiques. En effet, l'appartenance à une classe est fondée d'abord sur des critères de synonymie ou quasi-synonymie. Du point de vue morphologique, dans notre dictionnaire électronique, les unités lexicales étudiées sont caractérisées en genre et en nombre. De plus, nous présentons toutes les formes qu'un nom prédicatif donné peut avoir ; par exemple, les lexies *alloc* ou *allocs* sont souvent utilisées dans un registre de langue familier (comme nous avons pu l'attester dans la consultation de forums d'internet) au lieu des termes *allocation* ou *allocations* ; nous avons également recensé les sigles qui correspondent à certains noms d'aides financières, car elles sont souvent utilisés dans le discours (p. ex., *APL*). Enfin, du point de vue syntaxique, toutes les unités lexicales étudiées sont définies par leur schéma d'arguments et par leur combinatoire verbale.

4. Plan de la thèse

Dans le premier chapitre, nous abordons des notions fondamentales pour cerner l'objet de notre étude : le nom prédicatif d'<aide financière>. Ainsi, nous survolons différentes conceptions de la notion de nom prédicatif et nous expliquons comment il est conçu et décrit dans le cadre de la théorie des classes d'objets. Dans ce chapitre qui nous permet de cerner et de mieux comprendre l'objet de notre recherche, nous abordons également les caractéristiques des discours de spécialité pour observer en quoi ceux-ci diffèrent du discours général et nous faisons une brève approximation aux études en terminologie et en terminographie.

Le deuxième chapitre est consacré à la phraséologie. Nous partons d'un bref aperçu historique pour suivre le développement de cette discipline, ainsi que l'extension de la notion d'unité phraséologique. Les limites de la notion d'unité phraséologique restent floues, en fonction de l'approche linguistique adoptée. Il s'avère donc nécessaire de définir ce que nous comprenons par unité phraséologique. Nous nous attardons sur la

notion de collocation, puisqu'elle constitue l'un des éléments capitaux de notre recherche.

Dans le chapitre trois, nous présentons de manière détaillée la théorie des classes d'objets. Nous expliquons ses postulats fondamentaux et la manière de procéder à la description de la langue dans ce cadre théorique. Cette présentation nous permet de comprendre pourquoi cette approche théorique satisfait les besoins du traitement automatique.

Dans le chapitre quatre, nous faisons une brève incursion au traitement automatique, puisqu'il détermine le contenu et la forme de notre recherche. D'abord, nous analysons les deux grandes difficultés auxquelles se heurte le traitement automatique des langues naturelles : la polysémie et le figement. Ensuite, nous présentons les deux grandes méthodes d'analyse linguistique qui existent actuellement et nous précisons celle que nous adoptons, tout en justifiant notre choix. Pour avoir un bref aperçu des travaux en linguistique informatique en relation avec la lexicologie et la lexicographie, nous présentons quelques ressources lexicales pour l'étiquetage morphosyntaxique, sémantique, l'étiquetage de l'aspect et de la temporalité et l'informatisation des dictionnaires papier. En dernier, nous focalisons sur le traitement automatique dans le cadre théorique des classes d'objets. Ainsi, nous présentons certaines ressources linguistiques et les étapes du système de traitement automatique propre à cette théorie.

Dans le chapitre cinq, nous expliquons la méthodologie que nous avons suivie pour l'élaboration de notre dictionnaire électronique. Nous présentons d'abord les principes d'analyses qui ont régi nos choix méthodologiques. Ensuite, nous détaillons les étapes de travail que nous avons suivies et les outils d'analyse que nous avons développés pour mener notre recherche.

Enfin, dans le chapitre six, nous présentons le dictionnaire électronique que nous avons élaboré. Nous décrivons les propriétés syntacticosémantiques des prédicats nominaux d'<aide financière> et nous comparons les résultats obtenus dans les deux langues.

Finalement, dans la conclusion nous tirons les conclusions du résultat de notre recherche ainsi que nous faisons le point sur l'efficacité de la méthodologie de travail et le cadre théorique que nous avons adoptés.

CHAPITRE II : LES PRÉDICATS NOMINAUX DES <AIDES FINANCIÈRES>

Dans ce chapitre nous ferons une première approximation à l'objet de notre étude pour appréhender ce que nous entendons par prédicat nominal et ce que nous entendons aussi par nom d'<aide financière>. Ces noms appartiennent autant à la langue générale qu'au discours de spécialité. Ainsi, nous avons considéré nécessaire de mener une réflexion autour de cette distinction. Nous nous attarderons sur les caractéristiques lexicales et syntaxiques des discours de spécialité afin de pouvoir démontrer que les unités lexicales appartenant au discours de spécialité peuvent (voire doivent) être décrites avec les mêmes outils d'analyse que les unités lexicales de la langue générale. Enfin, nous ferons une incursion dans la terminologie, branche de la linguistique qui étudie les termes, et nous présenterons les bases théoriques sur lesquelles les spécialistes élaborent différents produits terminographiques.

1. La notion de prédicat nominal

Comme de nombreux termes du métalangage linguistique, le terme *prédicat* désigne différentes entités linguistiques. À première vue, ce terme a deux grandes significations. D'une part, dans la tradition logique classique, le *prédicat* désigne ce qui complète le sujet pour faire une proposition. D'autre part, la linguistique moderne, héritière de la logique rénovée de Frege (1867/1979), entend par *prédicat* le centre fonctionnel (ou organisateur) de la phrase, régissant des variables qui sont appelées des *arguments* (autant le sujet que les compléments sont des arguments)⁵. En d'autres termes, « le prédicat est une catégorie logicosémantique de nature relationnelle, une sorte de fonction⁶ sur la base de laquelle se calcule le sens de tout énoncé » (Mejri,

⁵ La vieille dichotomie aristotélicienne *sujet/prédicat* ou *thème/rhème* est remplacée par *prédicat/argument* (ou *fonction/argument*, d'après la terminologie utilisée par Frege). Pour une synthèse des emplois du terme *prédicat*, voir Merle (2009, p. 6-8) ; Muller (2013, p. 51-53).

⁶ F(x), où F est la fonction et x, la variable.

2019, p. 112). Dorénavant, c'est seulement cette deuxième signification que nous retiendrons dans notre travail.

Par ailleurs, le substantif⁷ occupe une place importante dans les travaux de linguistique, mais sa définition et classification font aussi l'objet d'interprétations divergentes.

Dans cette section, nous survolerons les problèmes définitionnels des notions *partie du discours* et *substantif*. Ensuite, nous présenterons les principales interprétations du terme *nom prédicatif* : pour certains, le nom prédicatif est un nom déverbal ; d'autres l'entendent comme un nom abstrait ; et d'autres encore considèrent que le nom prédicatif est un nom avec une structure argumentale. Enfin, nous définirons brièvement⁸ ce que l'on entend par prédicat nominal et nous présenterons les caractéristiques des prédicats des <aides financières>.

1.1. Mise en cause de la notion de *partie du discours*

Les parties du discours représentent des ensembles à fonctions hétérogènes. En ce qui concerne le substantif, cette catégorie recouvre des réalités très diverses : des arguments élémentaires (les substantifs concrets), des prédicats (*aide*), des déterminants (*un tas de, une foule de*), des parties constituantes d'adverbiaux (*de plain-pied*), des indicateurs de genre (*une souris mâle*), des valeurs adjectivales (*une braderie monstre*), etc. (G. Gross, 2004a, 2012).

Ainsi, la notion de partie du discours ou catégorie grammaticale a été mise en cause par de nombreux linguistes. Rappelons, que Tesnière (1959) a dénoncé le caractère infondé de la classification traditionnelle en parties du discours, du fait de l'imprécision des critères de définition de celles-ci. En effet, parmi les grammairiens modernes, il est communément admis que les diverses classifications en parties du discours sont basées sur un amalgame de critères définitionnels très varié : on utilise généralement des critères sémantiques pour définir les noms et les verbes, des critères

⁷ Nous utilisons de manière indistincte les termes *nom* et *substantif*.

⁸ Nous entrerons plus en détail sur la notion de prédicat dans le modèle des classes d'objets dans le Chapitre II.

positionnels pour les adjectifs et les prépositions, et enfin, un mélange de critères vagues pour les adverbes (Bosque, 1991). Par ailleurs, la grande dispute des spécialistes a toujours été de déterminer le nombre de parties du discours sous lesquelles regrouper le lexique d'une langue.

Béguelin (2000) signale que les grammaires traditionnelles ont souvent eu recours à des critères notionnels pour définir les différentes catégories grammaticales. Cette prépondérance de la sémantique est l'héritage de la philosophie aristotélique et de la scholastique. Leur notion des catégories grammaticales pourrait être résumée comme suit :

si el mundo físico consta de objetos (sustancias) que poseen propiedades (accidentes), mantienen relaciones, realizan acciones y experimentan procesos, la gramática debería reflejar esta misma concepción. Tendríamos sustantivos para designar los objetos, adjetivos para las cualidades, verbos para las acciones y los procesos, y partículas para las relaciones (Bosque, 1990, p. 35).

Ainsi, dans la *Grammaire de l'Académie française* (1932) on prétend que le nom « désigne les êtres vivants, les choses, les qualités [...] » (p. 67). En ce qui concerne la tradition grammaticale espagnole, Bello (1847/2004) soutient que le *nombre común* o *apelativo* est celui qui « conviene a todos los individuos de una clase, especie o familia, significando su naturaleza o las cualidades de que gozan » (p. 41). Plus tard, dans la *Gramática de la lengua española* de Emilio Alarcos Llorach (1994), on définit le substantif par des critères syntaxiques (a), morphologiques (b) et sémantiques (c) :

- a. El *sustantivo* es toda palabra capaz de cumplir en los enunciados llamados oraciones [...] la función de sujeto explícito (por ejemplo, *Sale humo*, *Pasan coches*) o la de objeto directo (por ejemplo, *Hay humo*, *Había coches* [...]) sin necesidad de ningún otro elemento. (p. 72)
- b. La estructura interna del sustantivo consiste, en general, en la combinación de un signo léxico, expresado por la raíz, y unos signos morfológicos, *accidentes* o *morfemas*, que suelen ser expresados en la desinencia. (p. 72)

- c. Designan objetos únicos [...] [...] los sustantivos comunes o apelativos, que clasifican los objetos de la realidad física o mental como pertenecientes a una determinada clase. (p. 83)⁹

Dans ces exemples, nous observons donc que, du point de vue sémantique, le nom est réduit à la désignation des êtres humains, des choses ou objets et des qualités¹⁰. Pourtant, le nom peut désigner aussi des procès (*vieillessement*), des actions (*destruction*), des résultats (*construction*), des états (*fatigue, innocence*), des portions de l'espace (*panorama*) et de temps (*matinée*), des quantités (*gramme*), etc. (Riegel et al., 1994). D'ailleurs, Flaux et Van de Velde (2000)¹¹ affirment même que, grâce à la nominalisation, qui permet de convertir en nom des adjectifs et des adverbes, les noms peuvent exprimer n'importe quel type de sens.

Toutefois, dans le cadre de notre approche théorique, la limite entre les parties du discours perd de l'intérêt. La distinction vraiment intéressante s'opère entre les trois fonctions primaires : prédicats, arguments et actualisateurs. Observons les deux phrases suivantes :

- (1) a. *L'équipage de l'Aquarius a aidé des centaines de migrants naufragés*¹².

⁹ Marengo (2011) souligne que ces critères sémantiques, syntaxiques et morphologiques, souvent utilisés dans les grammaires pour définir les parties du discours, sont rarement respectés par l'ensemble des unités linguistiques appartenant à une partie du discours donnée.

¹⁰ Pour approfondir sur le traitement de la notion de substantif dans la tradition grammaticale espagnole, voir Gómez Asencio (1981); Calero Vaquera (1986) et Ramajo Caño (1987). Du point de vue synchronique, pour une analyse comparée des catégories grammaticales en espagnol, voir Bosque (1990).

¹¹ Dans cet ouvrage, Flaux et Van de Velde proposent une classification sémantique des noms, sans observer la combinatoire de ceux-ci. Elles considèrent que le fait qu'un nom prédicatif donné sélectionne un verbe support ou un autre n'est pas pertinent du point de vue linguistique car ces verbes n'apporteraient aucune différence dans la nature des noms. Leur classement est basé sur des oppositions du type concret/abstrait, extensif/intensif, dénombrable/indénombrable, animé/inanimé, humain/non humain et naturel/fabriqué. Cette approche est donc, en partie, à l'opposé de la manière de travailler dans la méthode des classes d'objets, dont le postulat est que la phrase est l'unité minimale d'analyse (parce que c'est dans le cadre de la phrase qu'une unité linguistique donnée prend son sens).

¹² Les exemples ne contenant pas de référence bibliographique ont été construits par l'auteurice de cette thèse.

b. *L'équipage de l'Aquarius a apporté son aide à des centaines de migrants naufragés.*

Le substantif *aide* joue le même rôle que le verbe *aider* : il s'agit de deux prédicats qui relient le substantif collectif humain *équipage* au substantif humain *migrants*. Ils partagent donc le même schéma d'arguments¹³ et ils véhiculent le même sens.

Dans notre approche théorique, la notion syntaxique de prédicat est plus importante que la forme morphologique que celui-ci peut adopter. En effet, G. Gross et Vivès (1986) affirment : « Seuls sont pertinents le nombre et l'agencement des arguments » (p. 26). Pourtant, la notion de partie du discours est nécessaire dans notre description de la langue : nous nous en servons pour cerner l'objet de notre recherche et pour décrire le schéma d'arguments des noms prédicatifs qui composent notre dictionnaire.

1.2. Le prédicat nominal : différentes positions théoriques

Nous avons pu observer dans l'exemple (1b) précédent qu'il existe des substantifs qui ont un sujet et des compléments, c'est-à-dire ils possèdent une structure argumentale, comme les verbes. De même, le verbe *accorder* n'a pas de valeur prédicative. Il accomplit une fonction d'actualisation du prédicat nominal, puisque ce dernier ne peut pas porter les marques de temps.

La reconnaissance des prédicats nominaux a été tardive. Dans la tradition grammaticale européenne, on a toujours considéré que c'est le verbe qui porte la prédication dans le cadre de la phrase, ou bien l'adjectif dans le cas des phrases copulatives¹⁴ (Herrero Ingelmo, 2004). Actuellement, l'existence de prédicats nominaux

¹³ Force est de constater que les verbes prédicatifs peuvent sélectionner un complément *NI* avec ou sans préposition, alors que le *NI* des noms prédicatifs est toujours précédé par une préposition, c'est-à-dire ils n'ont pas de complément direct (Herrero Ingelmo, 2004 ; G. Gross, 2012).

¹⁴ Subirats (2001) a fait une révision du traitement de la notion de prédicat dans la tradition grammaticale et lexicographique espagnole depuis la *Gramática castellana* de la Real Academia Española, (1796). Il a pu observer certaines réflexions ou allusions à une possible fonction prédicative du substantif.

est communément admise. Pourtant, de même qu'il n'existe pas un consensus sur le nombre des catégories grammaticales ni leurs caractéristiques définitives, la notion de *nom prédicatif* fait l'objet aussi d'une grande controverse.

1.2.1. Nom prédicatif : un nom morphologiquement dérivé

Le terme *nom prédicatif* est souvent compris comme un nom morphologiquement dérivé d'un verbe. D'ailleurs, une étude approfondie du catalan menée par Badia (1994) a montré que la plupart des noms prédicatifs sont morphologiquement dérivés de verbes ou d'adjectifs. Cet auteur soutient que pour les noms qui ne seraient pas dérivés, il serait facile de créer un verbe ou un adjectif *ad hoc* dont ils dériveraient.

Par contre, les recherches en lexique-grammaire ont permis de montrer que les noms prédicatifs non dérivés représentent dans le lexique du français à peu près le double des noms déverbaux (G. Gross, 1996b). La dérivation verbale ne serait donc pas un critère de prédictivité. Comme G. Gross (1989) l'affirme, « un prédicat nominal est défini par le fait qu'il a des arguments et par la nature de ces arguments, et non par un lien morphologique avec un prédicat verbal » (p. 8). Dans la nomenclature de notre dictionnaire, nous trouvons, par exemple, les noms prédicatifs *bourse* (N0:Ncoll / N1:à Nhum / N2:de <montant> / N3:pour, sur<période_tps> / N4:pourVW)¹⁵ ou *allocation* (N0:Ncoll / N1:à Nhum / N2:de <montant> / N3:pour Vinf), qui n'ont pas de verbe morphologiquement associé¹⁶.

De surcroît, le fait que le nom prédicatif soit morphologiquement associé à un verbe ne garantit pas qu'il présente les mêmes emplois que le verbe. Giry-Schneider (1987b) signale l'existence de noms prédicatifs sans relation syntaxique avec la construction verbale ; par exemple, le substantif *escalade* et le verbe *escalader* (p. 66) :

¹⁵ Dans un souci de clarté, nous avons tenu à ajouter des espaces ne respectant l'encodage habituellement utilisé dans la méthode des classes d'objets.

¹⁶ C'est à dessein que nous n'employons pas ici le terme « dérivé » car parfois c'est le verbe qui dérive du substantif. Il n'est pas toujours facile de déterminer la séquence d'apparition de ces deux unités linguistiques. Herrero Ingelmo (2004) donne l'exemple des noms espagnols *viaje* et *consejo*, qui sont antérieurs aux verbes *viajar* et *aconsejar*. Nous préférons donc le terme « associé » au lieu de « dérivé ».

(2) a. *Les prix font une escalade vertigineuse.*

b. **Les prix escaladent (E + quelque chose).*

L'argument *prix* étant le sujet, il accepte la construction avec le prédicat nominal (2a), mais pas avec le prédicat verbal (2b).

Autrement dit, l'existence d'un rapport morphologique entre un substantif et un verbe, n'implique pas de rapport syntaxique et sémantique entre la construction verbale et la construction nominale. Un autre exemple illustratif est la paire *conférence/conférer* (Vivès, 1993, p. 10) :

(3) a. *Luc confère avec Max de l'amélioration possible de la situation.*

b. *Luc donne une conférence aux étudiants sur l'amélioration possible de la situation.*

Le verbe *conférer* et le substantif *conférence* n'ont ni le même sens ni les mêmes constructions prépositionnelles. Parmi les noms prédicatifs des <aides financières>, la paire *revenu/revenir* répond à cette même remarque :

(4) a. *L'État accorde un revenu à tous les chômeurs de longue durée.*

b. **L'État revient à tous les chômeurs de longue durée.*

Dans d'autres cas, le prédicat verbal et le prédicat nominal peuvent être reliés du point de vue sémantique et morphologique, mais les constructions syntaxiques dans lesquelles ils s'inscrivent sont différentes. Dans les exemples suivants on observera un changement de préposition (*à/pour*) :

(5) a. *Marie s'intéresse à la physique.*

b. *Marie a de l'intérêt pour la physique.*

Ainsi, Vivès (1993) conclut que :

la nominalisation n'est pas une transformation syntaxique générale, comme par exemple la passivation, mais un phénomène local qui met en jeu le lexique (la nominalisation est particulière pour chaque couple Verbe-Nom, Adjectif-Nom) et qu'elle ne peut être étudiée que dans le cadre des phrases simples à verbe support (p. 10).

En effet, la description des schémas d'arguments doit se faire sur la base des noms prédicatifs, et non sur celle des verbes auxquels ils sont associés.

Enfin, il existe des noms prédicatifs qui sont morphologiquement associés à des adjectifs, par exemple, *sérieux* et *optimiste*, qui partagent exactement la même forme morphologique que l'adjectif. Pourtant, nous n'en avons pas trouvé d'exemples parmi la nomenclature de notre dictionnaire.

En somme, les noms prédicatifs peuvent être autonomes (sans relation morphologique avec d'autres parties du discours), associés à un verbe ou associés à un adjectif.

1.2.2. Nom prédicatif : un nom abstrait

Dans la tradition grammaticale on a tendance à assimiler les noms prédicatifs à des noms abstraits. Nous trouvons deux problèmes majeurs à cette idée : il n'existe pas un accord sur la définition de la notion d'abstrait ; et il est prouvé que des noms concrets peuvent être prédicatifs et que des noms abstraits peuvent ne pas être prédicatifs.

D'une part, la dichotomie *concret/abstrait*, fondamentale dans la catégorisation des noms, semble être insaisissable. En effet, les définitions de ces deux notions dans la littérature sont loin de faire l'unanimité¹⁷. Ainsi, dans la *Grammaire méthodique du français* (Riegel et al., 1994) on peut lire : « La distinction entre noms concrets et noms abstraits est la plus problématique. Sémantiquement il est difficile de fonder cette dichotomie en termes suffisamment univoques » (p. 326). Kleiber et Vuillaume (2011) résument cet éventail d'interprétations en trois grands types d'opposition :

- (i) opposition entre référents matériels et immatériels ;
- (ii) opposition entre référents accessibles et inaccessibles aux sens¹⁸ ;
- (iii) opposition entre des référents autonomes et dépendants d'un point de vue ontologique, c'est-à-dire une entité dont l'existence implique l'existence d'une autre entité¹⁹.

¹⁷ Voir Bosque, (1983) ; Flaux et al. (1996) ; Flaux et Van de Velde (2000).

¹⁸ C'est, par exemple, le parti pris dans la *Grammaire méthodique du français* (Riegel et al., 1994).

Comme Huyghe (2015) le signale, ces trois définitions ne se recouvrent pas. Par exemple, le substantif *odeur*, référant à quelque chose de sensible, il est concret dans le sens (ii), mais pas dans les sens (i) et (iii) parce qu'il ne réfère pas à un objet matériel et qu'il ne s'agit pas d'une entité autonome (« l'odeur de quelque chose » ; ce substantif implique une entité qui exhale de l'odeur). Dans le cas de notre étude, tous les noms des aides financières seraient concrets dans les sens (i) et (ii), mais abstraits dans le sens (iii), étant donné que l'aide implique au moins l'existence de deux entités (celui qui apporte de l'aide, celui qui la reçoit).

Par ailleurs, Bosque (1990) met en question la pertinence de cette distinction :

Los gramáticos tradicionales se refieren una y otra vez a la dificultad de caracterizar con precisión la clase de los nombres abstractos, pero no parecen plantearse la pregunta que resulta natural en términos actuales: el problema no es exactamente qué definición debemos dar del concepto 'sustantivo abstracto'²⁰, sino más bien determinar si los nombres abstractos tienen entidad gramatical delimitada, esto es, si se trata o no de una clase asimilable a otras clases de sustantivos o a otras interpretaciones semánticas que no constituyen necesariamente clases léxicas (p. 47)²¹.

Bosque considère que l'opposition concret/abstrait n'est plus fondamentale pour la classification des noms, mais qu'elle est dérivée de facteurs lexicaux, aspectuels et phrastiques.

D'autre part, nous considérons que des noms concrets comme *aumône* ou *pourboire* sont prédicatifs car ils ont un schéma d'arguments (N0:Nhum / N1:à Nhum / N2:deN<montant>)²². En effet, ces noms entrent dans des constructions à verbe support. D'après Alonso Ramos (1998) quand on *donne une aumône* ou qu'on *laisse un*

¹⁹ C'est, par exemple, la position d'Alonso et Henríquez Ureña (1938) et de Flaux et Van de Velde (2000).

²⁰ Nous avons respecté les guillemets simples du texte original.

²¹ Rappelons que dans la classification que cet auteur proposait quelques années auparavant (Bosque, 1983), il considérait pertinentes seulement les oppositions continu/discontinu, concret/abstrait, singulier/collectif.

²² Pour plus d'informations à propos de la notion de nom prédicatif comme un nom ayant une structure argumentale, voir la section suivante : 1.2.3. *Nom prédicatif équivalent à nom avec structure d'arguments*.

pourboire, on est effectivement en train de donner quelque chose. Cependant, cette linguiste considère que ces verbes sont supports car ils n'apportent pas vraiment de sens ou qu'ils n'apportent qu'une sens très vague, et que leur fonction principale est d'actualiser le nom qu'ils accompagnent²³. Ces noms prédicatifs portent dans leur sens le signifié 'donner' :

même si l'on peut employer des prédicats qui soulignent leur caractère de chose, comme dans [...] *María garde les aumônes <pourboires>*²⁴ *dans une boîte*, il s'agit d'aumônes ou de pourboires déjà donnés, et donc une aumône ou un pourboire n'existe pas sans que quelqu'un les donne à quelqu'un (Alonso Ramos, 1998, p. 218).²⁵

Ainsi, le caractère concret d'un nom, ne l'empêche pas d'être un prédicat. De même, tous les noms abstraits ne sont pas nécessairement des prédicats, car l'autonomie syntactico-sémantique des noms n'implique pas l'autonomie ontologique de leurs référents (Huyghe, 2014).

Dans la perspective de la méthode des classes d'objets, les classifications ontologiques ont peu d'intérêt et n'ont pas de valeur linguistique. Comme nous le

²³ Le critère de l'absence de sens n'est pas accepté par toute la communauté linguistique. Par exemple, De Miguel (2008) soutient que le verbe qui entre dans une construction à verbe support « no se ha vaciado de significado sino que se ha rellenado del significado nominal » (de Miguel, 2008, p. 570). Elle considère que la fonction du verbe support n'est pas seulement d'actualiser le nom prédicatif, mais qu'il apporte du sens et contribue à la sélection sémantique des arguments. Nous considérons que ces apports sémantiques du verbe support restent très vagues et sont plutôt à caractère aspectuel, comme nous le démontrons dans la description des noms prédicatifs des <aides> de notre dictionnaire électronique. Pour une caractérisation plus détaillée sur la notion de verbe support, voir Chap. IV, section 5.2.1. *Les verbes supports*.

²⁴ Nous avons respecté les conventions d'écriture du texte original. Alonso Ramos utilise les chevrons pour indiquer que la construction avec le nom *pourboire* est possible aussi, c'est-à-dire ils indiquent une alternance. Dans notre travail, nous avons utilisé les chevrons pour indiquer qu'il s'agit d'une classe d'objets.

²⁵ Rappelons qu'Alonso Ramos développe sa recherche dans le cadre de la théorie Sens-Texte. Dans ce cadre théorique les noms *aumône* et *pourboire* sont considérés des *quasi-prédicats sémantiques* : des unités lexicales dont le sens « dénote une entité (comme un nom sémantique) et qui est liant (c'est-à-dire a des positions argumentielles – comme un prédicat) » (Iordanskaja et Mel'čuk, 2017, p. 399). Les quasi-prédicats sont donc des sens intermédiaires, qui tiennent en même temps des prédicats et des non-prédicats (Mel'čuk et Polguère, 2008).

montrons, autant les notions des trois fonctions primaires (prédicats, arguments et actualisateurs) que le classement en classes d'objets sont rigoureusement linguistiques.

1.2.3. Nom prédicatif : un nom à structure argumentale

Muller (2013) affirme que « l'une des utilisations majeures de la notion de prédicat est celle d'une métacatégorie universelle qui décrit les choix morphologiques du locuteur entrant dans des relations de dépendance en tant que recteur d'arguments » (p. 58). Ainsi, la caractéristique définitoire des noms prédicatifs serait qu'ils ont une structure argumentale²⁶. Il s'avère donc nécessaire d'établir des critères qui permettent de définir ce que l'on entend par *argument*, et qui permettent aussi de déterminer si un complément donné d'un nom est son argument ou non. Nous allons présenter de façon très succincte deux grandes approches de la linguistique actuelle qui nous intéressent spécialement : la théorie Sens-Texte et la méthode des classes d'objets²⁷.

Escandell (1995), pour sa part, considère que la structure argumentale d'un prédicat dépend du signifié lexical de chaque mot. Elle emploie les exemples d'*entrega* et *manzana* pour illustrer la différence entre un nom dont la définition n'inclut pas de référence à un participant (*manzana*) et un nom qui a une structure argumentale (*entrega*). Pourtant, Alonso Ramos (1998) met en évidence le mélange de critères sémantiques et syntaxiques : « Escandell (1995, p. 19 et 32) semble prendre la notion d'argument comme une notion sémantique. [...] Mais quand Escandell (1995, p. 18) présente la définition d'*argument*, elle commence par mentionner le terme *constituant*, c'est-à-dire une notion syntaxique » (p. 30-31).

Grimshaw (1990) établit une différence claire entre le *participant sémantique* d'un prédicat nominal et l'*argument syntaxique* d'un nom. Ainsi, les participants

²⁶ De nombreux travaux sont inscrits dans le cadre générativiste : Rappaport (1983) ; E. Williams (1981, 1987) ; Grimshaw et E. Williams (1993) ; Ingria et Leland (1993). Bierwisch (1990), dans son approche particulière, a fait aussi un travail remarquable. Dans le modèle de grammaires HPSG (*Head-Driven Phrase Structure Grammar*, parfois traduit « Grammaire syntagmatique guidée par les têtes »), nous soulignons la recherche de Badia et Colominas (1997).

²⁷ Pour une analyse critique des différentes interprétations du prédicat, voir les numéros 31 et 32 de la revue *Faits de langue* (2009) ; Muller (2013).

interviennent au niveau de la représentation lexicoconceptuelle (ou lexicosémantique), tandis que les arguments sont présents dans le niveau lexicosyntaxique. La notion de structure d'arguments renvoie donc à la représentation lexicosyntaxique des participants d'un prédicat donné. D'après Grimshaw (1990), tous les noms ont une structure lexicoconceptuelle, c'est-à-dire tous les noms ont des participants, mais seulement certains d'entre eux ont des arguments syntaxiques. Ces deux notions correspondent à ce que l'on appelle des *noms d'évènement* (les noms ayant une structure d'arguments) et des *noms de résultat* (les noms n'ayant pas de structure d'arguments)²⁸.

1.2.3.1. Le prédicat nominal dans la théorie Sens-Texte

Dans une perspective presque exclusivement sémantique, dans le cadre de la théorie Sens-Texte (cf. Mel'čuk et Polguère, 1987 ; Mel'čuk, 1997 ; Polguère, 1998a ; Wanner, 1996) on considère que les notions de prédicat²⁹ et d'actant sont avant tout des notions relevant de la sémantique. Mel'čuk et Polguère (2008) les définissent de la manière suivante :

- *sens prédictif* ou *prédicat*, désignera un sémantème³⁰ liant qui dénote un fait³¹ ('élire', 'élection', 'éligible', 'pendant [les élections]'...);
- *actant sémantique* d'un prédicat dans une phrase donnée désignera un sens qui correspond à un participant du fait dénoté par le prédicat en question et qui est exprimable dans la phrase auprès de ce prédicat. (p. 100)

²⁸ En fait, peu de linguistes sont satisfaits de ces termes. On trouvera alors dans la littérature une terminologie très variée à ce sujet.

²⁹ Dans le cadre de la théorie Sens-Texte, la notion de prédicat est empruntée de la logique mathématique (notion de prédicat logique). Mel'čuk et Polguère (2008) soulignent que cet emprunt n'est que métaphorique car le prédicat sémantique n'est pas une notion logique. Ainsi, les lois de la logique ne s'appliqueraient que de manière « métaphorique » sur la description de la sémantique prédictive.

³⁰ Dans la terminologie de la théorie Sens-Texte un sémantème est le signifié d'une lexie.

³¹ Selon Polguère (2008), « un fait est « quelque chose qui a lieu » ; par exemple : une action (manger, marcher, interroger...) ; un état (aimer, souffrir, être malade, se souvenir...) ; un évènement (exploser, tomber, naître...) ; etc. » (p. 132).

Ainsi, les prédicats, en tant que sens liants qui dénotent un fait, impliquent au moins un participant (actant)³².

En ce qui concerne le prédicat, nous avons expliqué ci-dessus qu'on s'intéresse à son expression dans la phrase et à l'expression de ses actants. Quant à ces derniers, ils remplissent, dans une structure sémantique donnée, une position actancielle sémantique associée au prédicat qui les contrôle³³. La notion de position actancielle est une variable sémantique qui correspond à un participant de la situation dénotée par le prédicat (point de vue sémantique), et qui est exprimé dans la phrase par un actant sémantique. Dans le cas des prédicats nominaux, l'actant peut apparaître soit comme dépendant syntaxique d'un collocatif verbal du prédicat (un verbe support ou un verbe de réalisation) dans des constructions à verbe support (6a), soit comme dépendant syntaxique directement du nom prédicatif (6b). Par exemple, observons la forme propositionnelle d'*ultimatum*³⁴ (Mel'čuk et Polguère, 2008) : 'ultimatum communiqué par la personne X à la personne Y pour demander l'accomplissement de Z avant le moment T'. Ainsi, dans les phrases :

(6) a. *L'UE a lancé un ultimatum aux politiciens espagnols pour qu'ils forment un gouvernement dans les plus brefs délais.*

b. *L'ultimatum de l'UE aux politiciens espagnols pour qu'ils forment un gouvernement dans les plus brefs délais me semble inacceptable.*

le système de dépendances sémantiques du prédicat *ultimatum* est le même dans les deux phrases, mais les dépendances syntaxiques sont différentes. Dans (6a) le verbe support *lancer* sert à inscrire dans la phrase l'actant X du prédicat nominal. Cet actant X

³² On notera qu'il existe des prédicats dénotant des faits, mais qui ne sont pas des sens liants. Par exemple, les verbes de phénomènes atmosphériques comme *tonner* ou *pleuvoir* ne contrôlent pas de structure actancielle.

³³ Voir Mel'čuk (2004a, 2004b) pour une théorie complète des actants.

³⁴ Dans la théorie Sens-Texte, les lexies sont encodées en petites capitales. Nous avons préféré d'utiliser les italiques pour être cohérent avec l'ensemble de notre travail.

devient le sujet du verbe support et *ultimatum* devient son complément d'objet³⁵. Dans (b) les actants sémantiques dépendent syntaxiquement du substantif *ultimatum*.

En somme, dans la théorie Sens-Texte, la notion d'actant est conçue sous une perspective sémantique, mais elle doit aussi être considérée sous l'angle des structures syntaxiques contrôlées par les prédicats³⁶.

1.2.3.2. Le prédicat nominal dans le modèle des classes d'objets³⁷

Dans le cadre du modèle des classes d'objets, la notion de prédicat est héritière de la notion d'*opérateur* de Harris (1976) : « Les mots qui n'ont aucune contrainte sur leurs arguments sont des arguments élémentaires. Les autres sont des opérateurs³⁸ » (p. 14). D'après cet auteur, ce qui définit un prédicat est le fait de nécessiter d'autres éléments pour le compléter. Dans le lexique-grammaire, on a considéré que ce seul critère définitionnel des prédicats ne permet pas de les distinguer, par exemple, des déterminants, puisqu'eux aussi, ils nécessitent la présence d'un substantif. Ainsi, on a ajouté une autre propriété : être susceptible d'être soumis au temps.

Par conséquent, selon cette approche théorique, les propriétés définitoires du prédicat sont qu'il possède un schéma d'arguments et qu'il peut recevoir des marques de temps (G. Gross, 1989, 2012). Le nom prédicatif est un prédicat au plein sens du terme : il sélectionne des arguments et assure une relation entre eux et, dans son actualisation, il est inscrit dans le temps (moyennant des verbes supports). Le nom prédicatif est donc une réduction d'une phrase élémentaire. Soit les phrases :

³⁵ Selon le modèle de la Lexicographie Explicative et Combinatoire, cette structure collocationnelle serait modélisée moyennant la fonction lexicale *Oper₁* (cf. Mel'čuk et al., 1984, 1988, 1992, 1999).

³⁶ Nous n'aborderons pas ici la distinction que Mel'čuk (2004a, 2004b) fait entre actants sémantiques, actants syntaxiques profonds et actants syntaxiques de surface.

³⁷ Dans cette section nous ne ferons qu'une petite approximation à quelques notions de la méthode des classes d'objets pour bien saisir l'objet de notre étude, les prédicats nominaux des <aides>. Pour un approfondissement théorique à cette approche linguistique, voir le *Chapitre IV : Le modèle des classes d'objets*.

³⁸ Le terme *opérateur* chez Harris est à peu près l'équivalent de ce que l'on entend par *prédicat*.

(7) a. *Léa fait un don à la Fondation du Patrimoine.*

b. *Le don que Léa fait à la Fondation du Patrimoine.*

c. *Le don de Léa à la Fondation du Patrimoine.*

par opposition à :

(8) a. *Luc fait un gâteau à Max.*

b. *Le gâteau que Luc fait à Max.*

c. **Le gâteau de Luc à Max.*

La nominalisation permet d'effacer les verbes supports, comme nous pouvons l'observer dans l'exemple (7c). Ainsi, le nom *don* est un nom prédicatif. Dans le cas des phrases (8), étant donné que l'effacement du verbe *faire* donne lieu à une phrase agrammaticale (8c), il s'agit d'un verbe prédicatif et le nom *gâteau* n'est pas un nom prédicatif. Seuls les noms prédicatifs peuvent entrer dans des constructions à verbe support.

Les noms non prédicatifs sont des éléments linguistiques autosuffisants. Ils ne sont pas associés à un ou plusieurs arguments parce qu'ils n'ont pas de schéma d'arguments : *chien, fleur, ordinateur*. Ils sont actualisés moyennant la quantification (*deux chiens*) et la détermination (*cet ordinateur*). Contrairement aux noms prédicatifs, ils n'acceptent pas la notion de temps. Or observons les syntagmes suivants : *les fleurs du printemps dernier ; l'ordinateur d'il y a cinq ans*. Si ces substantifs non prédicatifs sont compatibles avec des indicateurs temporels, c'est parce que ces derniers ne modifient pas les substantifs mais un prédicat effacé : *les fleurs qui ont fleuri le printemps dernier / les fleurs que nous avons cueillies le printemps dernier ; l'ordinateur que nous avons acheté il y a cinq ans*.

Les noms non prédicatifs ne peuvent jamais être des prédicats, quelle que soit la phrase où ils s'inscrivent. Par conséquent, ils ne peuvent être que des arguments d'un prédicat. On les appelle des *arguments élémentaires*. Par contre, les noms prédicatifs, comme nous l'avons déjà vu, non seulement ont une structure argumentale, mais ils

sont susceptibles de remplir la fonction d'argument dans le cadre d'une phrase complexe³⁹.

Les prédicats sont définis par leur schéma d'arguments et par leur actualisation. Comme tout prédicat, le nom prédicatif constitue le noyau d'une phrase élémentaire. C'est le sens du nom prédicatif qui détermine le nombre, la forme et la nature sémantique des arguments.

En ce qui concerne l'actualisation, les prédicats nominaux sont actualisés notamment par la détermination et par les verbes supports⁴⁰. La détermination subit une double restriction, car elle dépend autant du prédicat nominal que du verbe support utilisé. Ainsi, la détermination ne pourra pas être étudiée dans le seul cadre du groupe nominal.

Le verbe support permet de « conjuguer » le nom prédicatif, autrement dit le verbe support porte les marques de temps et d'aspect externe, que le nom prédicatif ne peut pas porter. Le prédicat nominal a donc une conjugaison lexicale, à la différence du verbe, dont la conjugaison est suffixale (les désinences verbales). Comme nous l'avons montré plus haut (exemple 7), l'actualisation peut être réduite au moyen d'une nominalisation :

(9) a. *Paul a apporté son aide à Max.*

b. *L'aide que Paul a apporté à Max.*

c. *L'aide de Paul à Max.*

Le verbe support, n'ayant pas de charge sémantique, peut être effacé après une relativation. On notera que l'effacement du verbe support (9c) implique la perte de la marque de temps.

C'est le prédicat nominal qui sélectionne le verbe support en fonction de sa nature prédicative. Nous pouvons distinguer trois grands types de prédicats nominaux selon le verbe support qui les actualise. Les prédicats d'action sélectionnent les verbes

³⁹ Sur la phrase complexe, voir la section 6. *La phrase complexe* du Chap. IV.

⁴⁰ Ils peuvent être actualisés aussi par certains adjectifs et certains adverbes (cf. section 5. *Description des actualisateurs* du Chap. IV).

supports *faire* ou *effectuer*⁴¹ en français (*faire un devoir, faire des achats, effectuer un travail*)⁴², et le verbe *hacer, dar*⁴³ o *efectuar* en espagnol (*hacer la compra, hacer una inspección, efectuar un pago*)⁴⁴. Les prédicats d'état sont conjugués à l'aide des verbes supports *avoir* ou *posséder* en français (*avoir de l'innocence, posséder de grandes qualités*), et *tener* en espagnol (*tener la gripe, tener sentido común*). Enfin, les prédicats d'évènement sont actualisés par le verbe *avoir lieu* en français (*un tremblement de terre a eu lieu ; la cérémonie aura lieu*), et *tener lugar* en espagnol (*el desfile tendrá lugar, el partido de tenis tuvo lugar*)⁴⁵. Cette typologie des noms prédicatifs est donc fondée sur des critères strictement syntaxiques (de combinatoire)⁴⁶.

Cependant, cette classification résulte simpliste. Ces trois grandes classes de noms prédicatifs sont trop générales et ne permettent pas de prédire le support adéquat. Ainsi, dans le cas de notre étude, les noms prédicatifs des <aides financières> sont des actions, pourtant un grand nombre de ces prédicats ne prend pas le verbe support *faire* : *faire* (**une aide, *une allocation, *un secours, un don*) ; *hacer* (**ayuda, *un subsidio,*

⁴¹ On notera qu'*effectuer* appartient à un niveau de langue plus soutenu que *faire*. G. Gross (2012) observe une autre différence entre ces deux verbes : *effectuer* n'a pas d'autre emploi que d'être support, tandis que *faire* a un emploi prédicatif (*faire un gâteau*).

⁴² Une analyse sur un vaste corpus journalistique a permis de constater que dans le cas de prédicats nominaux appartenant au discours spécialisé, le verbe support sélectionné sera plutôt *procéder à* (*procéder à une augmentation de capital, procéder à la réfection du système électronique*). Par contre, ce verbe support ne s'emploie guère avec des prédicats d'action très généraux : **procéder à* (*un voyage, un travail, une aide*) (G. Gross, 2012).

⁴³ En espagnol le verbe *dar* est sélectionné très fréquemment par les prédicats nominaux d'action. Colominas (2001) explique qu'il y a des verbes supports dont l'extension de sens est bien plus large que d'autres. Ainsi, elle compare l'espagnol *dar* (*una vuelta, un paso, un salto*) au catalan *fer* (*un volt, una passa, un salt*) et au français *faire* (*une promenade, un pas, un saut*) (p. 7).

⁴⁴ Force est de constater que le verbe *donner/dar* est aussi très fréquemment sélectionné par les prédicats nominaux des actions : *donner un coup, dar un beso*.

⁴⁵ On notera que les prédicats nominaux d'évènement occupent la place de sujet, tandis que les prédicats nominaux d'action occupent de préférence la place d'objet. Alors, la nature sémantique du nom prédicatif détermine la place que celui-ci occupe dans la phrase.

⁴⁶ Il ne faut donc pas chercher une correspondance avec les typologies sémantiques très célèbres d'auteurs comme Vendler (1967), Lyons (1977) ou, pour le contrastif français-espagnol, Emorine (1992).

donación). Les prédicats des <aides financières> vont plutôt sélectionner les couples *apporter/recevoir*, *prodiguer/bénéficier*, *prêter/recevoir*. Il s'avère donc nécessaire de mettre au point des classes sémantiques de prédicats beaucoup plus fines. Ces sous-classes sont ce qu'on appelle les *classes d'objets*⁴⁷.

Les noms prédicatifs d'<aide > sont des prédicats d'action. La notion d'action semble simple à appréhender de manière intuitive ; pourtant elle est très diverse du point de vue sémantique. Nous considérons donc nécessaire de préciser les propriétés communes à tous les prédicats d'actions. G. Gross (2012) a établi cinq critères qui servent à définir une action.

- 1) Un sujet humain en position thématique : un nom prédicatif d'action ne peut avoir qu'un humain en position de sujet. En effet, les prédicats nominaux des <aides> prennent comme sujet des humains et des collectifs humains.
- 2) Reprise par un classifieur d'action : les prédicats d'action peuvent être repris par le substantif *acte* ou *action*. Ainsi :
 - (10) a. *Luc a apporté de l'aide à ses voisins. Cet acte montre sa générosité.*
 - b. *Léa a porté secours à son cousin. Une telle action lui a valu la reconnaissance de toute la famille.*
- 3) Reprise par l'expression *le faire* : les prédicats d'action peuvent être repris par le pronom neutre *le* suivi du verbe prédicatif *faire* (Prandi, 2004). Observons les exemples suivants :
 - (11) a. *Il a donné un coup de main à ses parents dans l'entreprise familiale, comme il le fait tous les weekends.*
 - b. **Le match a eu lieu dimanche, comme il le fait toutes les semaines.*
 - c. **Il est triste, comme il le fait tous les dimanches soir.*

⁴⁷ Pour une explication plus détaillée de la description des prédicats moyennant les classes d'objets, voir la section 3. *Description des prédicats : la notion d'emploi* du Chap. IV.

Les prédicats d'évènement (11b) et d'état (11c) n'acceptent pas cette reprise par *le faire*. Le verbe *faire* de cette expression anaphorique n'est pas un verbe support. S'il était un verbe support, le pronom devrait garder le genre du nom prédicatif :

(12) a. *L'aide, il la prête tous les weekends.*

Dans la phrase (12a) l'antécédent du pronom *la* est *l'aide*. Il fonctionne comme objet du verbe *faire*, donc ce verbe est un verbe support. Pourtant, l'antécédent du pronom *le* dans (11a) n'est pas *un coup de main*, mais toute la phrase *Il a donné un coup de main à ses parents dans l'entreprise familiale*. Dans la phrase (11a) le verbe *faire* est donc un verbe prédicatif.

- 4) Verbes supports spécifiques : à part les verbes supports généraux *faire* ou *effectuer*, qui s'appliquent à une grande partie des prédicats d'action, il existe d'autres verbes supports réservés à des classes plus restreintes, c'est-à-dire ils s'appliquent à des sous-classes de noms prédicatifs d'action. Dans le cas des noms prédicatifs des <aides>, ces verbes supports spécifiques sont, par exemple : *accorder*, *allouer*, *apporter*, *attribuer*, *prêter*, etc. D'autres verbes supports spécifiques d'autres classes sont, par exemple, *commettre* et *perpétrer* pour la classe des <crimes>, et *mener* et *livrer* pour la classe des <combats>. Ces verbes supports sont appropriés à chaque classe :

(13) a. *allouer (une allocation, *un crime, *une bataille)*

b. *commettre (*une allocation, un crime, *une bataille)*

c. *livrer (*une allocation, *un crime, une bataille)*

- 5) Adverbes appropriés : les prédicats d'action peuvent recevoir l'actualisation d'adverbes orientés vers le sujet comme *lui-même*, *personnellement* ou *en personne* :

(14) a. *Ce coup de main, il le donnera en personne.*

b. *Luc a apporté de l'aide personnellement.*

Les noms prädicatifs des <aides financières>, qui constituent l'objet de notre étude, respectent les cinq critères présentés ci-dessus. Ils constituent une sous-classe des <aides> et, par conséquent, des actions.

Pour ce qui est de la modalité de l'action, les prédicats d'action se subdivisent en deux groupes : les actions volontaires et les actions qui ne sont pas volontaires. Les noms prädicatifs des <aides> sont des actions volontaires et conscientes. La preuve est qu'ils sont compatibles avec des adverbes indiquant l'intentionnalité (*intentionnellement, avec préméditation, consciemment, délibérément, de plein gré, exprès, volontairement, etc.*) ou avec des subordonnées exprimant la finalité⁴⁸ :

(15) a. *C'est à dessein que Léa a apporté de l'aide à ses nouveaux voisins.*

b. *Léa a apporté de l'aide à ses nouveaux voisins pour faire le déménagement.*

Nous avons montré que le nom, occupant une place centrale dans les grammaires, dictionnaires et études terminologiques, reste difficile à saisir. La preuve, les multiples interprétations quant à ses caractéristiques grammaticales, sémantiques et syntaxiques. Nous avons essayé de définir brièvement ce que nous entendons par *nom prädicatif*, passant en revue d'autres interprétations.

2. Le discours de spécialité

La nomenclature du dictionnaire que nous avons élaboré comporte des noms de la langue générale et des noms de ce qu'on appelle d'habitude la langue de spécialité. Nous considérons que rien n'empêche d'étudier ensemble toutes ces unités lexicales. Ainsi, nous avons décrit avec les mêmes outils d'analyse des unités lexicales telles que *soutien, coup d'épaule* ou *aumône* et des termes tels que *capital décès, parachute doré* ou *éco crédit*. En outre, la particularité de notre recherche réside dans le fait que ces noms prädicatifs relèvent de plusieurs domaines de spécialité tels que le droit civil

⁴⁸ Nous entendons que la finalité implique que le sujet maîtrise son action, celle-ci ne pouvant pas être délibérée.

(*legs*), le droit du travail (*congé de maternité*), la législation sociale (*allocation de logement familial*), les finances (*crédit zéro*), l'éducation (*bourse*), etc.

Cette section est divisée en deux volets. Premièrement, nous commencerons par apporter quelques précisions à propos des notions de *langue de spécialité* et *discours de spécialité*. Deuxièmement, nous ferons un tour d'horizon sur les caractéristiques lexicosyntaxiques des discours de spécialité qui font l'objet de notre recherche.

2.1. Langue de spécialité ou discours de spécialité ?

Le terme *langue de spécialité* et la définition de la notion à laquelle ce terme fait référence ont été longuement débattus dans la littérature linguistique⁴⁹. Rodríguez-Piñero Alcalá et García Antuña (2009) ont rassemblé certains termes utilisés :

Lenguas de especialidad (Gómez de Enterría, 2006), *lenguajes de especialidad* (Cabré, 1993)⁵⁰, *lenguas especializadas* (Lerat, 1995), *lenguajes especializados por la temática* (Sager et al., 1980), *lenguas especiales* (Rodríguez Díez, 1981), *lenguaje de la ciencia y de la técnica* (Gili Gaya, 1964), *tecnolecto* (Haensch, 1987), *lengua técnica* (Quemada, 1978), *microlengua* (Balboni, 1982), *lengua de minoría* (Hernán Ramírez, 1979), *lenguajes con fines específicos* (de Beaugrande, 1987), *lenguajes específicos*, *lenguaje científico* (Gutiérrez Rodilla, 2005), *lenguaje sectorial* (Beccaria, 1973), *discurso científico* (Grabarczyk, 1988), *sottocodice*⁵¹ (Berruto, 1997), *sublenguaje* (Kittredge et Lehrberger, 1982), etc. (p. 910).

Cette profusion terminologique est la preuve du dissensus qui existe sur l'étiquette et sur le concept précis qu'elle désigne.

L'International Standardization Organisation (ISO) est un organisme international qui s'occupe de la normalisation (dans le sens de 'standardisation) de la terminologie et de la représentation des symboles⁵². En 1990, l'ISO 1087 définissait la *langue de*

⁴⁹ Pour une révision de la notion de langue de spécialité et les termes employés, voir Peytard et al. (1984) ; Mortureux et Petit (1989) ; Cabré (1993 ; 2004) ; Lerat (1995 ; 2012) ; Balliu (2018).

⁵⁰ Cette autrice proposera plus tard les termes *usos especializados* et *discursos especializados* (Cabré, 2004).

⁵¹ En italien dans le texte original.

⁵² L'organisme national français est l'AFNOR (Association française de normalisation) ; en Espagne, c'est l'AENOR (Asociación Española de Normalización).

spécialité comme un « sous-système linguistique qui utilise une terminologie et d'autres moyens linguistiques et qui vise la non-ambiguïté de la communication dans un domaine particulier » (ISO 1087, 1990)⁵³. Pourtant, cette définition n'a pas recueilli l'accord d'une grande partie de linguistes ni de terminologues.

Ce manque de consensus provient, en partie, de la difficulté d'établir la limite entre ce qui est général et ce qui est spécialisé dans la langue. Ainsi, nous pouvons distinguer trois grandes conceptions de la relation entre la langue générale et les langues de spécialité (cf. Gómez González-Jover, 2005 ; Rodríguez-Piñero Alcalá et García Antuña, 2009) :

- i. Les langues de spécialité sont des codes linguistiques complets, différents de la langue générale. Elles constituent donc des sous-langages au même niveau que la langue générale. Cette autonomie des langues de spécialité par rapport à la langue générale réside sur leur lexique terminologique, transparent et monosémique, et sur des différences morphosyntaxiques et syntaxiques. Les principaux représentants de cette conception des langues de spécialité sont Kocourek (1982/1991), Wüster⁵⁴ (1996, 1998) et Hoffmann (1979, 1998).

⁵³ Une trentaine d'années plus tard, l'ISO a actualisé la définition de *langue de spécialité* en tenant compte de l'extralinguistique et en incorporant les résultats des dernières recherches en phraséologie (cette discipline ayant connu un essor foudroyant à partir des années 90) : « langue utilisée dans la communication entre experts dans un domaine et caractérisée par l'utilisation de moyens d'expression linguistiques particuliers. [...] Les moyens d'expression linguistiques particuliers englobent toujours une terminologie et une phraséologie propres au domaine et peuvent également présenter des traits stylistiques ou syntaxiques » (ISO 1087, 2019).

⁵⁴ Eugen Wüster est considéré le père de la terminologie moderne et il fut le fondateur de l'école de Vienne. Il promut l'entrée de la terminologie (en tant que discipline) dans l'ISO. Ainsi, en 1952, on créa le Comité technique 37, dont il fut le secrétaire pendant de longues années. On peut donc bien observer l'héritage de l'approche de Wüster dans la définition de l'ISO 1087 (1990) ci-dessus (idéal de monosémie et d'univocité). Tout au long de sa carrière il fit de grands efforts pour obtenir la reconnaissance de la terminologie en tant que discipline académique. En 1971, il obtint sa reconnaissance institutionnelle lors de la création par l'UNESCO de l'organisme INFOTERM (International Information Centre of Terminology), dont la tâche est d'encourager et coordonner la coopération internationale dans le domaine de la terminologie.

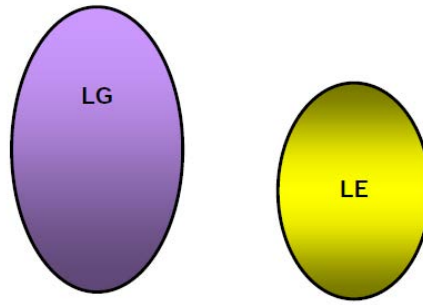


Figure 1 : *Relation (type i) entre la langue générale (LG) et la langue de spécialité (LE)*
(extrait de Gómez González-Jover, 2005, p. 38)

- ii. Les langues de spécialité sont des variantes lexicales de la langue générale. Ainsi, Mounin (1979) soutient que « au sens propre, il n'existe pas de langue du droit en soi, mais seulement, à l'intérieur de la langue française, un vocabulaire du droit et sans doute quelques tours syntaxiques spécifiques » (p. 13). La seule spécificité des langues de spécialité est donc le lexique. En ce qui concerne les caractéristiques, syntaxiques, sémantiques et énonciatives, rien ne les différencie de la langue générale⁵⁵. C'est l'optique de nombreux linguistes provenant de la linguistique théorique ou descriptive tels que Rey (1976), Quemada (1978), Mounin (1979) et Rondeau (1984).

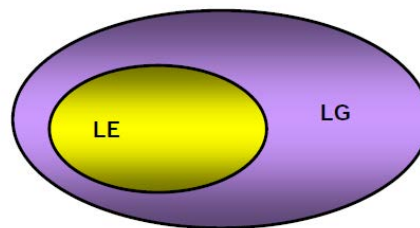


Figure 2 : *Relation (type ii) entre la langue générale (LG) et la langue de spécialité (LE)*
(extrait de Gómez González-Jover, 2005, p. 38).

- iii. D'autres linguistes ont pris une position intermédiaire selon laquelle les langues de spécialité sont des sous-ensembles dépendants de la langue générale, mais elles ont des éléments spécifiques propres, essentiellement lexicosyntaxiques. Comme le signalent Sager et al. (1980), ces spécificités relèvent de la pragmatique et de l'extralinguistique :

There is no absolute borderline between general and special language as the relative proportion of special terms may at best permit merely a distinction between

⁵⁵ À l'exception de quelques usages combinatoires propres à chaque langue de spécialité.

retrospective and innovatory speech acts. The threshold between general and special language can be delineated only by pragmatic criteria. (p. 68)⁵⁶

Il ne s'agirait donc plus d'une relation de juxtaposition entre langue générale et langue de spécialité (première conception) ni d'une relation d'inclusion (seconde conception), mais plutôt d'un continuum⁵⁷ entre langue générale et langue de spécialité. Cette optique est défendue par des auteurs comme Varantola (1986), Sager et al. (1980), Picht et Draskau (1985), Cabré (1993), Lerat (1995, 1997), Schifko (2001) ou Mejri (2012), entre autres.

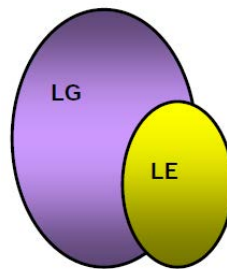


Figure 3 : Relation (type iii) entre la langue générale (LG) et la langue de spécialité (LE)

(extrait de Gómez González-Jover, 2005, p. 39)

C'est sous cette dernière conception des langues de spécialité que nous avons entrepris notre travail de recherche. Nous sommes d'accord avec Lerat (1997) lorsqu'il affirme que « les langues de spécialité ne sont rien d'autre que des usages spécialisés des langues naturelles » (p. 15). Une langue spécialisée n'est que l'usage des règles de la langue générale pour rendre compte techniquement de connaissances spécialisées (Lerat, 1995, p. 21). Cette optique pragmatique sera suivie aussi par Cabré (2004) lorsqu'elle affirmera : « no podemos hablar de 'lenguajes especializados' *strictu sensu*, sino de usos especializados a partir de una misma lengua o de discursos especializados » (Cabré, 2004, p. 23). Pour cette linguiste, l'utilisation du mot

⁵⁶ Nous tenons à préciser que ces auteurs séparent les langues de spécialité d'autres types de sous-ensembles tels que les dialectes et les sociolectes. Ainsi, on distingue d'une part les sous-ensembles géographiques et sociaux, qui font partie de la langue générale et qui sont acquis par le locuteur de manière inconsciente ; et d'autre part, des langues de spécialité, qui requièrent un apprentissage conscient.

⁵⁷ Selon Mejri (2012) la question de la spécialisation doit être appréhendée de manière scalaire : « entre le discours général et le discours spécialisé se situe le discours de vulgarisation qui, sans être totalement expert, essaie de transmettre des connaissances spécialisées à travers un discours plus ou moins général » (p. 19-20).

« langage »⁵⁸ est abusive car celui-ci est un ensemble d'unités linguistiques et de règles qui décrivent les textes (écrits et oraux) produits dans une situation de communication déterminée. Une langue possède une grammaire propre (une phonologie, une syntaxe, une morphologie et un lexique) qui la décrit. Cependant, l'observation des caractéristiques des ressources linguistiques utilisées dans des situations de communication spécialisées révèle qu'elles appartiennent à l'ensemble de ressources de la langue générale.

Dans ce sens, Mejri (2012) soutient que « la langue est commune et la spécialisation intervient uniquement lors des réalisations discursives. [...] Dans tout discours spécialisé, il y a du discours général, lequel discours sert de matrice au premier » (p. 19-20). Le code linguistique appartient donc à la langue générale. Les fondements du discours de spécialité sont plutôt d'ordre extralinguistique. C'est la situation de communication qui va déterminer la spécificité de la langue. Ainsi, Mejri (2012) définit le *discours spécialisé* comme :

une production réalisée par un locuteur expert à l'adresse d'un interlocuteur expert ; l'expertise étant mesurée par la compétence qu'ont ces interlocuteurs à produire et à interpréter ces discours conformément aux contenus conceptuels délimités dans le domaine par la communauté des experts à un temps T. (p. 22)

Étant donné que les spécificités linguistiques de ce qu'on a traditionnellement appelé les « langues de spécialité » sont en fait des particularités discursives, nous préférons donc le terme *discours de spécialité*⁵⁹.

⁵⁸ Il faut comprendre *langage* comme synonyme de *langue*.

⁵⁹ Nous utiliserons indifféremment *discours spécialisé* et *discours de spécialité*. Par ailleurs, Lerat (2012) fait la distinction entre les termes *technolecte* (« un *savoir-dire* qui comporte une terminologie, les collocations et la phraséologie pertinentes dans les limites d'une pratique langagière plus ou moins technique », p. 33), *langue spécialisée* (« une langue naturelle considérée en tant que vecteur de connaissances spécialisées », Lerat, 1995, p. 20) et *discours spécialisé* (« un ensemble d'énoncés pris comme corpus et formulé par un expert ou semi-expert à l'intention d'experts et/ou semi-experts », Lerat, 2012, p. 34). Autrement dit, il différencie la composante cognitive (le technolecte), la composante linguistique (la langue spécialisée) et la composante sociocommunicative (le discours).

2.2. Caractéristiques des discours de spécialité de notre corpus

L'étude du discours de spécialité ne se réduit pas à des nomenclatures de termes ou à des réseaux conceptuels. Dans ces textes, on utilise des dénominations dans des énoncés mobilisant les ressources de la langue générale. Ainsi, nous souscrivons à l'affirmation de Massoussi et Sfar (2009) : « une théorie des langues spécialisées est basée sur une conception multiple, à la fois morphologique, syntaxique, sémantique et énonciative⁶⁰ » (p. 66).

Les discours spécialisés présentent certaines caractéristiques linguistiques particulières. Nous n'entrerons pas dans le détail des caractéristiques linguistiques de chaque domaine de spécialité recouvert par la nomenclature de notre dictionnaire, parce que cette démarche risquerait de devenir répétitive et pénible. Nous ne ferons donc que survoler quelques caractéristiques communes. Considérant l'affirmation de Lerat (1997) : « [la spécificité des langues de spécialisés] est essentiellement lexicosyntaxique » (p. 15), nous allons présenter brièvement leurs caractéristiques au niveau lexical et au niveau syntaxique.

2.2.1 Caractéristiques lexicales des discours de spécialité

La marque la plus évidente de la spécificité du discours de spécialité est l'emploi d'un lexique particulier. La technicité du discours spécialisé est variable, en fonction des besoins de communication et de la situation de communication. Cabré (2005) précise que les situations de communication spécialisée ne se limitent pas à un échange entre spécialistes. La notion de situation de communication spécialisée englobe donc d'autres circonstances où la précision du contenu sera préservée, du fait que le producteur du discours est un bon connaisseur du domaine (quoique pas nécessairement un spécialiste). Ainsi, dans certains cas le locuteur fera un usage limité de la

⁶⁰ « L'énonciation peut être définie, avec E. Benveniste (1974), comme "la mise en fonctionnement de la langue au moyen d'un acte individuel d'utilisation", un acte pris en charge par un énonciateur donné, dans un cadre spatiotemporel donné, et destiné à un co-énonciateur (ou énonciatrice) donné » (Neveu, 2011, p. 143-144). Pour approfondir sur l'énonciation et l'analyse du discours, voir E. Benveniste (1966-1974), Charaudeau (1992), Charaudeau et Maingueneau (2002), Culioli (1990) et Jakobson (1963), entre autres.

terminologie spécialisée (par exemple, lorsqu'un spécialiste d'un domaine parle avec un non spécialiste), mais dans d'autres cas on utilisera même des signes non linguistiques (comme les formules de notation des sciences)⁶¹. Cabré (1993) observe une certaine gradation de spécialisation en fonction du domaine :

Los « lenguajes » de la física, la química, la biología, la geología, la matemática, la estadística, la lingüística, la antropología, la historia, la arquitectura, la estética, la economía teórica, etc., presentan muy alto grado de especialización: son claramente lenguajes de especialidad. Otros « lenguajes » como el de la banca, la bolsa, el derecho o la economía aplicada, constituyen un terreno intermedio entre los lenguajes más especializados y los más generales. Finalmente, los de la restauración, la peluquería, la ferretería, los deportes o los remedios caseros, presentan un grado muy menor de especialización, y se hallan en la frontera de la lengua común (p. 144).

En effet, le lexique de certains domaines de spécialisation hautement scientifiques atteint un niveau d'abstraction et de formalisation très élevé, tandis que d'autres présentent peu de spécialisation. De plus, les limites entre langue générale et langue spécialisée ne sont pas totalement tranchées. Cette linguiste préfère l'envisager comme un continuum entre les discours les plus généraux et les discours les plus spécialisés.

Dans le cas de notre étude, les termes qui constituent la nomenclature de notre dictionnaire n'appartiennent pas à des domaines hautement spécialisés. D'ailleurs, il existe des termes qui sont très communs (*allocation*). Nous comprenons que comme il s'agit d'aides, dont la raison d'être est que les gens puissent en bénéficier, on évite de compliquer la communication avec un discours hautement spécialisé et hermétique. Encore une fois nous montrons que le discours de spécialité est fortement marqué par la compétence pragmatique.

Par ailleurs, tout discours spécialisé comprend trois types de lexique (Català, 2003) :

- a) Environ les trois quarts de mots d'un texte spécialisé appartiennent à la langue générale.

⁶¹ Il ne faudrait pas confondre le fonctionnement des langues naturelles (que ce soit le discours général ou spécialisé) avec celui d'autres systèmes sémiotiques, comme celui de la chimie, qui utilise des formulations du type H_2O ou Na_2SO_4 . Ce n'est pas le cas des domaines abordés dans notre étude. Pour ça, nous ne dirons rien ici sur la spécificité sémiotique.

- b) Il existe un grand nombre de mots de la langue générale qui adoptent un sens spécifique lorsqu'ils sont utilisés dans un discours spécialisé.
- c) Le lexique spécialisé n'a qu'une faible incidence sur la totalité d'un texte scientifique ou technique.

Dans le discours spécialisé, c'est aussi le cas des discours relevant des domaines de spécialité recouverts par notre dictionnaire, il existe une profusion de noms composés, de néologismes et de sigles. Dans ce qui suit, nous allons présenter à grands traits les caractéristiques des noms qui constituent la nomenclature de notre dictionnaire électronique.

2.2.1.1. Présence de noms composés

Les textes de spécialité sont fortement constitués d'unités lexicales complexes dans un sens large : mots composés et d'autres types d'unités phraséologiques. Dans le cas de notre étude, les noms composés constituent la majorité des noms prédicatifs que nous avons recensés. Nous avons observé que la nature morphologique des substantifs est différente s'ils relèvent de la langue générale (ils sont en grande partie des unités lexicales simples) ou du discours spécialisé (ils sont, pour la plupart, des unités polylexicales⁶²).

La composition fait partie des procédés de formation de mots les plus exploités des discours spécialisés. La profusion d'unités polylexicales pourrait s'expliquer par la nécessité de préciser des concepts de manière transparente, ce qui demanderait des unités lexicales supérieures au simple mot. En effet, la création de nouveaux termes simples pour chaque nouveau type d'aide mis en place, donnerait lieu à une énorme prolifération de termes, ce qui compliquerait la communication dans ces discours de spécialité. Pour la dénomination des concepts scientifiques ou techniques, il est plus rentable d'utiliser des mécanismes de formation syntagmatique. Voici quelques exemples : *aide à l'embauche, aide au recrutement, aide au développement, aide à l'amélioration de l'habitat, aide aux chômeurs créateurs ou repreneurs d'une entreprise, aide à la réinsertion dans le pays d'origine*, etc. Même si l'on ne connaît pas

⁶² Mejri (2006) définit la polylexicalité comme « une caractéristique morphologique propre aux unités lexicales formées de plusieurs unités lexicales et dont les constituants sont, à l'origine, des unités autonomes » (p. 214).

exactement ce que chacun de ces noms d'aides désigne exactement, on peut bien se faire une idée plus ou moins précise grâce à leur formulation syntagmatique.

Sfar (2012) souligne que « la différence spécifique entre la dénomination polylexicale ou analytique par rapport à la dénomination monolexicale⁶³ ou synthétique se situe au niveau sémantique » (p.22). Ainsi, dans les discours spécialisés, les unités monolexicales tendent vers l'opacité, tandis que les unités polylexicales tendent vers la transparence sémantique. Des noms tels que *legs* ou *libéralité* semblent être plus opaques que *pension pour inaptitude au travail* ou *allocation compensatrice pour frais professionnels*.

Les unités polilexicales se situent entre le mot et le syntagme, c'est-à-dire elles fonctionnent comme un mot, mais elles ont la forme d'un syntagme. Nous pouvons observer que la structure syntaxique de ces noms composés respecte les propriétés typologiques et grammaticales de la langue générale. En effet, Lerat (1997) signale que ces noms composés n'ont pas de spécificité grammaticale par rapport à la langue générale. Cette caractéristique rend plus difficile la reconnaissance des noms composés. Rien n'indique que *bourse de thèse* (N de N) ou *aide alimentaire* (N + Adj) sont des expressions lexicalisées.

2.2.1.2. Présence de néologismes

La réalité est changeante, les types d'aides le sont aussi. Il est nécessaire d'attribuer un nom à chaque nouvelle allocation ou à chaque nouveau type de crédit. D'ailleurs, les termes ont une fonction dénominative⁶⁴.

Certains domaines de spécialité, comme le domaine de l'économie, comptent de nombreux emprunts, notamment de l'anglais⁶⁵. Nous n'avons pas repéré un nombre d'emprunts significatif du point de vue quantitatif. Parmi les quelques emprunts présents dans les discours de spécialité analysés, nous avons trouvé, par exemple,

⁶³ Mejri (2006) définit la monolexicalité comme « une caractéristique morphologique propre aux unités lexicales réductibles à une unité autonome, qu'elle soit simple ou dérivée par affixation » (p. 214).

⁶⁴ La fonction dénominative consiste à désigner les nouvelles réalités scientifiques ou techniques par des noms.

⁶⁵ Sur le discours de spécialité de la Bourse, voir Català (2000) et de Hoyos Puente (2002).

golden parachute ou *parachute doré/paracaídas dorado* (qui d'ailleurs est un terme du domaine de l'économie).

2.2.1.3. Présence de sigles

Les sigles sont souvent utilisés pour désigner certains noms d'allocations, notamment en français, dont l'emploi des sigles (en général) est beaucoup plus fréquent qu'en espagnol : *APL, aide personnalisée au logement* ; *AAH, allocation aux adultes handicapés* ; *ARE, allocation d'aide au retour à l'emploi*. Étant donné que ces noms d'aides sont des unités polylexicales, parfois très longues, les sigles se révèlent très utiles. Même si certains sigles sont largement connus, comme *APL*, la plus grande partie des sigles reste comprise seulement par les experts ou les personnes affectées.

2.2.1.4. Présence de langage figuré

Certains auteurs ont souligné le rôle fondamental des métaphores dans la création de termes et dans la conceptualisation même des domaines de spécialité (cf. Assal, 1995 ; Cortès, 2003 ; Kocourek, 1982/1991 ; Prandi, 2016 ; Prandi et al., 2013). Ce n'est pas le cas de la nomenclature de notre dictionnaire, car nous trouvons très peu de noms provenant d'une métaphore (p. ex., *parachute doré*).

De Hoyos Puente (2002) explique le caractère métaphorique de certains discours spécialisés, comme dans le domaine de la Bourse, par le fait qu'ils sont très souvent produits dans le milieu journalistique. Il explique que dans ce milieu il existe une tendance générale à utiliser un discours fortement figuré. Au contraire, dans le milieu académique la présence de métaphores est bien plus réduite car on privilégie un style plus sobre et épuré de figures littéraires. Dans le cas de notre recherche, nous avons trouvé très peu d'occurrences dans des textes journalistiques. Il a fallu consulter des textes produits par des organismes officiels et des associations pour pouvoir trouver un maximum d'occurrences. Ces textes ayant la finalité d'informer de la manière la plus claire et simple sur les caractéristiques, les conditions, les démarches, etc., de différents types d'aides, le langage figuré est très peu présent.

2.2.2. Caractéristiques syntaxiques des discours de spécialité

Comme Lerat (1995) l'affirme, « une langue spécialisée ne se réduit pas à une terminologie⁶⁶ : elle utilise des dénominations spécialisées (les termes), y compris des symboles non linguistiques, dans des énoncés mobilisant les ressources ordinaires d'une langue donnée » (p. 20-21). En effet, la production de textes⁶⁷ de spécialité demande la mobilisation de compétences linguistiques plus larges, qui sont celles de la langue générale. Au niveau syntaxique, il n'existe pas de règles propres à quelconque discours spécialisé.

Pourtant, ce linguiste soutient que dans le discours de spécialité on peut repérer des « habitudes d'expression statistiquement dominantes » (Lerat, 1995, p. 75), c'est-à-dire un style. Ainsi, les textes de spécialité ont d'habitude un style abstrait (obtenu par la présence de nominalisations et d'adjectivations), impersonnel (provenant notamment de l'emploi de la voix passive) et explicite (les articulations logiques sont explicitées par des connecteurs⁶⁸) (Lerat, 1997).

3. Terminologie et dictionnaires

Étant donné que l'objet de notre recherche comprend un grand nombre de termes, nous avons considéré nécessaire de présenter les bases théoriques à partir desquelles on entreprend ce type d'études actuellement.

Sager (1990) définit la terminologie comme « the study and the field of activity concerned with the collection, description, processing and presentation of terms » (p. 2). Le terme *terminologie* peut donc désigner l'ensemble de termes utilisés dans l'exercice d'une science ou technique (« the field of activity »), ou il peut désigner aussi l'étude des termes utilisés dans des discours spécialisés (« the study »). C'est notamment dans la seconde acception que nous allons utiliser ce terme dans cette section de notre travail.

⁶⁶ Ici, il faut entendre par *terminologie* un ensemble ou une liste de termes.

⁶⁷ Nous prenons *texte* dans le sens large de texte écrit ou oral.

⁶⁸ Nous tenons à signaler que les discours de spécialité n'ont pas de connecteurs spécifiques à leur domaine, sauf dans le domaine du droit, qui a un fort caractère archaïsant (p. ex., *en foi de quoi*).

L'étude des termes peut être appréhendée dans le cadre de deux approches théoriques différentes, donnant lieu à des produits terminographiques différents. Ainsi, dans l'*optique classique* ou *conceptuelle* on produit des glossaires et des bases de données terminologiques, tandis que dans l'*optique lexicosémantique*, les produits terminologiques ont un fort penchant lexicographique.

Dans cette section, nous allons faire un bref aperçu des deux grandes approches en terminologie. Ensuite, nous présentons les produits terminographiques dérivés de ces deux approches. Finalement, nous analysons les caractéristiques des dictionnaires spécialisés des domaines couverts par la nomenclature de notre dictionnaire.

3.1. La terminologie

Au XVII^e siècle la terminologie connaît un essor important. Les scientifiques commencent à faire des recherches sur les dénominations qui garantissent une communication univoque. Leur souci est d'identifier les objets, de décrire leurs propriétés et de les organiser ; autrement dit identifier, caractériser et classer (Depecker, 2005). Les révolutions industrielles du XIX^e siècle et la révolution technologique du XX^e ont conduit les linguistes à établir les bases théoriques de la terminologie en tant que discipline autonome et interdisciplinaire⁶⁹.

Dans cet élan linguistique de la terminologie nous distinguons deux grandes approches pour appréhender l'étude des termes : l'*optique classique* conceptuelle et l'*optique lexicosémantique*. Avant de nous attarder sur les différences entre ces deux approches en terminologie et les conséquences que leurs postulations théoriques auront sur leurs produits (les dictionnaires), nous considérons nécessaire d'explicitier quelques notions de base.

3.1.1. L'optique conceptuelle

⁶⁹ Cabré (2003) fait une intéressante synthèse sur les origines et l'évolution des différents modèles théoriques de la terminologie.

Cette optique, dite *classique*, est représentée par la Théorie Générale de la Terminologie (TGT), mise en place par Wüster (1996, 1998) vers les années 30⁷⁰. Dans cette approche, la terminologie est conçue comme une discipline autonome interdisciplinaire. L'ensemble des termes d'un domaine de spécialité donné sont le reflet de la structuration des connaissances dans ce domaine. On considère qu'un domaine de spécialité est un ensemble d'objets ou de concepts⁷¹ qui peuvent exister de manière indépendante à la langue ; les termes sont la réalisation linguistique du concept, c'est-à-dire une sorte d'étiquettes de la réalité. Alors, le terme est une unité sémiotique résultante de l'association entre un concept et une dénomination. Le concept précède la dénomination. Par conséquent, le point de départ de la démarche terminographique⁷² est le concept et les relations entre les concepts d'un même domaine. Il s'agit donc d'une démarche onomasiologique.

Ainsi, le terminographe isole un concept, il fixe son sens moyennant une définition et ensuite il cherche toutes les formes linguistiques utilisées pour nommer ce concept. Ces formes différentes sont traitées comme des synonymes véritables ou parfaits. Les concepts d'un même domaine de spécialisation entretiennent des relations. L'ensemble de ces relations constitue ce qu'on appelle la *structure conceptuelle* de ce domaine. C'est cette structure conceptuelle qui guide le terminographe dans son travail de description, spécialement dans la sélection des termes et dans l'élaboration des descriptions, puisque celles-ci sont basées sur les relations que les concepts entretiennent entre eux (L'Homme, 2004). La valeur d'un terme est donc établie à partir de la place que celui-ci occupe dans la structure conceptuelle.

⁷⁰ Sur les principes théoriques de la terminologie conceptuelle, voir Felber (1984), Picht et Daskau (1985) et Rondeau (1984).

⁷¹ Selon L'Homme (2004) « le concept est une représentation mentale qui retient les caractéristiques communes à un ensemble d'objets. Les objets du monde réel sont tous différents mais il est raisonnable de penser que la représentation que nous nous en faisons retient l'essentiel de leurs caractéristiques » (p. 25). La terminologie à optique conceptuelle considère le concept comme une abstraction qui existe de manière indépendante au terme et à la langue.

⁷² La terminographie est plutôt penchée sur l'élaboration d'un produit. Cette notion regroupe les différentes activités d'acquisition, de compilation et de gestion de termes. La terminologie est théorique : elle propose un cadre conceptuel pour l'étude des termes.

L'objectif de la terminologie conceptuelle est de normaliser ou standardiser les concepts et les termes afin de garantir la précision et univocité dans la communication professionnelle monolingue ou plurilingue. La terminologie cherche donc à supprimer les ambiguïtés produites par des phénomènes comme la synonymie et la polysémie. Cette optique privilégie la biunivocité : la relation entre le concept et la forme linguistique est de un à un, c'est-à-dire à une forme linguistique ne lui correspond qu'un concept, et un concept ne peut être exprimé que par une seule forme linguistique. Ainsi, le terme est considéré univoque, unidisciplinaire, monoréférentiel, uniforme (sans variation orthographique ou morphologique) et statique, il n'appartient qu'à un domaine de spécialité et il est universel, il ne dépend ni des cultures ni des langues.

3.1.1.1. Mise en cause de l'optique conceptuelle

L'on a accusé la TGT d'être réductionniste, idéaliste et ascientifique. Ainsi, Cabré (2003) signale : « Wüster developed a theory about what terminology should be in order to ensure unambiguous plurilingual communication, and not about what terminology actually is in its great variety and plurality » (p. 167).

En ce qui concerne l'approche onomasiologique, même si la délimitation d'un concept guide le terminographe dans sa démarche, normalement celui-ci suit le chemin inverse : d'abord il repère les termes dans des textes spécialisés et ensuite il appréhende le sens. Il adopte donc une démarche sémasiologique.

Le principe de biunivocité est aussi très critiqué. En effet, la monosémie est peu fréquente. D'ailleurs, elle n'est même pas nécessaire pour que la communication soit univoque. La même forme linguistique peut être le support de diverses significations, qui dépendent du contexte donné, et un même sens peut être exprimé sous différentes formes⁷³. Pour l'illustrer, L'Homme (2004) montre que même le domaine de la terminologie n'a pas réussi à neutraliser la polysémie et la synonymie :

⁷³ Les continuateurs du modèle de Wüster ont introduit la notion d'*univocité contextuelle*, afin de relativiser le concept de biunivocité. De cette manière, l'univocité se donne dans un usage concret du terme, dans un contexte donné.

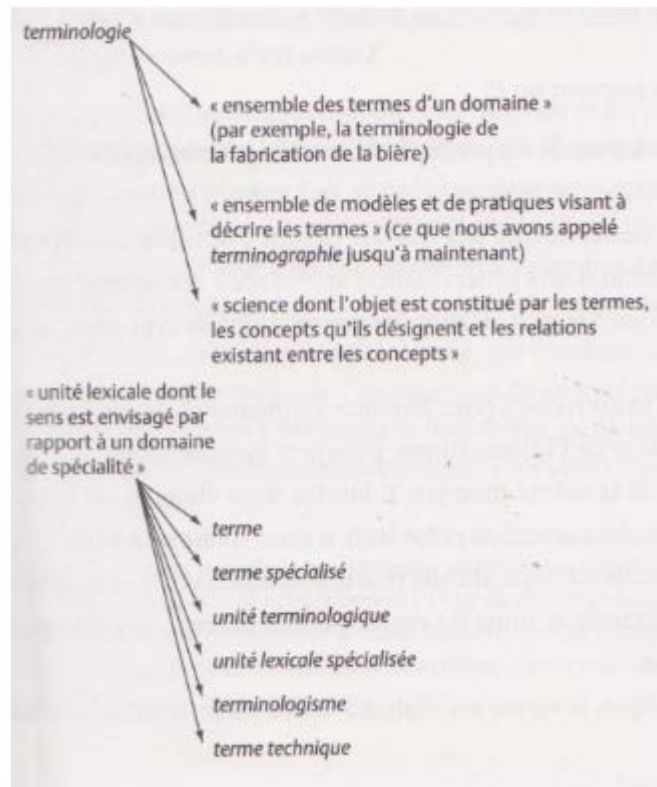


Figure 4 : *Polysémie et synonymie en terminologie* (extrait de L'Homme, 2004, p.31)

En outre, la structure conceptuelle d'un domaine de spécialité donné n'est pas composée de blocs statiques, uniformes et avec des limites précises. À ce sujet, Sager (1990) affirme : « knowledge structures are not absolute. [...] Conceptual systems are relatively fluid entities constantly undergoing change » (p. 13). Les structures conceptuelles sont donc dynamiques, elles varient avec le temps et les limites sont floues, en fonction du contexte de communication. D'ailleurs, Sfar (2010) souligne que les termes peuvent subir des modifications, voire ils peuvent perdre leur statut de terme lorsqu'ils sont utilisés en dehors d'un domaine de spécialité donné dans lequel on leur a assigné ce statut terminologique. Par exemple, plus haut, nous avons expliqué que la notion de prédicat de la logique mathématique a été empruntée par les linguistes de la théorie Sens-Texte pour désigner une autre notion avec laquelle elle garde un certain rapport, mais il s'agit d'une notion différente. Il s'est donc produit un glissement sémantique, car ce terme a été repris (moyennant une métaphore) dans un autre domaine de spécialité différent.

Quant à l'objectif de normalisation, on l'a considéré fictif. La diversité et la variation sont inhérentes aux langues naturelles, elles relèvent de l'utilisation réelle de la langue. Il ne semble pas possible de standardiser l'usage des locuteurs ni les

situations dans lesquelles la communication se produit. Ainsi, une normalisation quelconque ne pourrait jamais refléter le fonctionnement réel des termes. En outre, ce caractère prescriptif, selon certains auteurs, fait que cette optique conceptuelle soit ascientifique (Cabré, 2004).

Il est évident que la terminologie ne se compose pas seulement d'une dimension cognitive et d'une dimension linguistique. Il s'avère nécessaire d'y ajouter la dimension communicative, qui rendra possible l'étude des termes dans leur contexte réel, divers et changeant⁷⁴.

3.1.2. L'optique lexicosémantique

Depuis, des changements scientifiques ont influé sur le statut épistémologique de la terminologie. Parmi ces changements, l'ouverture de la linguistique à la sémantique et à la pragmatique a été capitale. La linguistique a abandonné son approche prescriptive du langage en faveur d'une approche descriptive. Elle a ainsi reconnu l'existence de la diversité et de la variation. De même, la terminologie a privilégié une approche moins restrictive. Elle s'est ouverte à la description de la complexité et la variabilité des termes dans un contexte global.

D'autre part, la puissance des outils informatiques a permis le développement de la linguistique de corpus. La facilité de stocker et d'exploiter des corpus de millions de mots a contribué à la mise en cause de la démarche onomasiologique. De ce fait, la terminographie se basera sur des corpus en format électronique et les termes seront analysés dans les contextes communicatifs dans lesquels ils apparaissent (Sager, 1990). Les corpus informatiques favorisent donc que les différentes approches linguistiques rendent compte de manière objective de la réalité langagière.

À la suite de ces innovations et aussi dans le but de combler les lacunes de la terminologie classique, de nouvelles approches en terminologie ont été proposées. La *théorie communicative de la terminologie* (TCT) de Cabré (1993, 1998-1999, 1999, 2000) met en relief l'importance de la dimension communicative du terme, ainsi que des aspects cognitifs et linguistiques. La *socioterminologie* de Gaudin (1993) met l'accent

⁷⁴ Pour approfondir sur les arguments qui mettent en cause l'approche conceptuelle, voir Gaudin (1993), Condamines (1995), Cabré (1998-1999), Bourigault et Slodzian (1999) et Temmermann (2000). Budin (2001) publie une réaction aux critiques faites à la TGT.

sur la dimension sociale de la terminologie. La *théorie sociocognitive de la terminologie* de Temmerman (2000) intègre des éléments de la théorie du prototype et propose la prise en compte des dimensions diachronique et sociale des termes. Enfin, Bourigault et Slodzian (1999) présentent une approche dans laquelle le texte constitue le point de départ pour décrire les termes ; c'est ce qu'on appelle la *terminologie textuelle*⁷⁵.

Le point commun des nouveaux modèles en terminologie est qu'on étudie le terme en contexte, dans le discours. Il s'agit donc d'approches résolument sémasiologiques dont l'objet d'étude est le terme conçu comme un type d'unité lexicale. Le terme fait donc partie du lexique d'une langue. Ce qui différencie les termes des autres unités lexicales est leur sens spécialisé. Autrement dit, le terme est une unité lexicale possédant une valeur spécialisée, qui sera activée en fonction des valeurs pragmatiques (des situations de communication spécialisée).

Depuis quelques décennies les terminologues ont élargi leur centre d'intérêt : ils ne se limitent plus au seul terme, mais ils tiennent compte aussi de l'environnement linguistique dans lequel un terme donné apparaît. Cet élargissement du champ d'étude s'est avéré pertinent, car la combinatoire des termes permet de distinguer des unités lexicales différentes (qui constitueront donc des entrées indépendantes dans le dictionnaire).

Le terme sera défini en observant l'ensemble des relations qu'il entretient avec d'autres unités lexicales, autant sur le plan syntagmatique que paradigmatic. En effet, le terme étant une unité lexicale, il appartient à un gigantesque réseau qui est le lexique. On s'intéresse à l'étude des représentations des relations entre les sens.

3.2. La terminographie

Dans les années 1970, Alain Rey propose de distinguer d'une part, la théorie de la terminologie et d'autre part, l'application de cette théorie pour l'étude des termes. La première s'appellera *terminologie*, et la seconde, *terminographie*. Nous citons L'Homme (2004) pour expliquer brièvement cette différence : « La terminologie regroupe les diverses activités d'acquisition, de compilation et de gestion des termes. La

⁷⁵ Pour un survol critique de ces nouvelles approches en terminologie, voir L'Homme (2005).

terminologie se penche sur les questions fondamentales que soulève l'étude des termes et propose un cadre conceptuel pour les appréhender » (p. 16).

Ainsi, sous le chapeau de *terminographie* on trouve deux types de produits différents qui relèvent de deux optiques différentes : l'optique conceptuelle et l'optique lexicosémantique. Ainsi, l'optique conceptuelle est appliquée à l'élaboration de terminologies, glossaires ou d'autres produits terminographiques, tandis que dans l'optique lexicosémantique on propose des produits terminographiques à caractère lexicographique, donnant lieu à des dictionnaires de spécialité.

Sager (1990) observe trois facteurs qui expliqueraient la différence entre la terminologie et la lexicologie :

Terminology is the study of and the field of activity concerns with the collection, description, processing and presentation of terms, i.e. lexical items belonging to specialized areas of usage of one or more languages. In its objectives it is akin to lexicography which combines the double aim of generally collecting data about the lexicon of a language with providing an information, and sometimes even an advisory, service to language users. The justification of considering it a separate activity from lexicography lies in the different nature of the data traditionally assembled, the different background of the people involved in this work, and to some extent in the different methods used. (p. 2)

Alors, pour Sager la différence réside dans le type de données recueillies, les profils des personnes qui travaillent dans l'élaboration des produits et les méthodes utilisées. Par contre, Cabré, (2005) met en cause cette différence :

L'on pourrait se demander pour quelle raison [...] on a insisté durant tant d'années sur la séparation des méthodologies entre la lexicographie et la terminographie, et on a tant souligné la différence radicale entre un produit lexicographique et un produit terminologique. La raison se situe à nouveau dans les positions initiales de la terminologie wülsterienne. (p. 199)

Cette autrice considère que ces différences ne se situent que sur les positions de l'optique conceptuelle classique. La méthodologie et les produits de la terminologie lexicosémantique ne seraient donc pas si éloignés des dictionnaires (des produits lexicographiques).

En effet, nous avons déjà abordé certaines différences entre l'optique classique et l'optique lexicosémantique et nous pouvons constater qu'elles correspondent à ce que pour Sager (1990) marque la différence entre terminologie et lexicologie. Ainsi, si l'objet d'étude de ces deux optiques est différent (le concept pour les uns et le terme

comme unité linguistique pour les autres), la démarche est différente (onomasiologique ou sémasiologique) et le but est différent (cerner des concepts ou décrire des sens), la méthodologie et le produit résultant seront forcément différents.

Nous présentons dans le tableau ci-dessous les différences fondamentales entre l'optique conceptuelle et l'optique lexicosémantique. Les informations concernant l'optique conceptuelle sont basées sur les préceptes de la TGT de Wüster⁷⁶, qui d'ailleurs n'ont pas toujours été mis en pratique par ses adeptes.

CRITÈRES	OPTIQUE CONCEPTUELLE	OPTIQUE SÉMANTICOLEXICALE
Objet d'étude	Le concept, auquel on associe un terme (qui n'est que la représentation linguistique de ce concept).	Le terme en tant qu'unité lexicale.
Démarche	Onomasiologique : on va du concept à la dénomination.	Sémasiologique : on va de la forme linguistique au sens.
But	Représenter des concepts spécialisés.	Décrire des termes (sémantique lexicale) et/ou fournir le terme équivalent en d'autres langues
Fonction	Normalisation (standardisation) des termes d'un domaine de spécialité	Description et/ou traduction des termes.
Sélection de la nomenclature	À partir de la structure conceptuelle du domaine donné.	Repérage des termes dans des corpus spécialisés.
Entrée	Un concept (il est parfois identifié moyennant un code).	Un terme (parfois appelé <i>lemme</i>).
Nomenclature	Généralement des noms (parfois	Généralement des noms (parfois

⁷⁶ Nous avons déjà expliqué (cf. note 73) que les continuateurs de Wüster ont introduit des modifications sur certains aspects ; par exemple, ils ont étendu leur domaine d'étude, mais ils ont conservé les fondements établis. Nous n'analyserons pas ici les différences entre la TGT « pure » (telle qu'elle avait été conçue par Wüster) et ce que Cabré (2005) a appelé le *modèle des portes* ou la théorie communicative de la terminologie (TCT). On trouvera dans Cabré (2005) une brève comparaison éclairante.

(catégories grammaticales)	des adjectifs et des verbes, mais en moindre mesure).	des adjectifs, des verbes et des adverbes mais en moindre mesure).
Variations⁷⁷	Il n'y en a pas.	Il y en a.
Ordre	Selon la structure conceptuelle systématique et avec des références à une liste par ordre alphabétique.	Alphabétique.
Polysémie	La relation entre le concept et le terme est univoque.	La polysémie est acceptée.
Information grammaticale	Indique seulement les différences grammaticales qui sont un rapport avec l'association du concept à un terme donné.	Information morphologique et grammaticale : genre et nombre.
Information syntaxique	Il n'y en a pas.	Inclut des informations phraséologiques de manière plus ou moins systématique.
Usage	Aide à la codification.	Aide à la décodification.

Tableau 1: La terminographie selon l'optique conceptuelle ou l'optique lexicosémantique

[adapté de Cabré (1993, p. 78-92), de Wright et Budin (1997, p. 328) et de Gómez González-Jover (2005, p. 118-120)]

En réalité, les différences entre les produits terminographiques ou lexicographiques élaborés depuis une optique conceptuelle ou depuis une optique lexicosémantique ne sont pas aussi tranchées. Dans la pratique il existe une sorte de mélange de démarches.

En fait, comme le signale Cabré (2005), autant le produit terminographique que celui lexicographique sont conditionnés par quatre facteurs essentiels, à savoir :

- a) le thème ou la portée thématique ;

⁷⁷ Nous faisons référence à des variations liées à la situation géographique, des variations liées au contexte social, des variations liées à la diachronie, des variations liées au mode de communication.

- b) les destinataires ;
- c) les fonctions couvertes par l'ouvrage, c'est-à-dire les attentes des destinataires au moment de le consulter ;
- d) la valeur sociale qu'on veut lui accorder.

En fonction de ces facteurs, il faudra ensuite prendre des décisions sur la macrostructure et microstructure. C'est à ce moment que les dictionnaires à termes vont acquérir leur spécificité par rapport aux dictionnaires du lexique général (Cabré, 2005). En ce qui concerne la macrostructure, les décisions à prendre sont les suivantes (Rey, 1995) :

- a) les sources ;
- b) la sélection des unités ;
- c) la représentation ;
- d) l'ordre.

La méthodologie de travail adoptée dépendra du type de dictionnaire que l'on veut élaborer, de la formation des terminographes ou lexicographes qui le feront et de l'usage qu'ils feront des outils informatiques. Ainsi, un dictionnaire (ou un autre type de produit terminographique) est le résultat de tout un ensemble de facteurs et de décisions.

3.2.1. Les dictionnaires spécialisés

Étant donné que dans notre travail de recherche nous avons suivi notamment l'optique lexicographique, nous allons présenter dans ce qui suit quelques produits concernant les dictionnaires spécialisés.

La terminologie a exercé une importante influence sur la lexicographie spécialisée⁷⁸. Nous avons analysé des dictionnaires des domaines de spécialité recouverts dans la nomenclature de notre dictionnaire. Cela nous a permis de dégager des caractéristiques générales sur certains aspects lexicographiques particulièrement intéressants pour l'élaboration de notre dictionnaire. D'abord, nous présenterons les

⁷⁸ Sur l'apport de la terminologie à la linguistique, voir Depecker (2005).

différences entre les dictionnaires monolingues et les dictionnaires bilingues ou plurilingues : objectif, public visé et contenu de l'article lexicographique ; ainsi que les problèmes spécifiques à la terminologie bilingue ou multilingue. Ensuite, nous relèverons quelques caractéristiques les plus saillantes de la nomenclature de ce type d'ouvrages lexicographiques : la primauté accordée aux substantifs et le traitement des unités polylexicales. Enfin, nous dénoncerons le manque d'attention porté aux relations que les termes entretiennent avec le reste de mots de la phrase (la syntaxe).

3.2.1.1. Dictionnaires monolingues, dictionnaires bilingues et dictionnaires multilingues

Les termes recueillis dans ces dictionnaires font appel à « un stock lexical dont ne dispose pas et où ne puise pas la majorité des locuteurs d'une langue donnée » (Balliu, 2018, p. 10). Ainsi, le public visé des dictionnaires spécialisés monolingues sont les spécialistes d'un domaine donné : des professionnels ou des étudiants. L'ordre de la nomenclature de ces dictionnaires traduit souvent la démarche onomasiologique suivie par les terminographes, c'est-à-dire les termes sont ordonnés en fonction de la structure conceptuelle systématique. On donne priorité aux données de type encyclopédique, comme la définition et les limites sémantiques du terme en question. Il ne s'agit donc pas de vrais dictionnaires, mais plutôt d'ouvrages encyclopédiques sur un domaine de spécialité.

En ce qui concerne la lexicographie spécialisée bilingue et multilingue, comme l'objectif et le public visés sont différents, cela implique des différences aussi dans le contenu de l'article lexicographique. L'objectif est de traduire des termes ; autrement dit ces ouvrages proposent des équivalents terminologiques en une ou plusieurs langues. De ce fait, la définition et d'autres types d'informations à propos du concept auquel le terme en question réfère s'avèrent inutiles. L'article du dictionnaire contient donc moins d'informations que dans les dictionnaires monolingues. Ces ouvrages sont souvent conçus par et pour des traducteurs. Ils se présentent sous forme de glossaires ou des bases de données terminologiques.

Quant à la méthode de travail pour l'élaboration de ce type d'ouvrages, L'Homme (2004) signale :

Si le projet terminographique porte sur plus d'une langue, la mise en forme du corpus, le repérage des termes, la collecte de données sur ces termes et l'analyse des données

terminologiques se font en principe dans chacune des langues séparément. À ces premières tâches succède un travail de fusion des données, c'est-à-dire d'établissement d'équivalences (p. 47).

Ainsi, le travail lexicographique est, dans un premier temps, monolingue. Ensuite, on établit les équivalences dans d'autres langues.

Lors de cette tâche d'établissement des correspondances entre les différentes langues les traducteurs ou terminographes se heurtent souvent à l'un des plus gros problèmes de la terminologie multilingue : les différences de type cognitif et culturel entre les cultures véhiculées par les langues. Comme le signale Lerat (1995), il se peut que les objets soient conceptualisés différemment d'une culture à une autre ; alors, leurs dénominations ne seront pas faciles à traduire⁷⁹. Dans le domaine des sciences sociales, qui est le domaine très général de notre nomenclature spécialisée, cette situation se révèle spécialement récurrente, étant donnée leur forte empreinte culturelle. Par exemple, les types d'allocations parfois ne correspondent pas exactement dans les différents pays, voire il n'y a pas de correspondance d'un type d'allocation donné car elle n'existe pas dans l'un des pays.

Des fois, faute de mieux, les lexicographes résolvent ce problème avec un équivalent approximatif. C'est ce qu'on appelle une *équivalence fonctionnelle* (Pigeon, 1982). Cette solution a été adoptée, par exemple, dans la banque de données terminologique TERMIUM⁸⁰.

⁷⁹ La normalisation que prônait Wüster résoudrait ce type de problèmes, mais elle reste un idéal. Dans certains domaines de spécialité il semble impossible pouvoir établir une normalisation internationale respectueuse des spécificités nationales et culturelles.

⁸⁰ TERMIUM est la banque de données terminologiques et linguistiques du gouvernement du Canada. Sa dernière version, TERMIUM Plus donne accès à des millions de termes en quatre langues (anglais, français, espagnol et portugais) et comprend un grand nombre de domaines de spécialité. Elle comporte des termes, des abréviations, des définitions et des exemples d'utilisation. On peut consulter cette banque de données sur <https://www.btb.termiumplus.gc.ca/> . En Europe, la Commission Européenne a créé l'IATE (Inter-Active Terminology for Europe). Cette base de données terminologique, disponible depuis 2004 (sa toute dernière version date de novembre 2018), est l'héritière de l'ancienne base EURODICAUTOM (créée en 1975). L'objectif de l'IATE est de recueillir et gérer la terminologie spécifique à l'Union Européenne. Il comprend donc les 24 langues officielles européennes, plus le latin. Il s'agit donc de la plus vaste base de données multilingue en terminologie. Elle est disponible sur le site

D'autres fois, pour respecter la spécificité culturelle, la solution adoptée est l'emprunt terminologique ; par exemple, la Commission Européenne conserve les dénominations *cour d'appel* (fr). et *corte d'appello* (it) dans leur langue d'origine pour respecter les différences de nature et de fonctionnement des juridictions française et italienne.

Kalonji (1993) a identifié un autre type de solution à ces problèmes : les *traduisants au sens*, c'est-à-dire « des mots ou des groupes de mots de la langue cible présentés comme des équivalents de la langue d'entrée » (p.177).

Enfin, la solution la moins efficace est de proposer une liste d'équivalents sans plus (Lerat, 1995). L'utilisateur sera donc obligé de chercher les articles consacrés à chacun de ces termes dans des dictionnaires monolingues.

3.2.1.2. La nomenclature des dictionnaires spécialisés

Les dictionnaires de spécialité et les banques de terminologie qui s'alignent sur l'optique d'une perspective conceptuelle accordent une importance démesurée aux substantifs par rapport à d'autres catégories grammaticales. Rappelons que la tradition grammaticale a défini le nom comme « le mot qui sert à désigner, à "nommer" les êtres animés et les choses » (Grevisse, 1980, p. 223), c'est-à-dire qu'il est la partie du discours qui sert à dénommer. De plus, la terminologie développée à partir des études de Wüster part du concept pour ensuite lui associer une forme linguistique. Donc, le nom est la seule catégorie grammaticale vraiment pertinente du point de vue terminologique. Ces deux facteurs pourraient expliquer la prédilection pour les noms en terminographie.

D'autre part, dans les dictionnaires spécialisés on ne précise pas pourquoi certaines unités polylexicales sont recueillies dans la nomenclature mais pas d'autres. Par exemple, en ce qui concerne le terme *congé*, aucun dictionnaire spécialisé ou glossaire terminologique ne coïncide sur la même série de composés de *congé* (*congé d'adoption, congé de conversion, congé de formation, etc.*).

<https://iate.europa.eu/home> . Ces deux bases de données ont été conçues par des documentalistes, plus que par des linguistes. Ainsi, elles sont dépourvues d'approche théorique terminologique ou lexicologique.

3.2.1.3. La syntaxe dans les dictionnaires spécialisés

La terminologie classique ne s'est intéressée qu'aux concepts et à leurs formes linguistiques, pas aux relations que les termes entretiennent avec les autres éléments de la phrase. Selon L'Homme (2004), « dans cette optique [l'optique classique], le terme est envisagé indépendamment du contexte linguistique dans lequel il est utilisé » (p. 37). En effet, dans les dictionnaires terminologiques ou glossaires consultés, le fonctionnement syntacticosémantique des termes est rarement abordé, voire il est ignoré, notamment dans les dictionnaires monolingues et multilingues. Dans les dictionnaires monolingues on suppose que l'utilisateur (un spécialiste) connaît déjà la distribution du terme, tandis que dans les dictionnaires multilingues l'absence d'informations syntaxiques est probablement due à des raisons d'espace et de format de présentation des équivalents en plusieurs langues. En ce qui concerne les dictionnaires bilingues, on trouve parfois des informations sur la combinatoire lexicale du terme en question, mais leur traitement n'est ni exhaustif ni uniforme.

Le manque d'intérêt pour la syntaxe fait que ces dictionnaires s'avèrent peu utiles pour la génération de textes, voire totalement inutiles dans le cas du traitement automatique des langues. Nous sommes donc d'accord avec G. Gross et Mathieu-Colas (2001) lorsqu'ils affirment que « la plupart des lexiques ne sont pas en mesure de rendre compte du fonctionnement d'une langue de spécialité avec des précisions nécessaires à des fins informatiques » (p. 70).

4. Conclusions

Après avoir passé en revue la notion de nom prédicatif et avoir analysé les caractéristiques du discours de spécialité et son traitement lexicographique, deux conclusions s'imposent.

Premièrement, le discours de spécialité peut être abordé sous l'angle de la prédication. En effet, le discours spécialisé est construit de prédicats, d'arguments et d'actualisateurs, tout comme le discours général. Nous avons montré que les termes peuvent avoir une structure argumentale, c'est-à-dire ils peuvent être des prédicats. Ainsi, autant les noms de la langue générale que les noms de la langue de spécialité que

nous avons retenus pour notre dictionnaire peuvent être décrits avec les mêmes outils d'analyse.

Deuxièmement, si le prédicat a une structure argumentale, il s'avère nécessaire de tenir compte de l'aspect syntaxique dans notre description des entrées du dictionnaire. L'étude des discours de spécialité ne peut pas se limiter au lexique, mais il faut étudier toutes les phrases qui sont propres à ces discours. Ainsi, nous pourrions décrire les emplois spécifiques et les usages que l'on fait de la langue générale dans des usages professionnels.

Pour entreprendre notre tâche, sans refuser les apports de la terminologie, nous avons opté pour une démarche lexicographique, qui s'adapte mieux à nos objectifs.

CHAPITRE III : LA PHRASÉOLOGIE

Depuis une trentaine d'années, la phraséologie est l'un des domaines les plus populaires en linguistique. En effet, elle est devenue une dimension fondamentale dans la description des langues. Longtemps délaissée par la tradition grammaticale, la description des unités phraséologiques est devenue l'un des enjeux centraux des théories linguistiques modernes, comme l'approche du lexique-grammaire (M. Gross, 1982 ; G. Gross, 1996a), le modèle Sens-Texte (Mel'čuk, 1973, 1988, 1995, 1997, 2013 ; Kahane, 2003b) ou les Grammaires de Construction (Croft et Cruse, 2004 ; Fillmore et al., 1988). Ce « tournant phraséologique de la linguistique » (comme le qualifient Legallois et Gréa, 2006, p. 5) est dû, d'une part, au développement des outils informatiques et de la linguistique de corpus (cf. Biber et al., 1998/2006 ; Habert et al., 1997), qui ont permis d'analyser des masses de données, et ont mis en évidence l'ampleur du phénomène phraséologique dans le système de la langue, et d'autre part, à l'essor des approches théoriques tendant à privilégier le lexique (Fillmore, 1988 ; Jackendoff, 2002 ; Sinclair 1991). En conséquence, ces dernières décennies la phraséologie a élargi notablement ses objets d'étude, ses approches et ses méthodes (cf. Legallois et Tutin, 2013). Ainsi, la notion d'unité phraséologique s'étend vers des phénomènes linguistiques tels que la collocation ou la colligation⁸¹.

⁸¹ Depuis l'apparition de la notion de *colligation* dans la linguistique contextualiste britannique (Firth, 1951), le domaine colligatif s'est étendu vers une portée textuelle (cf. Francis, 1991 ; Sinclair, 2003 ; Legallois, 2012 ; entre autres). De nos jours, la définition de colligation de Hoey (2005) semble faire autorité :

1. *the grammatical company a word or word sequence keeps (or avoids keeping) either within its own group or at a higher rank;*
2. *the grammatical functions preferred or avoided by the group in which the word or word sequences participates;*
3. *the place in a sequence that a word or word sequence prefers (or avoids).* (p. 43)

La colligation est donc un phénomène de comportement grammatical préférentiel, tandis que la collocation est un phénomène de comportement lexical préférentiel (cf. Legallois et Gréa, 2006). Étant donné que la collocation constitue l'objet de notre recherche, nous la décrivons minutieusement dans la section 4. *Les collocations* de ce chapitre.

Dans ce deuxième chapitre, nous proposons de faire un bref parcours de l'histoire de la phraséologie dans le but de mettre en évidence son évolution. Ensuite nous essayons de cerner l'objet d'étude de cette discipline, dont le domaine n'a cessé de s'élargir au fil des années, et de mettre en relief la présence massive de ce phénomène considéré même comme une propriété définitionnelle des langues naturelles (G. Gross, 1988). Malgré le grand nombre de travaux portant sur la phraséologie et son importante avancée théorique, la définition, délimitation et description de cette discipline est loin de trouver un accord parmi la communauté des linguistes. Néanmoins, il existe un certain nombre de caractéristiques des unités phraséologiques qui sont communément admises. Nous présentons donc certains critères définitoires de ce type de constructions et nous faisons un petit tour d'horizon de quelques classifications appartenant à différentes perspectives théoriques. Finalement, la dernière partie de ce chapitre est consacrée aux collocations, puisqu'elles font l'objet de notre travail de recherche.

1. Bref aperçu historique de la phraséologie

La phraséologie a parcouru un long chemin avant d'être considérée une discipline linguistique. Ignorée des linguistes pendant longtemps et ensuite rattachée aux travaux de lexicologie, de grammaire ou de stylistique, la phraséologie a eu du mal à trouver sa place comme discipline autonome. Selon Ruiz Gurillo (1997) « la fraseología ha sido desde siempre la *tierra de nadie* a la que acudían investigadores de todas las escuelas y signos movidos por el interés que despertaban en ellos las combinaciones fijas de palabras » (p. 17). Par ailleurs, le manque de consensus sur la délimitation de son objet d'étude, sur une classification et sur une terminologie rend très difficile la tâche de retracer l'histoire de cette discipline⁸².

Avant de devenir une discipline scientifique reconnue par les linguistes, la phraséologie se manifeste notamment dans des travaux de lexicologie ou de parémiologie. Dans la plupart des langues, les premières attestations de l'intérêt porté aux unités phraséologiques se trouvent dans des recueils de maximes, de proverbes, etc.

⁸² Pour une étude approfondie de l'histoire de la phraséologie, consulter Zuluaga (1980, chap. 2), Ruiz Gurillo (1997 : 17-32) et Mejri (1997). Pour une révision bibliographique de l'histoire de la phraséologie, voir García-Page (2008).

Au fil des années, ces ouvrages ont intégré toutes sortes d'expressions plus ou moins figées. Depuis le XIX^e siècle, avec l'essor de la lexicographie, ces suites figées intègrent les dictionnaires. Cependant, jusqu'à nos jours, le traitement des unités phraséologiques dans les dictionnaires généraux n'a jamais été ni systématique ni quantitativement significatif. Gómez Fernández et Uzcanga Vivar (2010b) notent que « les expressions idiomatiques sont déjà difficiles à repérer dans les dictionnaires monolingues, car elles sont généralement insérées parmi d'autres exemples qui sont des combinaisons libres, souvent sans aucun type d'étiquette indiquant leur statut différent » (p. 982). Les dictionnaires bilingues, de leur côté, manquent « de critère précis dans l'emplacement de ces expressions dans la microstructure de l'article ; il n'y a pas plus de marque qui permette à l'utilisateur de différencier les combinaisons libres des combinaisons restreintes » (Gómez Fernández et Uzcanga Vivar, 2010b, p. 982). Le traitement que la lexicographie aurait accordé à ces unités ne semblerait pas avoir contribué à mettre de l'ordre dans l'étude de la phraséologie. González Hernández (2016) souligne que le caractère fortement hétérogène des unités phraséologiques « justifica, en parte, que la tradición gramatical no concediera un tratamiento sistemático a dichas unidades, por considerarlas un foco de irregularidades sintácticas y semánticas ; quedando, por eso mismo, su estudio relegado, en muchas ocasiones, a un apéndice constituido por una especie de inventario o lista de secuencias prefabricadas » (p. 156).

Les premières allusions à l'existence de éléments phraséologiques datent de la fin du XIX^e siècle et apparaissent dans des travaux de linguistique générale en différentes langues (Hermann Paul, 1880, pour l'allemand ; Henry Sweet, 1891, pour l'anglais). En ce qui concerne les langues de notre étude, déjà en 1897 Michel Bréal, dans son *Essai de sémantique*, signale l'existence de groupes de mots figés, qu'il appelle *locutions*, *formules* ou *groupes articulés*, et dont il souligne le caractère préconstruit et idiomatique. Pour l'espagnol, il faudra attendre jusqu'en 1950 pour voir publié l'ouvrage de Julio Casares *Introducción a la lexicografía moderna*. Dans cet ouvrage, l'auteur essaie de déterminer et de faire un classement des unités qui constituent l'objet d'étude de la phraséologie.

Saussure (1916) aborde de façon marginale ce phénomène. Il fait référence à des *locutions toutes faites*, qui résultent d'une tendance mécanique des systèmes linguistiques. Ce linguiste s'intéresse au phénomène de l'agglutination, c'est-à-dire le

processus selon lequel deux ou plusieurs mots distincts se trouvent soudés en une seule forme lexicale. Pour sa part, Sechehaye (1921), élève de Saussure, a compris la grande portée de ce phénomène et fait allusion à un ensemble de parties significatives qui seraient considérées comme une seule signification, devenant ainsi l'équivalent à un signe simple. Il s'intéresse aux locutions, dont les composants perdent leur identité sémantique, et aux mots composés, dont les composants la conservent.

C'est un autre disciple de Saussure, le suisse Charles Bally (1909/1951), qui est souvent considéré le père de la phraséologie en tant que discipline linguistique⁸³ (Ruiz Gurillo, 1997 ; González Rey, 2002). En effet, Bally est le premier à utiliser le terme *phraséologie* dans un sens linguistique :

Si, dans un groupe de mots, chaque unité graphique perd une partie de sa signification individuelle ou n'en conserve aucune, si la combinaison de ces éléments se présente seule avec un sens bien net, on peut dire qu'il s'agit d'une locution composée. [...] C'est l'ensemble de ces faits que nous comprenons sous le terme général de phraséologie. (Bally, 1951, p. 65-66)

Nous observons que sa définition de la phraséologie est basée sur la compositionnalité sémantique⁸⁴. Bally est le premier à proposer une classification des phénomènes phraséologiques. Cet auteur distingue trois grands types de groupements de mots :

- les *groupements passagers* ou *associations occasionnelles* (p. ex., *avoir une maison*) ;
- les *groupements usuels* ou *séries phraséologiques* (p. ex., *avoir de la chance, chaleur suffocante, grièvement blessé*) ;
- les *unités indécomposables* ou *unités phraséologiques* (p. ex., *avoir maille à partir*).

⁸³ González Rey (2002) signale que bien que la plupart des linguistes considèrent Bally comme le fondateur de la phraséologie, il n'y a pas d'unanimité à ce sujet. Zuluaga (1980) trouve que cette place correspond à Hermann Paul, tandis que d'autres jugent que ce sont les idées de Saussure qui ont influencé son disciple Bally (Čermak, 1988).

⁸⁴ À part le critère de compositionnalité sémantique, Bally (1909/1951) établit d'autres indices morphologiques et syntaxiques pour déterminer si un syntagme est phraséologique. Cependant, il reconnaît lui-même que ceux-ci sont secondaires et ils ne sont ni suffisants ni déterminants.

Cette classification est basée notamment sur deux axes : le degré de cohésion (momentanée, fréquente ou permanente) et la nature compositionnelle (les éléments qui composent les *associations occasionnelles* et les *séries phraséologiques* sont sémantiquement reconnaissables et ils contribuent à la signification de l'ensemble, tandis que les *unités phraséologiques* sont non compositionnelles⁸⁵).

À partir de l'opposition entre les *unités phraséologiques* et les *associations occasionnelles*, Bally définit une diversité de *séries phraséologiques*⁸⁶ dont les constituants conservent leur autonomie tout en présentant des « affinités » : « Il y a série ou groupement usuel lorsque les éléments du groupe conservent leur autonomie, tout en laissant voir une affinité évidente qui les rapproche, de sorte que l'ensemble présente des contours arrêtés et donne l'impression du "déjà vu" » (Bally, 1951, p. 70).

Même si cette classification a le mérite d'être l'une des premières et la plus détaillée jusqu'à ce jour, la définition des syntagmes intermédiaires reste parfois trop sommaire⁸⁷. La contribution de l'œuvre de Bally dans l'élaboration d'une théorie de la phraséologie est essentielle. Pourtant ses idées ne trouveront d'écho que quarante ans plus tard, dans le cadre de la linguistique russe. En effet, malgré ces intuitions bien fondées de Saussure et de son disciple Bally, la phraséologie ne constitue pas encore l'objet central des théories linguistiques, car « on n'a jamais considéré le figement⁸⁸ comme une dimension fondamentale dans le fonctionnement du système linguistique » (Mejri, 2002, p. 213).

Dans les années 20, Polivanov est le premier à introduire les idées de Bally dans les travaux linguistiques russes. Quelques années plus tard, Vinogradov (1946, 1947)

⁸⁵ Les termes *compositionnel* ou *non-compositionnel* ne sont pas employés par Bally.

⁸⁶ Par exemple, les *séries d'intensité* comme *diamétralement opposé* et *chaleur suffocante* ou les *périphrases verbales* telles que *remporter une victoire* ou *prendre une décision* (Bally, 1909).

⁸⁷ Nous tenons à signaler que les exemples cités par Bally pour illustrer les séries phraséologiques sont encore présentés comme des prototypes de collocations en linguistique actuelle : *gravement malade / grièvement blessé ; désirer ardemment / aimer éperdument*.

⁸⁸ Nous devons comprendre ce terme comme synonyme approximatif de *phraséologie*.

s'inspire des apports de Bally⁸⁹ pour élaborer une classification des phénomènes phraséologiques en trois grands types⁹⁰ :

- les *groupes phraséologiques soudés* sont des unités de la langue sémantiquement immotivées, autrement dit leur sens « total » indivisible est complètement indépendant du sens de leurs composants (p. ex., *donde Cristo perdió el gorro, ser perro viejo*⁹¹);
- les *unités phraséologiques* sont aussi des unités de la langue à un sens « total » indivisible, mais elles sont sémantiquement motivées, car on peut y apercevoir certains indices de compositionnalité sémantique (p. ex., *tener letra de médico, pescar en aguas revueltas*) ;
- les *combinaisons phraséologiques* sont des combinaisons d'unités lexicales sémantiquement divisibles dont les sens sont liées, c'est-à-dire l'un des composants garde son sens, mais l'autre adopte un sens figuré, phraséologiquement contraint (p. ex., *hambre canina, cintura de avispa*).

Vinogradov propose donc une typologie des phénomènes phraséologiques essentiellement fondée sur le degré de cohésion sémantique et sur la notion de motivation sémantique⁹².

Les idées de Vinogradov ont aidé à élargir et mieux définir les limites de la phraséologie et les caractéristiques distinctives de son objet d'étude (les *phraséologismes*, selon le terme instauré par Vinogradov). Vinogradov, Polivanov et Amosova sont considérés comme les véritables pionniers de la constitution de la phraséologie comme discipline scientifique linguistique. Leur empreinte est déterminante dans la genèse de l'école phraséologique russe, dont les idées pénétreront dans l'étude phraséologique en Europe occidentale par le biais de la langue allemande

⁸⁹ Ultérieurement, les idées de Bally seront reprises de manière plus formelle par le linguiste russe Mel'čuk (2003b, 2011) et par le linguiste anglais Cowie (1991).

⁹⁰ Il s'agit d'une typologie des unités phraséologiques en langue russe qui s'est avérée constituer une typologie universelle.

⁹¹ Tous ces exemples en espagnol sont empruntés à Velasco Menéndez (2010, p. 127-131), qui propose des traductions ou des équivalences des exemples en russe donnés par Vinogradov (1946, 1947).

⁹² Sur la notion de motivation chez Vinogradov, voir Velasco Menéndez (2010, p. 128-130).

(Häusermann, 1977) et de l'espagnol (grâce aux cubains Carneado Moré et Tristá Pérez, 1983)⁹³.

Pendant les dernières décennies du XX^e siècle, la recherche en phraséologie se développe autour de trois grands blocs : le bloc oriental soviétique, le structuralisme européen occidental et la linguistique nord-américaine (Corpas Pastor, 1996). Les deux premiers blocs entretiennent des relations très étroites, comme le prouvent les travaux en phraséologie allemande, fortement influencés par les recherches soviétiques (cf. Dobrovol'skij, 1988 ; Häusermann, 1977 ; Mel'čuk, 1973 entre autres). Aux États-Unis, la grammaire générative a longtemps banni la description des constructions phraséologiques. En effet, elle s'est vue confrontée à de graves problèmes lorsqu'il a été question d'appliquer les transformations syntaxiques à ce type d'unités (cf. Chafe, 1968 ; Fraser, 1970 ; Ruwet, 1983 ; Weinreich, 1966/1969). C'est la grammaire transformationnelle de M. Gross (1982, 1988, 1993) qui réussit à intégrer la description des unités phraséologiques.

Ainsi, la prolifération de recherches⁹⁴ et notamment l'organisation de congrès et de séminaires internationaux⁹⁵ à ce sujet favoriseront le développement de la phraséologie dans toutes les langues. L'engouement croissant pour la phraséologie se traduit par l'effervescence de colloques, de publications de monographies et d'ouvrages collectifs consacrés à ce domaine⁹⁶, ainsi que la création de revues entièrement dédiées

⁹³ Pour une analyse des approches russes et leur contribution à la diffusion de la phraséologie, voir Telija (1996, chap. 1), Kótova (1998), Mokienko (2000a) et Dobrovol'skij et Filipenko (2007).

⁹⁴ Pour ne signaler que les débuts de la diffusion de la phraséologie à d'autres langues, nous ferons mention pour l'allemand des linguistes Burger (1973), Fleischer (1982) et B. Wotjak (1989), pour le français, Thun (1978) et Gréciano (1983), pour l'espagnol, G. Wotjak (1983, 1984) et Zuluaga (1980), et pour l'anglais, Weinreich (1966/1969) et Gläser (1986a, 1986b). Pour plus de références bibliographiques, voir González Rey (2021, p. 25-26).

⁹⁵ La première rencontre internationale autour de la phraséologie a lieu à Mannheim (Allemagne) en 1981. Ils s'ensuivront d'autres : Zurich (1984), Oulu (1986), Strasbourg (1988), Aske (1990), Saarbrücken (1992), Graz (1995), Liptovsky Ján (1997), entre autres.

⁹⁶ Legallois et Tutin (2013) en font un éclairant tour d'horizon.

à ce sujet⁹⁷ (Dostie et Tutin, 2019). C'est à partir des années 70 que la phraséologie suscite un grand intérêt et elle devient un véritable objet de recherche en linguistique théorique⁹⁸.

Les études en phraséologie traversent différentes étapes. Dans les années 70, on privilégie les classifications structurelles et fonctionnelles des unités phraséologiques. Ensuite, dans les années 80 et 90 on s'intéresse plutôt sur l'aspect sémantique et pragmatique de ces constructions. Dans les années 90, dans un contexte culturel d'ouverture européenne et de connaissance des langues et des cultures, la phraséologie contrastive se fait une place importante (cf. Gréciano, 1989a, 1989b ; Hausmann, 1991 ; G. Wotjak, 1992 ; Mellado Blanco, 1997).

Au tournant du millénaire, la création de l'association *Europhras* (*European Society of Phraseology*)⁹⁹ en 1999 et ses actions de diffusion scientifique vont favoriser l'échange d'idées et la consolidation du domaine de la phraséologie. Selon Corpas Pastor (2001a), « esta especie de "globalización fraseológica" está destinada a eliminar las barreras de cristal que han separado durante mucho tiempo a las distintas escuelas, algunas de las cuales han estado sumidas en un aislamiento egocéntrico apartado de la realidad » (p. 36). En effet, la phraséologie n'est plus une sous-discipline de la lexicologie, mais elle est devenue une discipline autonome¹⁰⁰. Ainsi, dans les travaux en phraséologie qui s'en sont suivis, on a essayé de définir cette discipline, son objet d'étude, ses critères d'inclusion, etc.

Depuis une trentaine d'années, les études actuelles s'orientent du côté de l'analyse sur de grands corpus. En phraséologie computationnelle, le traitement

⁹⁷ Par exemple, les *Cadernos de Fraseoloxía Galega*, en 2000, ou la revue *Yearbook of Phraseology*, en 2010.

⁹⁸ Voir, entre autres, Mel'čuk (1973, 1998) ; Fónagy (1982) ; Burger (1982, 1998) ; M. Gross (1982, 1988) ; Ruwet (1983) ; Rey (1985) ; Nunberg et al. (1994) ; Everaert et al. (1995) ; Klein et Lamiroy (1995) ; G. Gross (1996a) ; Falia et al. (1997) ; Mejri (1997, 2003a) ; Cowie (1998c) ; Hudson (1998) ; Moon (1998) ; González Rey (2002) ; Grossmann et Tutin (2003) ; Svensson (2004) ; Lamiroy et Klein (2005) ; Blumenthal et Hausmann (2006) ; François et Mejri (2006) ; Burger et al. (2007).

⁹⁹ L'objectif de cette association est de promouvoir l'échange scientifique et la coopération internationale dans le domaine de la phraséologie. Pour plus d'informations : <http://www.europhras.org/fr/europhras>.

¹⁰⁰ Nous développons ce sujet dans la section 2.1. *Discipline autonome ou dépendante*.

automatique de grands corpus a permis d'une part de confirmer et de nuancer les données obtenues auparavant moyennant des études linguistiques qualitatives¹⁰¹, et d'autre part, d'élargir le domaine de la phraséologie. À partir des études de Sinclair (1987, 1991), se développe une approche inductive basée sur le corpus (*corpus-driven*¹⁰²) connue comme *approche distributionnelle* (Evert, 2004, 2008) ou *frequency-based approach* (Nesselhauf, 2004). Cette nouvelle approche redéfinit les limites de ce qui était considéré phraséologique pour y inclure des suites traditionnellement considérées « libres » (Cowie, 1998a). En effet, les analyses de corpus ont révélé l'existence de combinaisons de mots très fréquentes qui ne rentraient pas dans les catégories prédéfinies par la phraséologie (Granger et Paquot, 2008). Néanmoins, tout en reconnaissant l'importance des résultats obtenus, certains auteurs, tels que Gaatone (1997) ou Stefanowitsch (2014), mettent en garde sur les mauvais usages des corpus et le risque de tout considérer comme phraséologique. Finalement, cette approche ne va certainement pas remplacer l'approche classique, qualitative (Kauffer et Keromnes, 2022).

L'imbrication de différents domaines linguistiques tels que la phraséologie, la linguistique de corpus, le traitement automatique des langues naturelles et la

¹⁰¹ En sciences humaines, la recherche qualitative désigne « any type of research that produces findings non arrived at by statistical procedures or other means of quantification » (Strauss et Corbin, 1998, p. 11). Nous utilisons « étude qualitative » ou « approche qualitative » au sens usuel où l'on entend les méthodes qualitatives en linguistique, par opposition à « étude quantitative » ou « approche quantitative » (études reposant sur des données chiffrées). Ainsi, nous entendons par « étude qualitative » l'étude de nature spéculative, reposant sur des jugements faits à partir de l'observation d'exemples (qu'ils soient fabriqués ou issus de corpus). Par contre, on a l'habitude de nommer « étude quantitative » celle qui repose sur des données chiffrées (cf. Kraif et Tutin, 2020).

¹⁰² Tognini-Bonelli (2001) introduit la différence entre l'approche *corpus-driven* et l'approche *corpus-based*. Les études *corpus-based* utilisent les corpus pour explorer une théorie ou une hypothèse, dans le but de la valider, la réfuter ou la redéfinir. Il s'agit donc d'une approche déductive. Selon cet auteur, la linguistique de corpus s'inscrit dans cette approche. En revanche, les recherches *corpus-driven* sont inductives, car on considère que les hypothèses linguistiques émergent du corpus. Ainsi, on prétend que le corpus lui-même incarne une théorie du langage (McEnery et Hardie, 2011). Il s'agit donc d'une approche heuristique.

linguistique computationnelle (ou linguistique informatique) a bouleversé les études en phraséologie¹⁰³, voire la conception même de la linguistique :

la phraséologie [...] est devenue la matière première de certaines conceptions linguistiques : elle propose un renouvellement de l'analyse syntaxique, sémantique, textuelle et psycholinguistique sur la base d'observations plus empiriques, grâce, notamment, au recours aux corpus de grandes dimensions (Legallois et Gréa, 2006, p. 5).

Au cours du XXI^e siècle, la phraséologie a étendu considérablement ses approches, ses méthodes et ses objets d'étude¹⁰⁴. Au-delà des disciplines traditionnelles telles que la lexicologie, la sémantique, la sémiologie ou la syntaxe, d'autres champs des sciences du langage comme la psycholinguistique (Conklin et Schmitt, 2008 ; Ellis et al., 2008 ; Edmonds, 2013b, 2016 ; Wray, 2002, 2008), la linguistique textuelle (Biber, 2009 ; Longrée et al., 2008 ; Longrée et Mellet, 2013 ; Quiniou et al. 2012), ou la linguistique informatique (Hoey, 2005 ; Langacker, 2013 ; Sinclair, 1991) ont contribué à une meilleure compréhension des unités phraséologiques¹⁰⁵. En contrepartie, cette extension du domaine de la phraséologie risque d'entraîner une « dilution du phénomène, une imprécision dans sa définition – ou plutôt de multiples définitions relatives à une diversité de points de vue » (Legallois et Tutin, 2013, p. 19).

Pendant ces dernières décennies, la phraséologie contrastive a connu un important essor, favorisé, d'une part, par la linguistique de corpus (cf. Baroni et al., 2009, 2010 ; Colson, 2003, 2012 ; Corpas Pastor, 2013) et, particulièrement ces dernières années, par le développement de la linguistique cognitive (cf. Dobrovolskij et Piirainen, 2005, 2018). La phraséologie contrastive constitue un domaine de recherche multidisciplinaire qui relève et entretient des liens très étroits avec plusieurs disciplines linguistiques, telles que le lexique, la sémantique, la syntaxe, la morphologie, la pragmatique, la prosodie, la lexicologie, la sémiotique, l'enseignement des langues ou la

¹⁰³ À titre d'exemple, nous citons quelques publications récentes telles que Corpas Pastor et al., (2021), Corpas Pastor et Colson (2020), Corpas Pastor et Mitkov (2019), Gray et Biber (2013), Mitkov (2017), Mitkov et al. (2018) ou les conférences bisannuelles organisées par l'association *Europhras* autour de la phraséologie computationnelle et la phraséologie basée sur le corpus.

¹⁰⁴ Voir Burger et al. (2007) ; Granger et Meunier (2008) ; Legallois et Tutin (2013).

¹⁰⁵ Le numéro 189 de la revue *Langages* (titré *Vers une extension du domaine phraséologique*) témoigne de cet élargissement.

traduction (théorique et pratique)¹⁰⁶. Elle offre donc un grand nombre de points de comparaison entre les différents systèmes linguistiques. La recherche en phraséologie contrastive a dépassé la simple constitution de listes d'équivalences d'unités phraséologiques, pour se centrer plutôt sur ce qui peut être universel dans les différentes langues et ce qui est idiomatique¹⁰⁷ (cf. Soutet et al., 2018). Cette universalité se présente sur le plan grammatical, sémantique ou culturel¹⁰⁸. En Europe, la recherche phraséologique s'est fortement intéressée sur une sorte d'uniformité de la phraséologie dans les langues européennes, basée sur l'héritage culturel commun¹⁰⁹. Ainsi, les linguistes européens ont développé de vastes projets tels que *Widespread Idioms in Europe and Beyond* (Pirainen, 2008, 2012, 2016) sur les expressions idiomatiques les plus répandues dans une soixantaine de langues européennes et une vingtaine de langues non européennes, ou, dans le cadre théorique des grammaires de construction, le projet *Frame (FRaseologia Multilingue Elettronica)*¹¹⁰ (Cotta Ramusino et al., 2014 ; Schafroth, 2019) propose une base de données de phraséologie multilingue incluant six langues européennes et une non européenne.

L'éclosion d'une « linguistique de la phraséologie » (Cavalla et Legallois, 2020, p. 25) et l'extension de son domaine d'étude a eu une portée remarquable sur la recherche en enseignement et apprentissage des langues. À l'aube de la naissance de la

¹⁰⁶ La relation entre la phraséologie et la traduction est bidirectionnelle. D'une part, le problème que pose la traduction des unités phraséologiques (notamment la traduction des collocations) exerce une certaine influence sur la théorie et la pratique de la traduction, parce qu'il est nécessaire de trouver des stratégies pour résoudre les problèmes de traduction ou de correspondance (Cotta Ramusino et Mollica, 2020, p. 5). D'autre part, la traduction a toujours contribué au repérage des unités phraséologiques (Colson, 2008, p. 196).

¹⁰⁷ En l'occurrence, nous comprenons *idiomatique* par ce qui est spécifique à une langue et qui ne peut pas être transféré tel quel à une autre langue.

¹⁰⁸ Ce dernier aspect, longtemps banni de la recherche linguistique, est l'un des plus récurrents dernièrement.

¹⁰⁹ Pour une révision des différentes étapes connues par la recherche en phraséologie en Europe à travers l'association *Europhras*, voir G. Gréciano (2018).

¹¹⁰ La ressource électronique est disponible en ligne sur <http://www.fraseologia.it/>.

phraséologie en tant que discipline, Bally (1909/1951) avait déjà attiré l'attention sur l'importance du bon maniement de phraséologie :

L'étude des séries, et en général de tous les groupements phraséologiques, est très importante pour l'intelligence d'une langue étrangère. Inversement, l'emploi de séries incorrectes est un indice auquel on reconnaît qu'un étranger est peu avancé dans le maniement de la langue ou qu'il l'a apprise mécaniquement. (p. 73)

Ultérieurement, de nombreux linguistes ont souligné l'intérêt, voire la nécessité, de l'enseignement de la phraséologie. Nous citons, par exemple, Mel'čuk (1993) :

Un natif parle en phrasèmes. Si ce postulat crucial est accepté, et nous l'acceptons, il apparaît alors clairement que l'apprentissage systématique des phrasèmes est indispensable dans l'enseignement d'une langue, que ce soit la langue maternelle de l'apprenant ou une langue étrangère, et indépendamment de l'âge ou du niveau d'éducation de l'apprenant. (p. 83-84)

Dans la même veine, Gréciano (2000c) soutient que « la Phras [= phraséologie] est fondamentale pour l'enseignement et la pratique des langues étrangères » (p. 91). Dans une perspective pragmatique, cette linguiste s'intéresse spécialement aux fonctions des unités phraséologiques pour assurer la cohérence textuelle (cf. Gréciano, 1984, 2000a, 2000c).

Pourtant, les unités phraséologiques, notamment les collocations¹¹¹, n'ont trouvé leur place dans l'enseignement que très tardivement (cf. Edmonds, 2013a, p. 132). Ce n'est qu'au cours de ce XXI^e siècle qu'elle connaît un moment de développement très remarquable¹¹². Gómez Fernández (2021a) regroupe les propositions didactiques existantes en trois grands types : « Dans le premier type, on sensibilise l'apprenant à l'existence des unités phraséologiques sans un enseignement explicite de ces unités (Kennedy, 2003) ; dans le second, on les enseigne en tant que blocs lexicalisés à mémoriser (Lewis, 2000) ; et dans le troisième, on enseigne des régularités qui sont la

¹¹¹ De nombreux linguistes s'intéressant à l'enseignement des langues ont signalé que la maîtrise des collocations constitue l'un des plus grands défis des apprenants d'une langue étrangère, mais celle-ci s'avère absolument nécessaire pour l'apprentissage d'une langue (cf. Durrant et Schmitt, 2009 ; Granger, 1998 ; Hausmann, 1984 ; Howart, 1998b ; Lewis, 2000 ; Wanner et al., 2013).

¹¹² González Rey (2012) attire l'attention sur l'apparition des premières études théoriques et pratiques précurseurs sur l'enseignement-apprentissage de la phraséologie dans les années 90 : Colson (1992) ; Ettinger (1992) ; Howart (1998a) ; Mel'čuk (1993). Pour un survol de l'histoire de la didactique de la phraséologie, voir González Rey (2012).

base pour pouvoir formuler des séquences appropriées (Liu, 2010). » (p. 66-67). L'intérêt pour la recherche en didactique de la phraséologie a donné lieu à de nouvelles approches qui abordent cette question dans des perspectives qui vont au-delà de l'aspect strictement linguistique, telles que les dimensions psycholinguistique, ethno-linguistique ou socioculturelle des unités phraséologiques.

À la croisée entre la phraséologie, la glotodidactique, la linguistique contrastive, la psycholinguistique, la neurolinguistique et la sociolinguistique se trouve la *phraséodidactique*. Selon González Rey (2012) « la fraseodidáctica consiste, ciertamente, en la didáctica de la fraseología de una lengua, pero también en la didáctica de una lengua a través de su fraseología » (p. 76). Ce courant didactique étudie les mécanismes d'acquisition des unités phraséologiques en langue maternelle et surtout les processus d'enseignement-apprentissage de ces constructions en langue étrangère (Sułkowska, 2016). Dans cette perspective on prône une approche actionnelle qui comprenne tous les aspects de la compétence communicative (linguistique, sociolinguistique et pragmatique)¹¹³. De nos jours la phraséodidactique fait l'objet de nombreux ouvrages monographiques (cf. González Rey, 2007 ; Sułkowska, 2013), d'ouvrages collectifs (cf. Henrot Sostero et González Rey, 2019 ; González Rey, 2014 ; Lorenz-Bourjot et Lüger, 2001), de méthodes d'enseignement comme *PHRASÉOTEXT – Le français idiomatique* (González Rey, 2016)¹¹⁴ ou de congrès exclusivement consacrés à la phraséodidactique (l'un des premiers étant *Repenser le figement : enjeux et perspectives en phraséodidactique des langues*, tenu à l'Université Sorbonne Nouvelle – Paris, en 2016).

Au début du millénaire l'approche onomasiologique à l'enseignement de la phraséologie a vu le jour (Cavalla, 2009 ; Cavalla et Crozier, 2005 ; Drouin, 2007 ; Pecman, 2007 ; Tutin, 2007). Dans cette perspective, le point de départ de l'apprenant n'est plus un ensemble de mots mais un ensemble de concepts (Picoche, 1992). Ainsi, la question qui guidera l'apprentissage sera : « De quel lexique a-t-on besoin pour exprimer ce concept ? » (Cavalla, 2009, p. 8). L'entrée par le concept permettrait à

¹¹³ Pour un bref aperçu des recherches en phraséodidactique, voir Ettinger (2008) et Sułkowska (2016).

¹¹⁴ Cette méthode s'inscrit dans le courant de la linguistique cognitive, appliquant les principes de la grammaire de constructions.

l'apprenant d'établir des liens entre le niveau lexical et les concepts sous-jacents, lui dotant ainsi d'une autonomie fonctionnelle (Frenck-Mestre et al., 1999, p. 61). Cette approche viendrait donc compléter l'approche sémasiologique¹¹⁵ en ajoutant le côté culturel et contextuel dont cette dernière manque parfois.

Dans une perspective psycholinguistique, certains spécialistes partent de l'idée que le fond langagier de l'apprenant n'est pas un simple dépôt. Il s'agit plutôt d'un, espace dynamique où les unités lexicales sont connectées en réseau grâce à des liens. Par conséquent, pour apprendre une langue il ne suffit pas d'apprendre des mots isolés, mais des mots employés sur l'axe syntagmatique, c'est-à-dire en combinaison avec d'autres unités lexicales. En ce qui concerne l'apprentissage-enseignement de la phraséologie d'une langue, ils trouvent que la fréquence est un facteur essentiel pour son acquisition aussi bien en langue maternelle qu'en langue étrangère (Cavalla et Legallois, 2020). En effet,

Forsberg (2008) a pu montrer que l'appropriation des SP [séquences préfabriquées] lexicales (comme *poser une question* ou *se mettre d'accord*) suppose une exposition fréquente et prolongée à la langue cible, alors que les SP discursives (p. ex. *parce que* et *il y a*) sont acquis plus facilement. (Falkert, 2013, p. 158)

Pour un apprenant d'une langue étrangère l'exposition quotidienne à la langue étudiée (et aux éléments phraséologiques de cette langue) peut ne pas être suffisante. On propose donc des méthodes didactiques basées sur l'exposition (de l'unité phraséologique dans son contexte), le repérage (de ces constructions) et la production (moyennant des exercices à contrainte lexicale, de reformulation ou de réflexion discursive¹¹⁶) (Cavalla et Legallois, 2020). Dans cette approche on va donc privilégier

¹¹⁵ L'approche sémasiologique procède plutôt par listes d'expressions regroupées par thématiques lexicales (p. ex., les expressions contenant des noms d'animaux) ne proposant pas d'informations sur l'utilisation discursive de ces éléments. Dans le cadre de cette approche, l'ouvrage *Les expressions imagées* de Galisson (1984b) a constitué une publication importante pour l'enseignement des unités phraséologique.

¹¹⁶ Voir Yan et al. (2018).

l'utilisation des corpus en classe de langue¹¹⁷, ce qui permettra à l'apprenant de construire ses propres savoirs de façon inductive. Enfin, l'utilisation des corpus peut se faire de deux manières : soit on exploite le corpus pour enseigner (Pecman et Klüber, 2011), soit on apprend à exploiter les corpus (Cavalla, 2018 ; Cavalla et Loiseau, 2014).

Par ailleurs, l'exploitation de corpus d'apprenants (parfois comparés à des corpus de locuteurs natifs) a permis de relever, voire de systématiser, les erreurs concernant l'utilisation des collocations (cf. Altenberg et Granger, 2001 ; Dagneaux et al. 1998 ; Futagi et al. 2008 ; Martelli, 2007 ; Nesselhauf, 2005 ; Orol et Alonso Ramos, 2013 ; Tono, 2003). Ces données objectives sur les difficultés des apprenants sont à la base de diverses propositions pédagogiques et outils informatiques visant l'acquisition de ces unités phraséologiques en L2 (cf. Granger, 2003 ; Bolly, 2011 ; Wanner et al. 2013).

L'extension des centres d'intérêt de la didactique de la phraséologie a favorisé sa rencontre avec l'intercompréhension. Ainsi, de nouvelles approches prônent une vision d'une phraséologie plurielle, dans le cadre d'une analyse contrastive, qui s'étendrait vers toutes les langues d'une même famille linguistique (cf. Gómez Fernández, 2021a), voire au-delà d'une seule famille (Kubeková, 2018 ; Mešková et Kubeková 2019). Les linguistes Uzcanga Vivar et Gómez Fernández proposent une approche à la didactique de la phraséologie en langues romanes basée sur deux piliers : l'inférence des collocations et l'exploitation du caractère universel des types de texte (cf. Gómez Fernández, 2011a, 2011b ; Gómez Fernández et Uzcanga Vivar, 2008 ; Uzcanga Vivar et Gómez Fernández, 2011). Leur point de départ est de « [considérer] les phrasèmes en tant que balises qui guident dans le parcours de la compréhension du texte » (Gómez Fernández, 2021a, p. 72). Autrement dit, le phrasème est le garant de la progression textuelle (Gómez Fernández, 2020). En enseignement-apprentissage des langues étrangères dans une approche intercompréhensive, il serait intéressant d'apprendre l'apprenant à inférer le collocatif si la base est transparente¹¹⁸ :

¹¹⁷ Cette approche didactique s'inscrit dans le cadre du *data-driven learning* (DDL), qui permet aux apprenants d'avoir un accès direct aux données linguistiques, en le plaçant ainsi au cœur de l'apprentissage (cf. Johns, 1991).

¹¹⁸ Pour les notions de *collocation*, *collocatif* et *base* de la collocation, voir la section 4. *Les collocations* de ce chapitre.

Ce parcours est très opératoire pour le domaine de la combinatoire lexicale restreinte, par exemple les collocations. En effet, si l'apprenant décode par transparence la base de la collocation, il parvient facilement à inférer le collocatif par transfert. (Gómez Fernández et Uzcanga, 2008, p. 4)

Un exemple d'application de cette approche de la compréhension et de l'enseignement phraséologique à partir des textes plurilingues est le projet Eurom.Com.Text (cf. Gómez Fernández et Uzcanga Vivar, 2010a), dans lequel l'apprenant travaille en cinq langues¹¹⁹.

Avant d'entamer une présentation succincte des études de phraséologie française et de phraséologie espagnole, nous tenterons une brève incursion dans la linguistique anglo-américaine, une « école » traditionnellement isolée. En effet, en Europe continentale, la recherche en phraséologie s'est orientée vers l'identification des critères linguistiques permettant de distinguer les différents types d'unités phraséologiques (basés notamment sur les notions de figement et d'idiomaticité). Par contre, dans les études en langue anglaise, ayant suivi une approche inductive *corpus-driven*, le phénomène phraséologique se définit par la fréquence de coapparition. Même si ces linguistes se tiennent un peu à l'écart des courants phraséologiques européens (la phraséologie russe, slave¹²⁰, allemande et française), les recherches en langue anglaise auront une importante influence sur les études phraséologiques en Europe.

Les études en langue anglaise se regroupent en deux pôles : les contextualistes britanniques et les cognitivistes américains. Ces deux grands courants ont une vision phraséologique de la langue. Malgré les points communs qu'ils présentent, les travaux britanniques et américains se sont longtemps ignorés. Les contributions de ces deux pôles vont mener à l'élaboration de deux dictionnaires combinatoires en anglais exerçant une énorme influence dans les études de phraséographie dans d'autres langues : le *Oxford Dictionary of Current Idiomatic English* (Cowie et al., 1983) et *The BBI Combinatory Dictionary of English* (Benson et al., 1986).

¹¹⁹ Le projet a démarré en quatre langues (portugais, italien, catalan et français) et il a incorporé ces dernières années au roumain.

¹²⁰ Pour une étude approfondie de la phraséologie slave, voir Mokienko (2000b).

Fondée sur l'analyse de corpus et dans une perspective lexicologique, l'école contextualiste britannique (cf. Firth, 1957b ; Hallyday, 1966 ; McIntoch, 1966 ; Sinclair, 1991)¹²¹ s'intéresse aux phénomènes phraséologiques étudiés du point de vue de la fréquence de coapparition des mots. Ils privilégient par principe une méthode a-théorique, c'est-à-dire pour ces linguistes, les seules données véritables sont celles fournies par les grands corpus. Cette approche donnera donc naissance à la linguistique statistique de corpus (cf. Sinclair, 1966, 1991, 2004). Pour Sinclair, l'analyse basée sur des corpus permet de se libérer de tout a priori théorique sur la langue¹²².

Ces travaux fortement appuyés sur de vastes corpus ont mis en évidence la dimension phraséologique de la grammaire et du lexique. Les contextualistes considèrent que la phraséologie constitue un principe organisateur du discours et de la langue. Pour Sinclair, le discours est régi principalement par deux principes qui opèrent en parallèle, mais qui sont contradictoires :

- le principe de libre choix (*open-choice principle*) : « this is a way of seeing language as the result of a very large number of complex choices. At each point where a unit is completed (a word or a phrase or a clause), a large range of choice opens up and the only restraint is grammaticalness... Virtually all grammars are constructed on the open-choice principle » (Sinclair, 1991, p. 109 - 110) ;
- le principe d'idiomaticité (*idiom principle*) : « The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they light appear to be analysable into segments » (Sinclair, 1991, p. 110).

¹²¹ Nous tenons à signaler que cette approche linguistique n'a pas vraiment fait école. Les linguistes que nous citons partagent un point de vue relativement homogène sur la méthode de description linguistique, mais il ne s'agit pas d'une théorie unifiée. Pour une présentation détaillée et documentée sur le contextualisme linguistique britannique, voir G. Williams (2003). Pour une présentation plus succincte mais introduisant la novatrice théorie du *lexical priming* de Hoey (2005) et une remarquable comparaison entre le contextualisme et les grammaires de construction nord-américaines, voir Legallois (2006)

¹²² D'ailleurs, il refuse aussi l'étiquetage catégoriel des corpus et met en question l'utilité de la lemmatisation. Son outil privilégié est le concordancier. Dans le Chap. V nous présentons quelques outils d'étiquetage de corpus.

Alors que le premier principe met en lumière l'idée d'une combinatoire libre, il existe dans les langues une tendance des unités lexicales à coapparaître dans le discours (le principe d'idiomaticité). Selon les propres termes de Sinclair (1991) : « the choice of one word affects the choice of others in its vicinity » (p. 173). Ainsi, le locuteur sélectionnerait des pans de langue préconstruits qui intègrent le lexique et la grammaire. Le second des principes étant prioritaire sur le premier, pour Sinclair la phraséologie n'est plus seulement une discipline linguistique ou une catégorie de la langue, mais un principe qui régit le fonctionnement du discours. En effet, la linguistique de corpus a permis de révéler que le discours (son organisation et le rapport du locuteur au discours) présente des caractéristiques inhérentes aux unités phraséologiques telles que l'habitude, la routine, la globalité ou l'usage. Dans cette même veine, s'appuyant sur les observations statistiques, ce linguiste étaye une thèse sur l'importance du cotexte dans la construction du sens : « choices at word rank co-ordinate with other choices round about in an intricate fashion, suggesting a hierarchy of units of different sizes sharing the realization of meaning » (Sinclair, 2004, p. 140)¹²³. Ainsi, la signification lexicale ne réside pas dans le mot ; elle se distribue sur les cooccurrents du mot.

Toujours dans une perspective lexicologique et dans la continuité des apports de Sinclair, les études inscrites dans le cadre de la théorie des *patterns* cherchent à identifier et à analyser les patrons distributionnels des mots. Les travaux sur de grands corpus menés par Hunston et Francis (2000) ont permis d'identifier des *patterns*. Un *pattern* est un patron distributionnel qu'intègrent certaines classes de mots sémantiquement homogènes¹²⁴. Par exemple, le patron *v-link ADJ about n*¹²⁵ (*Everything is excellent about this golf course*) comprend les adjectifs *adult*, *beastly*,

¹²³ Nous ne développons pas ici les quatre niveaux de cotexte dans la construction du sens autour d'un *core* (un mot-noyau ou base) : *collocation*, *colligation*, *semantic preference* et *semantic prosody* (Sinclair, 2004, p. 140-148). Nous invitons le lecteur à consulter également les analyses critiques de ces notions faites par Legallois (2006, p. 29-33) et par Veniard (2012, p. 68-70).

¹²⁴ Nous tenons à signaler que le patron ne correspond pas nécessairement à un syntagme.

¹²⁵ *V-link* symbolise un verbe d'état, et *n*, un groupe nominal. Nous tenons à signaler que les *patterns* sont caractérisés à partir des parties du discours et non à partir des fonctions syntaxiques. Ces linguistes considèrent qu'une caractérisation en fonctions n'apporte rien de significatif et pourrait compliquer la méthode d'analyse. Il est important de souligner également que la dimension du patron est indépendante de la notion de syntagme.

brave, brilliant, cool, excellent, fine, foolish, funny, good, gracious, great, heavy, lovely, marvellous, mature, nice, odd, ok / okay, reasonable et *sweet* (Hunston et Francis, 2000, p. 87). Il semble que ces adjectifs n'aient pas de propriétés sémantiques communes, mais employés dans le patron indiqué ci-dessus, ils dénotent tous que quelqu'un réagit d'une certaine façon à une situation. Ce sens repose sur la structure entière dans laquelle ces adjectifs sont employés¹²⁶. Chaque patron est associé à un sens pragmatique défini. Ces patrons constitueraient des sortes d'unités phraséologiques schématiques. Nous observons donc que le principe d'idiomaticité de Sinclair (1991) est fondamental pour la théorie des *patterns*. Pourtant, pour Sinclair il est mesurable statistiquement puisque l'analyse porte sur le lexème, tandis que ce principe reste diffus lorsque le point de vue de l'analyse porte sur des patrons.

Outre-Atlantique, les générativistes américains font des efforts pour intégrer l'étude de l'aspect sémantique et de la structure des unités phraséologiques dans la méthodologie de la grammaire générative. Comme le signale Ruiz Gurillo (1997), « el modelo generativo no es apropiado, en principio, para dar cuenta de forma adecuada de su estructura, ya que su comportamiento diverge del de las combinaciones libres » (p. 27). Ainsi, certains générativistes adoptent une approche sémantique interprétative pour essayer de surpasser le modèle générativiste chomskyen, qui avait présenté des failles dans l'étude de la phraséologie (cf. Katz et Postal, 1963 ; et les travaux déjà mentionnés de Chafe, 1968 ; Fraser, 1970 ; Weinreich, 1966/1969)¹²⁷.

Dans les années 80, un nouveau courant naît de la critique du générativisme (notamment sur le rapport entre syntaxe et sémantique) : la linguistique cognitive (cf. Lakoff, 1987 ; Langacker, 1987 ; Talmy, 1983, 1988). Ces linguistes renégocient la coupure entre syntaxe et sémantique. Leur thèse fondamentale est de considérer que la grammaire est symbolique par nature, c'est-à-dire qu'elle constitue une paire entre la structure sémantique et la structure phonologique. La grammaire est donc aussi symbolique que le lexique. L'étude des phénomènes phraséologiques dans le cadre de la

¹²⁶ L'idée qu'une structure syntaxique est dotée de signification est déjà présente, de manière plus ou moins explicite, dans la grammaire transformationnelle et dans le modèle du lexique-grammaire de M. Gross (1975). Néanmoins, il y a des différences de fond, puisque la théorie des *patterns* n'est dérivationnelle, ni transformationnelle.

¹²⁷ Pour des références bibliographiques plus étendues, nous renvoyons à Ruiz Gurillo (1997, p. 27-28).

linguistique cognitive a mis en évidence l'existence de *constructions* possédant des propriétés grammaticales, sémantiques et pragmatiques qui ne sont pas prédictibles à partir de leurs composants (cf. Fillmore et al., 1988 ; Goldberg, 1995, 1998 ; Kay et Fillmore, 1999)¹²⁸. Étant donné que ces constructions ne sont pas sémantiquement prédictibles, le sens qu'elles véhiculent est en partie indépendant du lexique qu'elles intègrent. Les cadres syntaxiques seraient donc dotés de signification. C'est la genèse des grammaires de construction¹²⁹. Dans ce cadre, on va, par exemple, calculer les verbes les plus fréquemment attirés par une structure argumentale donnée (cf. Stefanowitsch et Gries, 2003), ou les cadres prédicatifs des verbes de production de parole qui présentent un même prototype sémantique (cf. François et Sénéchal, 2006). Ces constructions possèdent un degré d'enracinement (*entrenchment* en anglais) plus ou moins fort dans le savoir grammatical des locuteurs et dans le discours¹³⁰. L'*entrenchment* présentant des degrés, il s'avère impossible de délimiter de manière précise les unités linguistiques d'une langue. En effet, les grammaires de construction élargissent le domaine de la phraséologie vers un grand nombre de syntagmes et de structures argumentales idiosyncrasiques qui n'avaient pas encore été repérés ni décrits. Par ailleurs, dans cette optique, les phénomènes phraséologiques font partie d'un continuum entre le lexique et la syntaxe : « lexicon and grammar glide into one another

¹²⁸ Pour une comparaison succincte entre la notion de *pattern* (Hunston et Francis) et celle de *construction* (Goldberg), voir Lagallois (2006, p. 36-42).

¹²⁹ Nous préférons le pluriel au singulier dans la mesure où il ne s'agit pas d'une théorie unifiée mais plutôt d'une « famille » de théories développées au sein de la linguistique cognitive. D'ailleurs, il existe plusieurs conceptions de la notion de construction. À titre d'exemple, nous reprenons la définition de Goldberg (1998) : « a construction is [...] pairing of form with meaning/use such that some aspect of the form or some aspect of the meaning/use is not strictly predicable from the component parts or from other constructions already established to exist in the language » (p. 205).

¹³⁰ L'*entrenchment* met en relation les notions de fréquence d'occurrence et d'unité linguistique. Comme l'expliquent Lagallois et Gréa (2006) « plus la fréquence d'occurrence d'une structure est grande, plus son degré *entrenchment* est grand, et plus elle est perçue comme unité à part entière » (p. 19). Nous tenons aussi à signaler que dans ce cadre, une unité linguistique est « a structure that a speaker has mastered thoroughly, to the extent that he can employ in largely automatic fashion, without having to focus his attention specifically on its individual parts or their arrangement » (Langacker, 1987, p. 59-60). La notion d'*entrenchment* est très proche de celles de figement et de lexicalisation. Pour approfondir sur la relation entre ces trois notions, voir Lagallois et Gréa (2006, p. 18).

so that any specific line of demarcation would be arbitrary » (Langacker, 2000, p. 33)¹³¹. Ainsi, la frontière entre structure syntaxique et structure phraséologique devient floue.

1.1. La phraséologie française¹³²

Dans la première moitié du XX^e siècle des réflexions sur les combinaisons stables de mots s'amorcent. Nous avons signalé plus haut les premières références de la part de linguistes tels que Bréal (1897/1976), Bally (1909/1951)¹³³, Saussure (1916/1978) ou Sechehaye (1921). À cette époque correspond aussi le travail du linguiste hollandais de Boer (1922), qui définit la différence entre la syntaxe figée et ce qu'il nomme la syntaxe vivante ou mobile.

Comme le signale González Rey (2021, p. 26), « c'est en russe qu'on trouve le plus grand nombre d'études menées sur la phraséologie française »¹³⁴. Par ailleurs, il existe un important essor de contributions autour de la phraséologie française provenant de linguistes d'origines diverses, tels que le roumain Coseriu (1966), le russe Mel'čuk (1981), les hongrois Fónagy (1982, 1997) et Bárdosi (1989) et l'allemand Hausmann (1979), entre autres.

¹³¹ Pour approfondir sur l'articulation des expressions phraséologiques dans le continuum entre lexique et syntaxe, voir Croft et Cruse (2004).

¹³² La littérature sur la phraséologie française est énorme, ce qui nous oblige à en rendre compte ici que de manière superficielle. Pour un historique de la phraséologie française, voir Bárdosi (1990a, 1990b) et Mejri (2007, 2018b).

¹³³ Legallois et Tutin (2013) signalent que les idées de Bally sont peu présentes dans la tradition phraséologique française. En effet, les linguistes français, influencés par le structuralisme et le générativisme, adoptent une perspective plus formelle de la langue et définiront les phénomènes phraséologiques notamment dans leur rapport au figement. Néanmoins, depuis quelques années les *collocations* (dénomination moderne des *séries phraséologiques* de Bally) font l'objet de nombreuses recherches en phraséologie.

¹³⁴ L'autrice fait allusion à des thèses de doctorat citées par Thun (1975, p. 53, note I), introuvables hors de l'URSS, à Bárdosi (1989, p. 7), qui cite Nazarian (1976), et à Kótova (1998, p. 253-254), qui mentionne aussi des travaux de Nazarian (1965, 1976, 1978, 1981), Gak et Jodos (1981) et de Gak et Jodos (1981) (González Rey 2021, p. 26).

L'extension en Europe occidentale des recherches russes en phraséologie est passée par les germanistes. Ainsi, Gréciano, linguiste française spécialiste de la langue allemande, a été une figure très importante pour les études en phraséologie en France. Son approche contrastive lui a permis d'observer l'universalité du phénomène phraséologique, tout en s'opposant à la tendance de considérer les séquences figées comme des exceptions ou des irrégularités d'un système linguistique donné, c'est-à-dire comme des unités idiosyncratiques (Gréciano, 1991, 1999). Dans ses travaux, le contenu et la fonction des unités phraséologiques utilisées dans le discours l'emportent sur leur forme et leur structure (Gréciano, 1999). Cette linguiste a donc contribué à l'extension de la phraséologie vers la pragmatique. Elle s'est intéressée à l'usage que le locuteur fait des unités phraséologiques, car c'est le locuteur qui, en fonction du but communicatif poursuivi, des informations à transmettre et des conditions discursives, fait le choix entre des combinaisons libres et des unités phraséologiques. Ainsi, cette linguiste a proposé d'appréhender l'étude des unités phraséologiques « depuis la globalité du texte, sa structure, sa thématique et son intentionnalité » (Gréciano, 2000c, p. 102).

Les travaux de Greimas (1960, 1970) sur les locutions, les proverbes et les dictons, et les deux publications de la collection « Que sais-je ? » de la part de Guiraud en 1961 (*Les locutions françaises*) et, en 1967, de Pineaux (*Proverbes et dictons français*) ont fortement contribué au développement théorique de ce domaine en français. Du point de vue structuraliste et transformationnaliste, les travaux de M. Gross (1971, 1975, 1982, 1988) et G. Gross (1988, 1996a, 1997) ont visé l'obtention d'un lexique-grammaire cohérent du français (et ultérieurement d'autres langues). M. Gross a prôné l'intégration de la description des unités phraséologiques dans la description de la syntaxe générale, car « elles ne sont exceptionnelles ni par leur syntaxe, ni par rapport au lexique » (M. Gross, 1982, p. 152), mais il souligne également, la nécessité d'élaborer des tests spécifiques à ce genre d'unités. G. Gross a défini ces tests, a établi la première liste de critères pour reconnaître les unités phraséologiques et a introduit la notion de *degrés de figement* : « Notre démarche consiste à montrer que le figement n'est pas une valeur absolue mais relève d'une gradation correspondant à des propriétés transformationnelles potentielles réalisées à des degrés différents » (G. Gross, 1988, p. 63). Ultérieurement, Mejri (1997) a réalisé l'un des travaux les plus complets sur le

figement du français ; un sujet qui occupera le centre de ses recherches au cours du XXI^e siècle.

Pour sa part, Ruwet (1983) s'inspire des générativistes américains, tout en refusant les thèses chomskyennes relatives à la phraséologie. Il a abouti à la conclusion que les contraintes associées au fonctionnement des unités phraséologiques ne sont pas spécifiques à ce type de constructions (cf. Mejri, 2005).

La phraséographie connaît un important essor dans les dernières décennies du XX^e siècle. Les recherches des lexicographes Rey et Galisson donneront lieu à l'élaboration de dictionnaires de phraséologie (Rey et Chantreau, 1990 ; Galisson, 1984a). Dans le cadre de la théorie Sens-Texte, les dictionnaires de combinatoire (Mel'čuk et al., 1984, 1988, 1992, 1999 ; Mel'čuk et Polguère, 2007) font partie des ouvrages lexicographiques théoriques les plus remarquables dans ce domaine. Toujours dans une perspective combinatoire, les dictionnaires de Grobelak (1990), Beauchesne (2001), Zinglé et Brobeck-Zinglé (2003) et Le Fur (2007) constituent aussi de très importants ouvrages. Une autre tendance en dictionnaires phraséologiques est le recensement d'expressions, par exemple dans Duneton et Claval (1990), Henry (1992) et Lafleur (1991).

À part la parution d'ouvrages lexicographiques, l'intérêt pour la phraséologie se manifeste également dans la publication de numéros de revues de linguistique entièrement consacrés à cette question, comme *Cahiers de Lexicologie*, n° 82 (2003), *Syntaxe et sémantique* (2004), *Revue de linguistique et de didactique des langues*, n° 53 (2016), *Études de linguistique appliquée*, n° 186 (2017), *Le français moderne* (2018-1), *Langages*, n° 225 (2022), etc.

1.2. La phraséologie espagnole¹³⁵

En Espagne, l'incorporation de la phraséologie à la linguistique est assez tardive. Même si déjà en 1950 Casares élabore, pour la première fois en espagnol, une classification des unités phraséologiques d'un point de vue linguistique (il suit des

¹³⁵ Pour un parcours plus complet de l'histoire de la phraséologie espagnole, voir Ruiz Gurillo (1997, 2002), Corpas Pastor (2003) et Penadés Martínez (2005b). Pour une révision bibliographique, consulter García-Page (2008).

critères grammaticaux et sémantiques), ses idées tomberont dans l'oubli pendant des décennies et ce n'est que dans les années 80 que l'on trouvera des travaux remarquables en phraséologie espagnole¹³⁶. L'espagnol Zuluaga (1980) et les cubains Carneado Moré et Tristán Pérez (1983) ouvrent la phraséologie espagnole à la tradition phraséologique nord-américaine, russe et allemande. L'empreinte de Zuluaga dans la phraséologie espagnole a été très importante. Dans sa thèse (1980), il élabore une classification des expressions figées de l'espagnol dans laquelle on peut observer l'influence des travaux de Casares, Bally, Coseriu et des générativistes, et qui en même temps a laissé son empreinte dans des classifications ultérieures, comme celle de Corpas Pastor (1996). En outre, Zuluaga a distingué différents degrés de figement et il a fait remarquer que l'idiomaticité ne caractérise pas toutes les unités phraséologiques. Alors, il a déjà signalé que le figement et l'idiomaticité sont des caractéristiques graduelles¹³⁷.

C'est dans les années 90 que la phraséologie prend son élan en Espagne. Ainsi, la phraséologie intéresse dans toute sa complexité : les ouvrages de Corpas Pastor (1996), de Ruiz Gurillo (1997, 1998) et de Mendivil Giró (1999) apportent d'importantes contributions en phraséologie descriptive ; Martínez Marín (1996, 1998) propose une perspective pragmatique ; les linguistes Conca i Martínez (1999) et Guía (Conca i Martínez et Guía, 2000) s'intéressent à la phraséologie diachronique ; les lexicographes Varela et Kubarth (1994) élaborent un dictionnaire phraséologique ; les ouvrages collectifs de G. Wotjak (1998) et de Luque Durán et Bertrán (1998) proposent des travaux dans une perspective appliquée et lexicographique ; les travaux de Sevilla Muñoz (1988, 1992a, 1992b, 1993) portent essentiellement sur la parémiologie et la phraséologie contrastive espagnol-français (Sevilla Muñoz et Arroyo Ortega, 1993 ; Sevilla Muñoz et González Rodríguez, 1994) ; l'approche contrastive intéresse également Alonso Ramos (1998) pour l'espagnol-français ; pour l'espagnol-anglais, les

¹³⁶ Ruiz Gurillo (1997) souligne que l'œuvre de Casares (1950/1992) a été traduite en russe vers 1958 et il est fort probable qu'elle a été connue du linguiste soviétique Vinogradov. Les idées de Casares auraient donc pu influencer les recherches russes en phraséologie.

¹³⁷ Dans la section 3.2. *Caractéristiques des unités phraséologiques*, nous montrons que ces deux notions et leur caractère graduel sont très importantes pour la caractérisation de ce qui est phraséologique dans la langue.

travaux de Corpas Pastor (1998, 2000) sont une référence incontournable ; la didactique de la phraséologie est abordée par Penadés Martínez (1999), etc.¹³⁸

Les débuts du XXI^e siècle dessinent un avenir prometteur pour la phraséologie en langue espagnole. Ainsi, on continue à publier des travaux en phraséologie descriptive tels que les ouvrages monographiques de Koike (2001), González Rey (2002, 2021), Iñesta Mena et Pamies Bertrán (2002), Alonso Ramos (2004b) ou García-Page (2008), entre autres, ou les deux volumes de la revue *Lingüística española actual*, 23.1 (2001) et 24.1 (2002) entièrement consacrés à la notion de collocation. En lexicographie, à part la publication de dictionnaires phraséologiques à caractère plutôt traditionnel (Buitrago Jiménez, 2002 ; Cantera, 2007 ; Penadés Martínez, 2002, 2005a y 2008 ; Seco et al., 2004), on a vu apparaître les remarquables dictionnaires de combinaisons de mots *Redes* (Bosque, 2004) et *Práctico* (Bosque, 2006)¹³⁹. Dans un autre domaine qui nous intéresse spécialement, les études contrastives français-espagnol, il y a eu de nombreuses contributions, par exemple : Blanco (1999, 2013), Alonso Ramos (2001a), Yllera (2001), Mogorrón (2002), Sevilla Muñoz et Cantera Ortiz de Urbina (2004), Gómez Fernández (2021b), Gómez Fernández et Uzcanga Vivar (2010b), González Royo et Mogorrón (2010), Blanco et Catena (2015), Uzcanga Vivar (2017), Navarro-Brotos (2023). Dans une perspective didactique, nous tenons à souligner les travaux de Gonzalez Rey (2006, 2012) en phraséodidactique.

Actuellement, le nombre de publications et de congrès autour de la phraséologie ne cesse d'augmenter en Espagne. Comme le signale Timofeeva (2008), « pese a su extremada juventud, hoy por hoy debemos reconocer que la fraseología española supo

¹³⁸ Pour une présentation plus riche des travaux en phraséologie espagnole et l'évolution des axes de recherche pendant les années 90, voir Corpas Pastor (2001a).

¹³⁹ Le dictionnaire *Práctico* est la version pratique du *Redes*. Les différences de ceux deux dictionnaires dérivent du but poursuivi. *Redes* a été conçu comme un outil de réflexion autour des restrictions de sélection lexicales. Il est plutôt destiné à des linguistes ou des spécialistes de la langue. Par contre, le dictionnaire *Práctico* est un outil d'aide pour trouver la combinaison de mots la plus appropriée. Il vise un public plus large. Le *Práctico* est une version simplifiée et en même temps augmentée (en nombre d'entrées) du *Redes*. Pour une étude comparative détaillée entre ces deux dictionnaires, voir Barrios Rodríguez (2007).

sacar provecho del camino recorrido por otras lenguas y situarse, en poco tiempo, a la vanguardia del estudio fraseológico en Europa » (p. 119).

2. Le statut de la phraséologie

2.1. Discipline autonome ou indépendante

L'étude de la phraséologie en tant que discipline autonome commence à se développer en Europe de l'Est. Le grand nombre de travaux en phraséologie publiés en URSS fait que certains linguistes, comme Kunin, Mauer ou Eisman, considèrent que la phraséologie est une discipline indépendante d'autres branches de la linguistique¹⁴⁰. Cette idée d'autonomie se répand dans l'école russe et donne lieu à une réflexion autour de l'indépendance de cette discipline. Par exemple, des spécialistes comme Vinogradov ou Kunin soutiennent que le caractère lexicosyntaxique des éléments phraséologiques les sépare de la lexicographie et de la grammaire.

Le passage vers une conception de la phraséologie au sens restreint¹⁴¹ a renforcé son autonomie. Ainsi, Éckert (1976, cité dans Ruiz Gurillo, 1997) considère que cette discipline est indépendante car elle étudie un domaine relativement autonome de la langue et qu'elle dispose d'un objet d'étude et d'une méthodologie scientifique indépendantes. Certains linguistes soutiennent que la phraséologie est un domaine doté de toutes les caractéristiques propres d'un système (polysémie, homonymie, synonymie, antonymie et variations syntagmatiques) et de tous les niveaux d'analyse (phonétique, morphologie, lexicologie et syntaxe).

Isačenco (1948) défend l'autonomie de cette discipline en s'appuyant sur le caractère improductif des unités phraséologiques qu'il étudie (les *clichés-phrases*, des groupes improductifs équivalents à des phrases). Il sépare la phraséologie de la lexicologie et de la grammaire car les groupes improductifs qui ont des équivalents lexicaux appartiennent au domaine d'étude de la lexicologie, tandis que ce qui est

¹⁴⁰ Ruiz Gurillo (1997, p. 34) signale qu'en 1970 Eisman estimait que la phraséologie soviétique comptait déjà quelques 10.000 travaux publiés.

¹⁴¹ Pour une étude plus approfondie sur la conception de la phraséologie au sens large ou au sens restreint, voir la section suivante.

productif est étudié par la syntaxe régulière de la langue. D'autres linguistes, comme G. Wotjak (1983), tout en adhérant à l'indépendance de la discipline, reconnaissent qu'elle est étroitement liée à la lexicologie (cf. Ruiz Gurillo, 1997).

Dans ses débuts, la phraséologie a essayé de se constituer en tant que discipline indépendante en se basant sur une conception « mathématique » du langage opposant le lexique (les variables) à la syntaxe (ensemble de règles combinatoires) (Kauffer et Keromnes, 2022). Ainsi, « initialement, la phraséologie semble avoir pour « raison d'être » de rendre compte des irrégularités et particularités propres de chaque langue » (Kauffer et Keromnes, 2022, p. 10).

Le caractère idiomatique des éléments phraséologiques est un autre argument pour revendiquer l'indépendance de cette discipline (cf. Makkai, 1975). En étudiant à fond l'idiosyncrasie de ces éléments, Čermak (1988) propose de construire une théorie sur les procédés d'idiomatization de la langue. Il semble être d'accord sur l'autonomie de la phraséologie, même s'il reconnaît que les caractéristiques qui définissent les éléments phraséologiques, ne sont pas exclusives à ceux-ci¹⁴².

En revanche, des positionnements contraires à l'autonomie de la phraséologie la situent sous la dépendance d'autres domaines. Anisi, Bally (1909/1951) inclut la phraséologie dans une approche stylistique de la langue. Pour sa part, Gläser (1986a) décrit un niveau phraséostylistique au sein de la stylistique. Ce niveau intermédiaire permettrait d'expliquer les effets communicatifs (des effets stylistiques) produits par les unités phraséologiques dans un texte.

Par ailleurs, les fonctionnalistes européens préfèrent concevoir l'étude de ces combinaisons stables de mots comme relevant de la lexicologie, la lexicographie ou la parémiologie (Ruiz Gurillo, 1997). Certains linguistes (cf. Coseriu, 1964/1986 ; Pottier, 1974) décrivent les unités phraséologiques avec les outils de la lexicologie. En effet, il existe de nombreuses relations entre les objets d'étude de ces deux disciplines : les unités phraséologiques ont très souvent une correspondance monolexicale, elles entretiennent des relations sémantiques de synonymie, antonymie, etc., elles ont une

¹⁴² Sur la non spécificité des critères qui caractérisent les constructions phraséologiques, notamment la non-compositionnalité sémantique, la non-substituabilité paradigmatique et la non-modifiabilité morphosyntaxique, voir Lamiroy (2008).

autonomie semblable à celle des mots et elles sont conçues par les locuteurs comme un bloc. Dans cette même perspective, des linguistes tels que Hausmann (1979, 1989) et Corpas Pastor (1996) trouvent que les unités phraséologiques sont stockées en mémoire par les locuteurs d'une langue comme les unités simples du lexique. Elles feraient donc partie, selon ces auteurs, du vocabulaire. Par conséquent, la phraséologie serait une sous-discipline de la lexicologie. D'autre part, la phraséologie a été traditionnellement rattachée à la lexicographie et à la parémiologie. *L'Introducción a la lexicografía moderna* de Casares (1950/1992), où l'auteur décrit les unités pluriverbales, en constitue un exemple relevant dans la littérature linguistique espagnole. Casares considère que seulement les locutions devraient être étudiées par les disciplines linguistiques, tandis que les dictons et d'autres éléments phraséologiques devraient être considérées dans la parémiologie. Malgré l'énorme tâche de compilation accomplie pendant des siècles dans les domaines de la lexicologie et de la parémiologie, ces travaux n'ont pas suffisamment contribué à une description des faits phraséologiques dans toute leur complexité (leur fonction, leur structure interne, les outils sémantiques et pragmatiques intervenant dans leur formation, etc.). Enfin, des linguistes tels que Gaatone (1997), M. Gross (1982) ou Rey (1985) estiment que les éléments phraséologiques relèvent en même temps du lexique et de la grammaire.

Actuellement, la tendance est d'entreprendre l'étude de la phraséologie dans une perspective intermédiaire. D'une part, la thèse de l'autonomie de cette discipline répondait plutôt à une nécessité de défendre que ces recherches devaient être abordées depuis une approche propre et spécifique. Cependant, certains de ces linguistes (tels que G. Wotjak ou Čermak) ont gardé une certaine réserve en considérant que la phraséologie entretient des relations directes avec la lexicologie et la syntaxe. D'autre part, le point de vue adopté est de considérer ces relations que la phraséologie entretient avec d'autres disciplines non pas comme une dépendance ou inclusion, mais plutôt comme la preuve de la place intermédiaire de la phraséologie : « aujourd'hui [la phraséologie] n'est plus à considérer comme une discipline distincte et autonome (cf. Telija, 1971), mais plutôt comme un carrefour intra- et interdisciplinaire, concernée qu'elle est par la morphosyntaxe, la pragmasémantique, le texte et le discours » (Gréciano, 2000c, p. 91). Ainsi, l'étude des unités phraséologiques relève, en partie, de la syntaxe, car il s'agit de combinaisons de mots (Zuluaga, 1980), qui d'ailleurs ne respectent pas souvent les règles syntaxiques. Elle relève aussi du lexique, car ces

éléments sont conçus par les locuteurs comme des blocs entretenant des relations sémantiques avec d'autres mots. Ces expressions sont donc semblables aux mots de la langue. L'étude des unités phraséologiques relève également de la morphologie, car il s'agit d'un phénomène de créativité lexicale, comme la dérivation ou la composition (cf. Fleischer, 1992 ; Polguère, 2007 ; Legallois, 2012 ; Uzcanga Vivar, 2020a). Enfin, elle relève encore de la pragmatique, car certaines unités phraséologiques assument une fonction discursive et interactionnelle (cf. Bardovi-Harlig, 2012 ; Blanco et Mejri, 2018 ; Coulmas, 1981 ; Cowie, 2001 ; Dostie, 2019, 2021 ; Dostie et Tutin, 2019, 2023 ; Fonagy, 1997 ; Kauffer, 2011, 2018 ; Mel'čuk, 2013 ; Tutin, 2020).

Mejri (2003a) explique pourquoi l'étude des phénomènes phraséologiques implique toutes les dimensions de la langue :

[...] c'est parce qu'il s'agit d'un fait lexical. Le lexique représente le siège vers lequel convergent toutes les dimensions : les phonèmes ne peuvent avoir d'existence hors des mots ; la syntaxe dite libre ne se conçoit qu'au moyen des unités lexicales véhiculant leurs constructions, traduites en termes de contraintes ; la sémantique prend racine d'abord dans le lexique, etc.¹⁴³ (p. 32-33)

Ainsi, la phraséologie se développe du point de vue interne par le biais de différentes approches (des études linguistiques en diachronie, synchronie, traductologie, terminologie, didactique, pragmatique, stylistique, textologie, typologie) et du point de vue externe, en relation avec d'autres branches de la linguistique et de la science (l'informatique, la sociolinguistique, la psycholinguistique, etc.). Ce positionnement intermédiaire implique, selon Ruiz Gurillo (1997), « reconocer la autonomía de la fraseología para poder dar una explicación satisfactoria de sus mecanismos, y al mismo tiempo, reconocer su capacidad de proyección hacia el exterior como recurso ejemplificador y explicativo de un buen número de fenómenos del lenguaje » (p. 42). Cette perspective intermédiaire permet à la phraséologie de garder son idiosyncrasie,

¹⁴³ Cet auteur signale que dans cet article il s'est limité aux principales dimensions, n'ayant pas évoqué d'autres aspects comme les caractéristiques pragmatiques, la valeur stylistique et les emplois discursifs des constructions phraséologiques.

tout en conservant des liens étroits avec d'autres disciplines de la linguistique ou de la science en général¹⁴⁴.

Dans ce même ordre d'idées, Mel'čuk affirmait déjà en 1995 :

There is no independent linguistic discipline phraseology similar to semantics, syntax, or morphology, each of which study a particular component of the language. Phraseology is rather a particular field of interest that concentrates on a particular type of linguistic signs and has to deal with everything, starting with semantics and ending with phonetics (particular pronunciation or prosody of phrasemes). That is why phraseology is so difficult, but so appealing! (Mel'čuk, 1995, p. 227)

2.2. Objet d'étude et limites de la phraséologie

La délimitation de l'objet d'étude de la phraséologie est controversée, car elle dépend directement du sens large ou restreint attribué au domaine (cf. Hundt, 1994b ; Corpas Pastor, 1995, 1998 ; Ruiz Gurillo, 1997 ; García-Page, 2004). Autrement dit, si la phraséologie peut être vaguement définie comme « the study of the structure, meaning and use of word combinations » (Cowie, 1994, p. 3168), il s'avère nécessaire de déterminer quelles « combinaisons de mots » sont des faits phraséologiques.

Dans un sens restreint, la phraséologie étudie des suites figées et idiomatiques équivalentes à un mot ou à un syntagme, tandis que dans un sens large, la phraséologie inclut aussi l'étude des suites figées, pas nécessairement idiomatiques, ayant une structure syntagmatique ou phrastique, voire supraphrastique.

Cette distinction est basée sur la thèse de l'École de Prague selon laquelle il existe des éléments centraux, qui sont bien adaptés aux paradigmes théoriques, et des éléments périphériques, éloignés des normes générales du système. Chaque catégorie contient des éléments prototypiques et des éléments périphériques. Le passage d'une catégorie à une autre est graduel, moyennant des transitions du centre vers la périphérie. Ainsi, les classes et sous-classes ne sont pas étanches, mais il existe une gradation et des interrelations entre elles (cf. Daneš, 1966).

¹⁴⁴ Pour approfondir sur les relations de la phraséologie avec toutes les dimensions de la langue, voir Mejri (2003a).

Ce modèle centre/périphérie de l'École de Prague a été repris en phraséologie pour distinguer les éléments phraséologiques figés et idiomatiques (le centre) de ceux qui sont seulement figés mais ne présentent pas ou ne présentent guère d'idiomaticité (la périphérie). Alors, la phraséologie conçue dans un sens restreint ne tient compte que du centre (cf. Fraser, 1970 ; García-Page, 2008 ; Ruiz Gurillo, 1997 ; Weinreich, 1966/1969) ; tandis que, dans un sens large, la phraséologie inclut le centre et la périphérie (cf. Casares, 1950/1992 ; Corpas Pastor, 1996 ; Coseriu, 1966 ; Fleischer, 1982/1997 ; Gläser, 1986b ; M. Gross, 1982, 1988 ; Tristá Pérez, 1980 ; Vinogradov, 1946/1977 ; G. Wotjak, 1983 et 1988 ; Zuluaga, 1980)¹⁴⁵.

Ainsi, dans une conception large du système phraséologique, les locutions occupent le centre, et les éléments supérieurs au syntagme occupent la périphérie. González Rey (2021) signale que parmi ces derniers, on trouve les collocations (Mel'čuk ; Tutin), les parémies (Gremias), les slogans et les titres de livres, de films, de presse (Galisson), les textes faits pour la récitation comme les prières, les poèmes, les chansons, les citations, etc. (Coseriu), les énoncés pragmatiques (Corpas Pastor), et des schémas textuels (Gréciano), entre autres. La conception la plus large de la phraséologie correspond à celle qui est adoptée en linguistique de corpus (G. Gross, Sinclair).

D'autre part, Ruiz Gurillo (1997) signale que les objets d'étude de la phraséologie dans un sens restreint sont :

las combinaciones de palabras tenidas como más representativas de las UFS [unidades fraseológicas], es decir, las expresiones idiomáticas (idioms) que incluyen los grupos equivalentes a nombres, adjetivos, verbos y adverbios cuyo significado complejo no se deriva del significado de sus constituyentes. (p. 69)

Donc, la conception restreinte de la phraséologie est basée sur les notions d'idiomaticité et de figement.

¹⁴⁵ Parmi ces auteurs, que ce soit dans la conception large ou restreinte de la phraséologie, il n'existe pas un accord unanime sur les types de structures qui doivent faire l'objet d'étude de la phraséologie. Par exemple, Lamiroy (2008, p. 89) attire l'attention sur la conception largement maximaliste de M. Gross, car pour ce linguiste l'expression « prendre son parapluie » dénote le sens 'prévoir qu'il va pleuvoir' ; étant un sens non compositionnel, cette expression figure parmi les expressions figées de la base de données établie par M. Gross.

Il y a quelques décennies, ces deux conceptions de la phraséologie étaient très distinctes. La conception large était la représentation privilégiée en URSS et d'autres pays de l'Est, tandis que la conception restreinte s'imposait dans le courant linguistique occidental (l'ancien structuralisme européen, notamment l'École de Genève, et les linguistes anglo-américains) (cf. Corpas Pastor, 2003). Cependant, l'échange scientifique entre ces deux blocs, la consolidation de la phraséologie en tant que discipline et le développement de la linguistique de corpus (notamment des études statistiques) ont entraîné le rapprochement de ces deux perspectives et un élargissement à de nouveaux types d'unités phraséologiques. En effet, l'objet d'étude phraséologique s'est élargi vers des combinaisons syntagmatiques telles que les *séquences récurrentes*¹⁴⁶, la *prosodie sémantique*, la *colligation* ou les *patrons (patterns)* (cf. Bolly, 2008, 2010, 2011)¹⁴⁷.

Comme le soulignent Legallois et Tutin (2013), du moment où les objets de la phraséologie s'émancipent de la lexicologie, ils « prennent des dimensions plus larges, mais aussi plus floues, contribuant ainsi à l'extension des disciplines concernées » (p. 9). Certes, nous avons déjà signalé que les recherches actuelles en phraséologie s'étendent vers de nombreuses disciplines de la linguistique, chaque discipline favorisant un type particulier d'unités. Ainsi, la syntaxe pointe les schémas phrastiques récurrents (cf. Fillmore et al., 1988 ; Francis, et al., 1996 ; Stefanowitsch et Gries, 2003) ; en linguistique textuelle, on s'intéresse, par exemple, aux séquences de cohésion discursive ou à des suites discontinues propres à des genres spécifiques (cf. Biber, 2009) ; en linguistique didactique, on centre l'enseignement sur la maîtrise des

¹⁴⁶ Ces éléments sont parfois appelés *segments répétés* (Habert et al., 1997).

¹⁴⁷ Le caractère phraséologique de ces combinaisons pourrait être résumé en termes de préférences : préférences lexicales (la séquence récurrente, p. ex. : *il y a*), préférence pour certaines classes sémantiques (la prosodie sémantique, p. ex. : *commettre + crime, meurtre, faute*, etc.), préférence grammaticale (la colligation, p. ex. : [*Il + (Pron. pers.) + semble + Adj. + que*]) et préférence pour certaines structures syntaxiques (le patron, p. ex. : en anglais, le verbe *to leave* apparaît à l'infinitif après le verbe *to decide*, mais il prend la forme du gérondif en *-ing* après *to hate*). Pour une présentation succincte de ces phénomènes phraséologiques, voir Bolly (2010, p. 17 ; 2011, p. 54-63). Cette extension du phénomène phraséologique se produit notamment en linguistique statistique, héritière de la linguistique de corpus anglo-saxonne, la fréquence de co-apparition étant calculée sur des corpus.

séquences répétées ayant une fonction interpersonnelle et textuelle (cf. Scott et Tribble, 2006), etc.

2.3. Phénomène marginal ou central de la langue

Les spécialistes en phraséologie se sont souvent demandé quelle est la place de la phraséologie dans le système de la langue. Déterminés par les principes de centre et périphérie de l'École de Prague, notamment dans les études soviétiques, et aussi par les thèses chomskiennes de *grammaire nucléaire* et *périphérie*, surtout chez les générativistes (cf. Mendívil, 1998), on a souvent perçu les phénomènes phraséologiques comme marginaux et relevant d'un certain dysfonctionnement de la langue. Ainsi, selon Leska (1966) le centre de la langue est caractérisé par sa productivité et sa régularité (la langue réglée), tandis que les éléments linguistiques de la périphérie sont des signes non productifs (les éléments phraséologiques).

Pourtant, les études sur corpus ont mis en lumière l'importance, du point de vue quantitatif, des constructions figées. Déjà en 1970, il a été démontré que 20 % des textes est composé de segments répétés (Fiala et al., 1978). Plus tard, les travaux en lexique-grammaire ont révélé que le nombre d'unités phraséologiques est bien supérieur aux unités monolexicales appartenant à la même partie du discours (M. Gross, 1982, 1988). Pour sa part, Mel'čuk (1993) estime que le nombre de *phrasèmes*¹⁴⁸ en français est de « dizaines de milliers » (p. 83) : 40 % d'un texte et 90 % d'un dictionnaire (Mel'čuk, 1994). Dans un ouvrage ultérieur, il affirmera : « Les phrasèmes abondent dans la langue ; les textes de tous les genres en regorgent » (Mel'čuk, 2013, p. 129). Depuis des décennies, on s'accorde à dire que la phraséologie occupe une place centrale dans la langue (cf., entre autres, G. Gross, 1996a ; Hausmann, 1997 ; Mejri, 1997). Ainsi, selon G. Gross (1997) :

« il s'agit d'un fait central dans les langues : il y a près de 200 000 noms composés ou dont la combinatoire n'est pas libre, près de 15 000 adjectifs et au moins 30 000 verbes figés. On voit qu'on est en face d'une réalité de première importance et qui est une caractéristique des langues naturelles au même titre que la double articulation dont parlait Martinet ». (p. 201)

¹⁴⁸ Dans la terminologie de la théorie Sens-Texte, « un *phrasème* de la langue L est une expression multilexémique de L qui ne peut pas [...] être produite, à partir d'une situation donnée ou d'un sens donné, selon un dictionnaire de mots de L et à partir des règles générales standard de L » (Mel'čuk 1993, p. 83).

En effet, les travaux sur des corpus textuels de Wray et Perkins (2000) ont démontré que 70 % de la production langagière (écrite ou orale) d'un natif peut être « formulaic »¹⁴⁹. Ces chiffres ne sont pas absolus, puisqu'ils ont été calculés à partir de conceptions différentes de la phraséologie. Néanmoins, elles permettent de souligner l'importance de phénomène.

À part ces données quantitatives, des changements théoriques avaient commencé à s'opérer, depuis les remarques de Chafe (1968) jusqu'à l'extension de la notion d'idiomaticité proposée par Hausmann (1997). Loin d'être un fait accidentel de la langue, la phraséologie serait donc inhérente aux langues naturelles, c'est-à-dire que toute langue en usage par une communauté produit des séquences figées. La description des langues ne peut pas se passer de la phraséologie. Le caractère improductif qu'on a traditionnellement associé à la phraséologie est de nos jours absolument révolu (cf. Anscombe, 2019 ; Uzcanga Vivar, 2020a). La phraséologie est un processus dynamique (Gréciano, 1983 ; Mejri, 1997, 2003a) et, comme le soutient Mejri (2002), elle « occupe une place privilégiée parmi les procédés et les processus à l'œuvre dans le renouvellement du lexique » (p. 215).

Par ailleurs, ce développement des outils informatiques a permis aussi de remettre en question la régularité de la langue normée et l'idée d'irrégularité associée à la phraséologie. Ainsi, une étude du lexique-grammaire portant sur plus de 30 000 entrées figées et près de 12 000 phrases libres a montré que pratiquement toutes les structures figées ont une structure syntaxique régulière et qu'elles subissent les mêmes règles de syntaxe que les phrases libres (M. Gross, 1988).

¹⁴⁹ Wray (2002) définit la *formulaic sequence* (une définition dominante dans la phraséologie anglophone) comme une « sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at time of use, rather than being subject to generation or analysis by the language grammar » (p. 9).

3. Qu'est-ce qu'une unité phraséologique ?

3.1. La terminologie phraséologique

La terminologie est un indice relativement fiable dans la fixation et stabilisation des objets de la recherche (Mejri, 2005). Or, l'un des maillons faibles de la phraséologie, qui aurait pu entraver sa constitution et son développement en tant que discipline, a été la profusion et la confusion terminologique caractérisant ce domaine (Granger et Paquot, 2008). Dans la même veine, Cowie (1998b) affirme que la phraséologie est « a field bedevilled by the proliferation of terms and by the conflicting uses of the same term » (p.210). En effet, d'une part, il existe plusieurs termes pour désigner un seul type de séquence figée, et d'autre part, un seul terme peut se référer à plusieurs faits phraséologiques différents. À titre d'exemple, Martins-Baltar (1997) a repéré 68 appellations différentes dénommant les objets d'étude de la phraséologie¹⁵⁰ :

à-peu-près (Heinz, Rastier), allusion (Rey), aphorisme (Mochet / Cintrat, Rey), circonlocution (Grimaldi), citation, (Rey, Candel), cliché (Rey), collocation (Candel, Chanier / Fouqueré / Issac, Fónagy, Gentilhomme, Hausmann, Léon / Mazière, Leroy-Turcan, Rey), combinaison idiomatique (Bennet), composé (Gross, Schön), construction (François / Grass, Hausmann, Rey), construction figée (Gross), dicton (Candel), énoncé lié (Fónagy, Martins-Baltar), expression (Grimaldi, Léon / Mazière, Rey), expression idiomatique (Candel), expression proverbiale (Mochet / Cintrat), expression semi-figée (Chantier / Fouqueré / Issac), façon de parler (Grimaldi, Leroy-Turcan), fonction lexicale ([notion introduite par Mel'čuk, voir p. ex. Mel'čuk *et al.*, 1992, p. 127 ss.]¹⁵¹ Gentilhomme, Gross), forme de langage particulière (Grimaldi), formulation (Hausmann), formule (Léon / Mazière, Rey), fragment lié (Fónagy), gallicisme (Rastier), groupement discursif (Lewin / Mochet / O'Neil), idiome (Bennet, Léon / Mazière, Rey), idiotisme pragmatique (Burger, cit. par Güllich / Krafft), langue de bois (Branca-Rosoff, Fónagy), lexie complexe (Rey), locution (Grimaldi, Rey), locution nominale (Gross), locution plébee (Leroy-Turcan), locution terminologique (Candel, Rey), locution toute faite (Candel), manière de s'exprimer (Grimaldi), maxime (Rey), modèle locutionnel (Martin), mot (d'auteur) (Rey), mot composé (Corbin, Meunier-Crespo, Piguët), mot syntaxique (Corbin), palimpseste (Galisson), phrase faite (Grimaldi), phrase figée (Fotopoulou, Rey), phrase

¹⁵⁰ Parmi ces termes, il y en a qui désignent un type spécifique de combinaisons de mots, par exemple *aphorisme* ; d'autres comprennent plusieurs types, comme *expression semi-figée* ; et d'autres les recouvrent tous.

¹⁵¹ Nous tenons à signaler que le terme *fonction lexicale* ne désigne pas un type d'unité phraséologique mais un outil formel servant à modéliser les liens lexicaux. Voir Mel'čuk (1996).

idiomatique (Kleiber, cit. par Martin), phrasème (Hausmann, Gréciano, Rey), phrasème de spécialité (Gréciano), phraséolexème (Gréciano), phraséologisme pragmatique (Burger *et al.*, cit. par Gülich / Krafft), phraséoterm (Gréciano), phraséotextème (Gülich / Krafft), proverbe (Candel, Mochet / Cintrat, Rey), schéma (Grunig, cit. par Gülich / Krafft), séquence figée (Candel), séquence polylexicale (Gréciano), série phraséologique (Bally 1909, cit. par Gréciano), slogan (Rey), stéréotype (Mochet / Cintrat, Rey), structure préformée / préfabriquée (Gülich / Krafft), suite composée (Gross), syntagme figé (Corbin, Rey), syntagme idiomatique (Bennet, Corbin), syntagme terminologique (Candel, Rey), terme complexe (Candel, Rey), tour (Grimaldi, Léon / Mazière, Rey), tournure (Hausmann, Rey), unité phraséologique (Bally 1909, cit. par Gréciano), unité polylexématique (Corbin). (p. 23-24).

Nous tenons à ajouter à cette liste des termes très enracinés dans la tradition linguistique en langue française tels que *synapsie* (Benveniste, 1967), *synthème* (Martinet, 1967), *lexie complexe* ou *lexie composée* (Pottier, 1974 et 1987), *unité multilexémique*, *phrasème* ou *syntagme phraséologique* (Mel'čuk, 1993), *séquence figée* (Mejri, 1999), *séquence polylexicale* (Mejri, 2004) ou *phraséologisme* (Mejri, 2018).

En phraséologie espagnole, on trouve aussi une grande abondance terminologique. Ainsi, Corpas Pastor (1996) dresse une liste des termes utilisés en espagnol pour se référer à tous les types de combinaisons de mots :

- *Expresión pluriverbal* (Casares, 1950/1992); *unidad pluriverbal lexicalizada y habitualizada* (Haensch et al., 1982; Corpas Pastor, 1995 [1994]) o *unidad léxica pluriverbal*, en Hernández (1989). [...]
- *Expresión fija* (Zuluaga, 1980; García-Page, 1990; Martínez Marín, 1991). [...]
- *Unidad fraseológica* o *fraseologismo* (Zuluaga, 1980; Haensch et al., 1982; Carneado Moré, 1983; Tristán Pérez, 1988; Martínez López, 1996). (p. 17)

De nos jours, le terme le plus répandu en espagnol est *unidad fraseológica*¹⁵², car ces éléments fonctionnent comme un bloc (une unité) et ils sont formés d'au moins deux mots¹⁵³.

¹⁵² Ce terme avait déjà été utilisé par Bally en 1909 pour désigner un type de combinaison stable de mots. D'ailleurs, Corpas Pastor (1996) signale que ce terme est très répandu aussi dans deux grands blocs de la recherche linguistique : l'Europe occidentale et les anciens pays soviétiques. En espagnol, le premier auteur qui utilise ce terme est Zuluaga (1980).

¹⁵³ À l'exception de certains pragmatèmes composés d'un seul lexème tels *amitiés* ou *salut*.

Pour notre travail, nous avons choisi d'employer les termes *unité phraséologique* et *séquence figée* dans un sens large. Le premier est la dénomination la plus courante actuellement, et le second est le terme de préférence dans le modèle des classes d'objets.

D'après Mejri (2005), « cette situation [de multiplicité terminologique en phraséologie] traduit le flottement conceptuel dans ce domaine. Il y a une instabilité terminologique qui concerne aussi bien le phénomène que le produit. » (p. 184). En effet, comme nous le montrerons à continuation, les spécialistes ne se sont pas mis d'accord sur une définition de l'unité phraséologique ni sur une classification des différents éléments phraséologiques¹⁵⁴.

3.2. Caractéristiques des unités phraséologiques

Au cours de toutes ces décennies d'étude de la phraséologie, le concept d'unité phraséologique a subi des changements sensibles. Celui-ci dépend directement de la conception large ou restreinte de la phraséologie. Il n'existe pas encore une définition claire et univoque qui soit largement acceptée dans la communauté linguistique. L'affirmation de Mel'čuk (2013) : « aucune définition rigoureuse de ce qu'est un phrasème n'est universellement acceptée par notre communauté » (p. 129) reste vraie en 2023. Néanmoins, certaines caractéristiques des phénomènes phraséologiques sont traditionnellement admises depuis les travaux de Bally (1909/1951) et de Bréal (1897/1976), notamment la polylexicalité, le figement, l'idiomaticité et la fréquence d'emploi.

Corpas Pastor (1996), à partir de l'étude de la littérature sur ce sujet, relève les aspects sur lesquels il existe un certain consensus :

- (a) Se trata de una expresión formada por varias palabras (Casares, 1950/1992; Matešić, 1983; Cowie, 1985, 1989, 1991; Braasch, 1988; Gates, 1988; Hernández, 1989; Jackson, 1989 [1988]).

¹⁵⁴ D'autant plus que les limites entre ce qui est phraséologique et ce qui ne l'est pas ne sont pas claires du fait des différentes conceptions de la phraséologie. En outre, nous avons déjà signalé que la multiplicité de disciplines (linguistiques et au-delà de la linguistique) qui participent à la description des faits phraséologiques apportent une énorme richesse, mais en même temps elles apportent aussi une certaine imprécision dans la définition et délimitation du phénomène phraséologique (Legallois et Tutin, 2013).

- (b) Esta se caracteriza por estar institucionalizada (Cowie, 1985, 1989, 1991; Gläser, 1986; Hougaard, 1986; Alexandrova y Ter-Misanova, 1987; Braasch, 1988; Gates, 1988).
- (c) Por ser estable en diverso grado (Cowie, 1985, 1989; Kühn, 1985; Gläser, 1986b; Braasch, 1988; Jackson, 1989; Carter, 1989 [1987]; Nuccorini, 1992; Moon, 1992; Verstraten, 1992; Winter, 1992).
- (d) Por presentar cierta particularidad sintáctica o semántica (Cowie, 1985; Kühn, 1985; Gläser, 1986b; Braasch, 1988; Gates, 1988; Carter, 1989 [1987]; Nuccorini, 1992; Moon, 1992; Verstraten, 1992; Winter, 1992).
- (e) Y por la posibilidad de variación de sus elementos integrantes, ya sea como variantes lexicalizadas en la lengua o como modificaciones ocasionales en contexto (Gläser, 1986b; [M.] Gross, 1988; Sypnicki, 1991; [B.] Wotjak, 1992). (Corpas Pastor, 1996, p. 19-20)

En se basant sur ces caractéristiques générales, l'autrice propose cette définition de l'unité phraséologique (Corpas Pastor, 1996) :

[...] son unidades léxicas formadas por más de dos palabras gráficas en su límite inferior, cuyo límite superior se sitúa en el nivel de la oración compuesta. Dichas unidades se caracterizan por su alta frecuencia de uso, y de coaparición de sus elementos integrantes; por su institucionalización, entendida en términos de fijación y especialización semántica; por su idiomatidad y variación potenciales; así como por el grado en el cual se dan todos estos aspectos en los distintos tipos. (p. 20)¹⁵⁵

D'autre part, González Rey (2021), à partir de la consultation de 18 dictionnaires, dresse une liste des caractéristiques pouvant contribuer à la reconnaissance des unités phraséologiques :

- la polylexicalité ;
- la fréquence ;
- le figement ou lexicalisation ;

¹⁵⁵ Dans un travail ultérieur, Copras Pastor (2003) nuance cette définition : « Entendemos por unidad fraseológica (UF) una combinación estable de unidades léxicas formada por al menos dos palabras gráficas, cuyo límite superior se sitúa en el nivel de la oración compuesta. Presenta los siguientes rasgos distintivos: polilexicalidad, alta frecuencia de aparición y de coaparición, institucionalización, estabilidad (fijación y especialización), idiomatidad y variación potenciales) » (p. 131). D'une part, cette nouvelle définition lui permet de rendre compte de la conception restreinte de la phraséologie (« combinación estable de al menos dos palabras, que, en virtud de las distintas corrientes, tendrá como límite superior el sintagma o la oración compuesta »). D'autre part, l'autrice introduit la notion de stabilité, à laquelle elle relie les propriétés de figement et de spécialisation sémantique.

- le défigement , désautomatisation ou délexicalisation ;
- l'institutionnalisation ;
- l'idiomaticité ;
- la figuralité ;
- l'iconicité ou figuration ;
- l'opacité ;
- l'ambiguïté ;
- l'écart ou dérivation ;
- la moulabilité ou productivité ;
- la répétition ;
- la reproduction ;
- les différents registres ;
- la réductibilité ;
- l'arbitrariété, la motivation et la démotivation ;
- la valeur métaphorique ;
- la remétaphorisation ;
- les éléments expressifs et les procédés productifs. (p. 53)

Dans ce qui suit, nous présentons une brève description des propriétés permettant de reconnaître les unités phraséologiques. Nous avons considéré différentes approches à la phraséologie, ayant différents objectifs. Aucune des propriétés présentées n'est nécessaire ni définitive ; d'autant plus si nous concevons la phraséologie dans sa conception large. Comme le signale Gréciano (2000c) : « il faut retenir aujourd'hui la nature "prototypique" des caractéristiques phraséologiques qui se contentent de ressemblance de famille (p. 92). Par ailleurs, la plupart de ces propriétés sont gradables. Nous ne recueillons ici que les propriétés communes ou les plus fréquemment signalées¹⁵⁶.

¹⁵⁶ Nous tenons à signaler que ces caractéristiques ne constituent pas des compartiments étanches mais fonctionnent de manière collective. De plus, par souci de praticité, nous avons été obligée de schématiser la description de ces caractéristiques, ce qui entraîne une certaine perte de nuances.

3.2.1. La polylexicalité

Pour Gréciano (1995, 2000c) la polylexicalité, constitue, avec le figement et la figuration, l'un des traits définitoires des unités phraséologiques¹⁵⁷. En effet, la phraséologie concerne des constructions syntaxiques composées de plusieurs mots. D'une part, la limite inférieure est généralement placée à deux mots (*regard noir*). D'autre part, comme nous l'avons signalé plus haut, il n'y a pas d'accord sur la limite supérieure. Dans une conception restreinte de la phraséologie, la limite supérieure est le syntagme, tandis que dans une conception plus large, cette limite atteint la phrase (*Pierre qui roule n'amasse pas mousse*), voire tout un texte (p. ex., dans le cas des chansons, des prières, des cérémonies de mariage ou des procès de justice, etc.)¹⁵⁸.

Pourtant le concept de polylexicalité n'est pas de nature à couvrir toute la réalité phraséologique car un seul lexème pourrait former une expression phraséologique, notamment dans une situation d'énonciation (*Santé !*). La polylexicalité n'est donc pas un critère définitoire.

D'autres questions sur le degré d'extension surgissent, comme le cas de l'acceptation des mots composés séparés par des traits d'union (*compte-gouttes*) ou des apostrophes (*aujourd'hui*) en tant que des unités phraséologiques.

3.2.2. La fréquence

Les études en linguistique de corpus ont permis d'observer deux variants de cette caractéristique : la fréquence de coapparition des éléments formatifs d'une combinaison, et la fréquence d'apparition de ces expressions dans le discours. Quant à la fréquence de coapparition, l'emploi réitéré d'une séquence de mots dans le même ordre donne lieu à un processus de figement formel et sémantique (Corpas Pastor, 1996). Plus on répète cette combinaison de mots (fréquence), plus elle aura de chances de devenir une séquence figée que les locuteurs natifs vont stocker dans leur mémoire comme une unité lexicale (cf. Pawley et Sider, 1983).

¹⁵⁷ Cette linguiste n'utilise pas le terme *unité phraséologique* ; elle préfère les termes *phrasème*, *phraséologisme* ou *phraséolexème* (Gréciano, 2000c, p. 92).

¹⁵⁸ Gülich et Krafft (1997, p. 245) mentionnent comme des exemples de « modèles de texte » des faire-part de mariage ou de décès, la recette de cuisine, le résumé d'un article linguistique, les discours de bienvenue, etc. (cité dans Schmale, 2013, p. 32).

En ce qui concerne la fréquence d'apparition, nous avons déjà signalé que les unités phraséologiques constituent un phénomène massif dans les langues naturelles. D'ailleurs, les relevés de fréquence ont montré que certaines expressions sont particulièrement fréquentes dans certains genres de textes ou en situations de communication orale (González Rey, 2021).

3.2.3. L'institutionnalisation

García-Page (2008) définit l'institutionnalisation de la manière suivante :

Entendemos esencialmente por *institucionalización* el proceso por el cual una comunidad lingüística adopta una expresión fija, la sanciona como algo propio, como moneda de cambio en la comunicación cotidiana, como componente de su acervo lingüístico-cultural, de su código idiomático, como cualquier otro signo convencional, y pasa a formar parte del vocabulario. (p. 29)

Notons que cette caractéristique est intimement liée à la fréquence, que nous venons d'évoquer. Ainsi, la fréquence d'emploi et la répétition sont à la base du processus d'institutionnalisation : ces expressions répétées sont reconnues et mémorisées par la communauté de locuteurs d'une langue, elles deviennent conventionnelles et acquièrent un statut linguistique qui leur permet d'être répertoriées dans le dictionnaire (cf. Bolly, 2010 ; Corpas Pastor, 1996 ; González Rey, 2021 ; Grunig, 1997 ; Moon, 1998)¹⁵⁹. L'institutionnalisation est donc une notion diachronique car ces expressions « requièrent un certain temps de soudure pour se constituer en tant qu'expressions figées » (Lamiroy, 2008, p. 91). Les unités phraséologiques seraient des produits diachroniques qui nous sont parvenus à l'état synchronique actuel de la langue (Ruiz Gurillo, 1997).

3.2.4. Le figement

Indépendamment des approches linguistiques et des terminologies utilisées, les spécialistes sont d'accord sur le fait que le figement est l'une des propriétés essentielles des unités phraséologiques (Bolly, 2010, p. 12 ; Corpas Pastor, 1996, p. 23-24 ; Casares, 1950/1992, p. 170 ; García-Page, 2008, p. 25 ; G. Gross, 1996a, p. 198 ; Ruiz Gurillo, 1997, p. 15), voire une caractéristique universelle des langues naturelles (G. Gross,

¹⁵⁹ Ce facteur de répétition ou reproduction de ces expressions mène certains linguistes à identifier *reproductibilité* et *institutionnalisation* (Zuluaga, 1980 ; Gläser, 1986b ; Montoro del Arco, 2006).

1988, p. 70). Pourtant, la vaste littérature scientifique à ce sujet témoigne de la difficulté de définir cette notion avec précision (Lamiroy, 2008, p. 85), d'où la boutade de Robert Martin (1997) : « Nous sommes nombreux à trouver que c'est un thème admirable, sans pouvoir dire avec netteté ce que c'est » (p. 291).

Des auteurs tels que González Rey (2021) considèrent que le terme *figement* peut dénoter à la fois un sens inchoatif (le procédé de lexicalisation) et un sens résultatif (le résultat de ce procédé). Par ailleurs, de nombreux linguistes identifient le degré de figement avec le degré de phraséologisation. Polguère (2011b) ou Mel'čuk et Milićević (2014) préfèrent réserver le terme *figement* pour désigner la caractérisation d'un phrasème selon un certain nombre de paramètres (des transformations), et le terme *phraséologisation* pour dénoter « le processus par lequel un syntagme libre devient un phrasème » (Mel'čuk et Milićević, 2014, p. 302)¹⁶⁰. Pour d'autres auteurs, le terme *phraséologisation* est synonyme de *lexicalisation* (Coulmas, 1985 ; González Rey, 2002). Et encore, il y a des linguistes qui trouvent que le *figement* serait « une étape dans un processus dont la lexicalisation serait le terme » (Mortureux, 2003, p. 12). Il y en a d'autres qui identifient le *figement* à l'*institutionnalisation* (Baránov et Dobrovol'skij, 1998 ; Iñesta Mena et Pamies Beltrán, 2002). Un grand nombre de spécialistes utilisent de manière indistincte les termes *figement* et *stabilité* (Tristá Pérez, 1988 ; Hundt, 1994a ; Čermak, 2001). Cependant, selon Corpas Pastor (1996), suivant en cela l'optique de Bauer (1983), la *stabilité* comprend le *figement* et la *spécialisation sémantique* (ou *lexicalisation*). Elle considère que « ambos aspectos [la fijación y la especialización semántica] están relacionados, pues la fijación formal conduce al cambio semántico » (p. 23). Ces quelques exemples illustrent les différentes terminologies et interprétations de la notion de figement dans la littérature

¹⁶⁰ Mel'čuk considère que le terme *figé* ou *figement* est trop vague, car il peut référer à la cooccurrence syntagmatique, au manque de liberté syntaxique, à la sélection paradigmatique ou à la non-compositionnalité sémantique. En outre, étant une propriété gradable (une unité phraséologique peut être plus ou moins figée), elle ne devrait pas faire partie du concept de phrasème. Cet auteur reconnaît qu'il s'agit d'une « caractéristique importante, mais pas définitoire » (Mel'čuk, 2011, p. 48). Ainsi, il aborde la description de la phraséologie à partir de la notion de *contrainte*. Pour une réflexion autour des notions de *contrainte* et de *figement*, voir Mel'čuk (2011, p. 41-48 ; 2013, p. 145-147).

scientifique¹⁶¹ et, en conséquence, la difficulté à trouver une définition complètement "étanche" (Lamiroy, 2008, p. 96).

Un autre aspect qui complique la tâche de trouver une définition simple et univoque est le caractère polyfactoriel du figement. Le figement implique toutes les dimensions du système de la langue (cf. Mejri, 2002, 2008b). Ainsi, dans la littérature scientifique on relève des critères très hétérogènes tels que la fixité syntaxique, la non-compositionnalité du sens, la non actualisation des composants des unités phraséologiques, l'aréférenciation des éléments, le blocage de la variation paradigmatique, le sens figuratif, la préfabrication, la conventionnalité, l'impossibilité d'une traduction mot par mot, parfois même les traces d'une langue ancienne, l'ancrage pragmatique, etc¹⁶². Néanmoins, nous pouvons dégager trois propriétés essentielles communément acceptées : le figement formel, le figement sémantique et le figement pragmatique¹⁶³.

Avant d'entamer la présentation de ces trois grands volets du figement, il s'avère nécessaire de préciser la notion de construction libre, car le figement est souvent défini

¹⁶¹ Pour une présentation plus riche de ces divergences, voir García-Page (2008). Pour une étude détaillée de la notion de figement, voir Mejri (1997) ou le numéro 82 de *Cahiers de lexicologie* (Mejri, 2003b).

¹⁶² La littérature scientifique consacrée à la question du figement est immensément riche. Nous ne signalons ici que quelques références en littérature francophone : François et Mejri, 2006 ; Gaatone, 1997 ; G. Gross, 1996a ; M. Gross, 1982, 1988 ; Gülich et Krafft, 1997 ; Martin, 1997 ; Mejri, 1997 et 2003a ; Schapira, 1999 ; Svenson, 2004 ; en littérature hispanophone : Casares, 1950/1992 ; Corpas Pastor, 1997 ; García-Page, 2008 ; Mendívil Giró, 1999 ; Montoro del Arco, 2006 ; Ruiz Gurillo, 1997 ; Tristán Pérez, 1988 ; Zuluaga, 1989 ; et en littérature anglophone : Everaert et al., 1995 ; Hudson, 1998 ; Hunston et Francis, 2000 ; Makkai, 1972 ; Moon, 1998 ; Nunberg et al., 1994 ; Tomasello, 2002 ; Wray, 2002.

¹⁶³ Selon Thun (1978), les unités phraséologiques peuvent présenter un *figement interne* et un *figement externe*. Le figement interne comprend le *figement matériel* et le *figement de contenu*. Il serait l'équivalent de ce que nous avons appelé figement formel et figement sémantique. Quant au *figement externe*, il englobe le *figement situationnel* (contraint par la situation de communication), le *figement analytique* (l'emploi de certaines unités phraséologiques pour l'analyse déjà établi du monde), le *figement « pasémique »* (contraint par le rôle du locuteur dans l'acte de communication) et le *figement positionnel* (l'emploi de certaines unités phraséologiques dans des positions déterminées dans l'élaboration de textes). Nous verrons ultérieurement que ce figement externe recouvre ce que nous avons appelé le figement pragmatique.

par opposition à cette dernière. Nous considérons que la définition proposée par Mel'čuk et Iordanskaja (2017) est particulièrement pertinente, parce qu'elle aide à comprendre l'ampleur de la notion de figement (ou de contrainte)¹⁶⁴ :

Un syntagme **AB** est libre [= non contraint] si et seulement si :

- 1) **AB** est libre paradigmatiquement – du point de vue de sa sélection par le Locuteur :
 - a. le syntagme **AB** est sélectionné indépendamment de la situation qu'il décrit et de la situation de son énonciation ;
 - b. le lexème **A** est sélectionné indépendamment du lexème **B** et vice versa.
- 2) **AB** est libre syntagmatiquement – du point de la combinaison de ses composants : **A** et **B** sont réunis strictement selon les règles générales de la langue :

$AB = A \oplus B$, ce qui veut dire que **AB** est compositionnel. (p. 95)

En somme, un syntagme est libre s'il est construit librement du point de vue syntaxique, sémantique et pragmatique.

Pour revenir à la description des trois propriétés essentielles déjà mentionnées (le figement formel, le figement sémantique et le figement pragmatique), du point de vue formel le figement concerne autant l'axe syntagmatique (l'aspect morphosyntaxique) que l'axe paradigmatique (l'aspect lexical). Comme nous venons de le signaler, on considère qu'une séquence est figée lorsqu'elle n'est pas librement construite par le locuteur¹⁶⁵. Il s'agit de structures dont l'ordre interne est pratiquement

¹⁶⁴ La citation reproduit l'orthotypographie et les symboles utilisés par les auteurs.

¹⁶⁵ Rappelons que les travaux de M. Gross ont mis en évidence que les phrases dites libres sont construites sur la base d'importantes contraintes morphosyntaxiques. Toute construction possède toujours un certain degré de contrainte sur le plan sémantique et sur le plan morphosyntaxique, ne serait-ce que par le fait qu'elle est conditionnée par les règles de grammaire de chaque langue ou par la structure actancielle des prédicats : « Les constructions libres sont caractérisées par l'existence de paradigmes permettant des substitutions définies par les contraintes d'arguments et par des modifications et des restructurations qui dépendent de la nature sémantique et syntaxique de la relation existant entre le prédicat et ses arguments » (G. Gross, 1996a, p. 19). Ainsi, on ne mange que ce qui est mangeable : *manger un train de chocolat* vs **manger un train* (Lamiroy, 2003). À la limite, la combinatoire libre pourrait être considérée comme un degré minimal de figement, puisqu'il existe toujours des contraintes de sélection (Lamiroy et Klein, 2005). D'ailleurs, Mejri (2011b) dénonce que ce qui pose problème dans la combinatoire libre est la dénomination même, car elle « laisse entendre que cette combinatoire repose sur la liberté absolue dans l'emploi des unités lexicales dans la chaîne syntagmatique ; ce qui donne lieu à

inaltérable. Ces séquences fonctionnent comme des éléments préfabriqués de la langue et elles sont stables comme le reste des unités lexicales de la langue (cf. Edmonds, 2013b ; Schmale, 2013). Le figement formel se traduit par le blocage des propriétés transformationnelles telles que la passivation (**La tangente a été prise par Luc*¹⁶⁶), la pronominalisation (**Luc l'a prise*), l'extraction (**C'est la tangente que Luc a prise*), la relativation (**La tangente que Luc a prise*), etc., ainsi que le blocage (ou du moins, la limitation) de la variation paradigmatique (*Luc a pris / emprunté / *saisi la tangente*)¹⁶⁷. Cependant, le figement n'est pas un phénomène absolu. D'ailleurs, l'équipe du lexique-grammaire a pu démontrer que le figement total est très peu fréquent : « Quand on examine les listes de séquences contraintes élaborées au LADL et au LLI, on constate que les suites totalement figées sont très minoritaires par rapport à celles qui ont des restrictions partielles » (G. Gross, 1996a, p. 22). Dans cette même veine, Mejri (2005) soutient que « le figement absolu ne touche en moyenne qu'une expression sur dix » (p. 187). L'étude de grands corpus a permis de relever le caractère graduel du figement. Il s'inscrit donc dans un continuum qui présente un très grand nombre de variations (Mejri, 2002). Ces variations peuvent être de différente nature : graphiques, phoniques, morphologiques, grammaticales ou lexicales (*no ser nada del otro mundo/jueves ; tener (muy) mala uva*)¹⁶⁸. Moins une séquence accepte de variations, plus elle est figée. Ainsi, le figement peut être décrit en termes de degrés. Par exemple, Mel'čuk (1995) distingue

une représentation qui ne correspond pas à la réalité de son fonctionnement » (p. 64). Cette liberté est donc relative.

¹⁶⁶ Tous ces exemples ont été empruntés à G. Gross (1996a, p. 12-13).

¹⁶⁷ Notons que de nombreuses unités phraséologiques jouissent d'une certaine liberté combinatoire. G. Gross (1996a) affirme que « les variantes sont plus fréquentes que le figement total » (p. 16). En effet, Mogorrón (2010) a démontré que 50 % des constructions verbales figées présentent des variantes. Dans la même veine, Navarro-Brotons (2013) a mené une étude exhaustive et systématique des parémies en français et en espagnol mettant en relief les nombreuses variations de ces unités phraséologiques auxquelles on a toujours associé un degré de figement très élevé. D'ailleurs, Anscombe (2011, 2019) introduit la notion de *matrice lexicale* : « Une matrice lexicale sera un schéma comportant des unités linguistiques fixes et des variables linguistiques » (Anscombe, 2011, p. 25).

¹⁶⁸ Pour plus de détails sur la variation, voir Corpas Pastor (1996, p. 27-30) ; García-Page (2008, p. 65-67, note 38) ; Mejri (2011b, p. 72-76).

entre phrasèmes complets, semi-phrasèmes et quasi-phrasèmes¹⁶⁹. Il est évident que les cas qui posent vraiment problème sont ceux qui représenteraient la limite inférieure¹⁷⁰ :

La difficulté de décider du caractère figé d'une expression provient de son caractère graduel, en particulier dans la zone des figements minimaux où se situent les cas moins prototypiques. [...] (ex. *je vous avoue que, attirer les foudres de qqn, se dessiner à l'horizon*, etc.). (Lamiroy et Klein, 2005, p. 136).

Ainsi, nous ne pouvons pas considérer le figement comme un critère définitoire, puisqu'il présente une gradation (Casares, 1950/1992 ; Mel'čuk, 2011).¹⁷¹

Bien qu'il s'agisse d'une caractéristique quantitative (Mel'čuk et Milićević, 2014), puisqu'elle est gradable, nous considérons que les systèmes de pondération calculant le degré de figement proposés par certains chercheurs comme Guimier et Oueslati (2006) ou Bolly (2008) sont difficilement applicables lorsqu'on travaille avec de grands corpus de dizaines de milliers de constructions (Lamiroy, 2008). En outre, nous estimons que les résultats obtenus ne sont pas satisfaisants car il est extrêmement difficile de mesurer le paramètre sémantique.

Le figement formel a été traditionnellement associé à des anomalies syntaxiques ou lexicales. Par exemple, Tristá Pérez (1988) considère que les anomalies sont des *indicateurs phraséologiques*. En effet, certaines unités phraséologiques présentent soit un lexique vieilli qui n'est plus utilisé en dehors de l'expression (*peu ou prou*), soit des structures syntaxiques écartées de la norme, s'agissant souvent de structures archaïques (*baisser pavillon*). Cependant, l'étude de grands corpus en linguistique informatique a démontré que la plupart des unités phraséologiques sont formées régulièrement (Burger, 1998 ; G. Gross, 1996a). Ainsi, Nárday-Szabó (2002) attire l'attention sur la différence

¹⁶⁹ Dans la section 3.3.2. *Typologie de Mel'čuk* le lecteur trouvera une présentation succincte de ces notions.

¹⁷⁰ Sur le manque de limite précise entre les constructions libres et les constructions figées ou contraintes, voir Fraser (1970) ; G. Gross (1996a, 1997) ; M. Gross (1982) ; Kauffer (2013) ; Klein (2007) ; Lamiroy et Klein (2005) ; Mel'čuk (1993, 1995) ; Martin, (1997) ; Mejri (1997, 2002) ; Weinreich (1966/1969), entre autres.

¹⁷¹ Mel'čuk (2011) pousse cette réflexion à l'extrême en considérant que le figement, étant gradable, ne peut pas faire partie du concept de phrasème. Il trouve qu'il s'agit d'une caractéristique importante, mais pas définitoire.

entre le figement syntaxique et l'anomalie syntaxique : « [Celle-ci] se distingue de la notion de figement syntaxique car [elle] sert à désigner une irrégularité dans la syntaxe et non une restriction des procédés syntaxiques normaux, caractérisant en général les syntagmes libres » (p. 76). Le figement ne suppose pas l'irrégularité, mais cette dernière peut constituer un indice de figement.

Du point de vue sémantique¹⁷², le sens global des séquences figées reste inchangé, « quels que soient l'époque ou le milieu » (González Rey, 2021, p. 55). Cette non-évolution sémantique est la preuve d'un figement sémantique. Certaines séquences figées sont composées de mots préalablement autonomes dont le sens, en dehors de la séquence figée, peut évoluer diachroniquement. Ces « anomalies lexicales » sont donc « des irrégularités dues à des archaïsmes ou à des lexèmes incompréhensibles » (Náray-Szabó, 2002, p. 74)¹⁷³. Par exemple, dans les locutions *à sa guise* ou *en guise de*, le mot *guise* n'est plus compris par la population en général et n'est plus employé, sauf par effet de style ; d'ailleurs le dictionnaire Petit Robert 2002 le considère un archaïsme.

En fait, le figement du sens est un phénomène beaucoup plus large. Svensson (2004, 2008) affirme que c'est un amalgame de quatre notions dichotomiques : motivation *vs* non-motivation, sens propre *vs* sens figuré, transparence *vs* opacité, et analysabilité *vs* inanalysabilité.

Enfin, du point de vue pragmatique, l'emploi des unités phraséologiques dans le discours peut être figé. Schmale (2013) parle d'une *empreinte pragmatique* dont toutes les unités phraséologiques seraient touchées plus ou moins fortement : « elles [les expressions préformées¹⁷⁴] sont rattachées pragmatiquement à des contextes

¹⁷² Certains auteurs décrivent ensemble le figement et l'idiomaticité (cf. Tristá Pérez, 1988 ; Corpas Pastor, 1996 ; Montoro del Arco, 2006). Ils trouvent qu'il n'est pas toujours possible de différencier objectivement le processus de figement formel et le changement sémantique de la construction, puisque souvent ils se produisent de manière simultanée. Nous sommes d'accord avec leurs arguments mais, par souci de clarté expositive, nous avons tenu à créer une section à part spécifiquement dédiée à l'idiomaticité ou la non-compositionnalité sémantique (cf. Chap. III, section 3.2.5. *L'idiomaticité ou non-compositionnalité*).

¹⁷³ Ce linguiste souligne la valeur stylistique supplémentaire des anomalies lexicales, de sorte qu'elle contribue à l'expressivité et à la fonction évocatrice de la phraséologie (Náray-Szabó, 2002, p. 74).

¹⁷⁴ Nous employons la terminologie de l'auteur.

d'utilisation, leur réalisation adéquate nécessitant le respect de conditions séquentielles, contextuelles, sociales, stylistiques, etc. très spécifiques » (p. 36). Selon cet auteur, la dimension pragmatique de la phraséologie ne serait pas exclusive des pragmatèmes¹⁷⁵. Il s'agit donc d'un figement discursif. D'ailleurs, des approches plus récentes de la phraséologie considèrent cette fonction discursive. À part leur fonction référentielle, les unités phraséologiques ont une fonction organisationnelle dans le texte, en tant que marqueurs discursifs (p. ex., *autrement dit*), et une fonction interactionnelle entre les locuteurs, en tant que routines conversationnelles (p. ex., *Tu parles !*) (Wray, 2002). Ces expressions sont clairement figées au niveau lexical et morphosyntaxique, mais elles le sont un peu moins du point de vue sémantique (Lamiroy, 2008).

3.2.5. L'idiomaticité ou non-compositionnalité

L'idiomaticité est considérée l'une des propriétés essentielles des unités phraséologiques. Certains auteurs la considèrent l'un des universaux du langage (Makkai, 1978) ou « l'essence même de la langue » (Hausmann, 1997, p. 284). Cependant, comme le signale García-Page (2008), « la idiomaticidad es, seguramente, la propiedad fraseológica que más críticas e hipótesis ha suscitado, partiendo, por ejemplo, del propio concepto, pues hay disparidad de opiniones acerca de qué debe entenderse por *idiomaticidad* » (p. 27)¹⁷⁶.

Les termes *idiomatique* ou *idiomaticité* ont deux grandes significations en linguistique. D'une part, il peut désigner ce qui est propre à une langue, c'est-à-dire « un modo de hablar que da lugar a términos o expresiones reconocidas como propias de una lengua » (Ruiz Gurillo, 1997, p. 93). L'idiomaticité relève donc de la culture d'une

¹⁷⁵ Terme introduit par Mel'čuk (1995) pour désigner le phrasème compositionnel restreint par la situation de communication extralinguistique dans laquelle il est énoncé. Fónagy (1982) avait déjà parlé d'unités qui ont une fonction essentiellement pragmatique (*unités pragmatiques*). Pour plus de détails sur cette question, voir Beliakov et Mejri (2015) ; Blanco et Mejri (2018) ; Corpas Pastor (1996) ; García-Page (2008) ; Moon (1998).

¹⁷⁶ Pour plus de détails sur les différends conceptuels et terminologiques à ce sujet, voir García-Page (2008, p. 70, note 45) ; Legallois et Tutin (2013, p. 6-7) ; Svensson (2004, p. 71-99) ; Tamba (2014).

communauté. Elle est le reflet d'un mode de pensée spécifique¹⁷⁷. D'autre part, ce terme réfère à la propriété sémantique de certaines unités phraséologiques selon laquelle le sens de celles-ci ne correspond pas à la somme des signifiés des éléments constitutifs (cf. Corpas Pastor, 1996 ; Fernando y Flavell, 1981 ; González Rey, 2002 ; G. Gross, 1996a ; Hudson, 1998 ; Makkai, 1972 ; Moon, 1998 ; Svenson, 2004 ; Zuluaga, 1980 ; etc.)¹⁷⁸ ; par exemple, l'unité phraséologique *poser un lapin*, dont le sens n'a pas de relation avec les sens des lexies *poser* et *lapin*. Dans cette seconde acception, le terme *idiomaticité* est habituellement utilisé comme équivalent de *non-compositionnalité sémantique* ou *sens non compositionnel*¹⁷⁹.

Traditionnellement la non-compositionnalité a été considérée l'une des propriétés fondamentales, voire nécessaires, des unités phraséologiques¹⁸⁰. Actuellement, les spécialistes soutiennent plutôt que la non-compositionnalité constitue une caractéristique potentielle mais pas essentielle car il existe des unités phraséologiques dont le sens est compositionnel (p. ex., *avoir peur*). L'écart sémantique qui existe entre les sens des constituants de l'unité phraséologique et le sens global de celle-ci est de nature scalaire. Ainsi, la non-compositionnalité de ces constructions peut être totale (*casser du sucre dans le dos de qqn*), partielle (*partir les pieds devant*) ou nulle (*être rouge comme une tomate*)¹⁸¹. Il s'agirait donc d'une caractéristique

¹⁷⁷ Neveu (2011) souligne qu'on peut parler également de « l'idiomaticité d'un texte ou d'un discours spécialisé où se reconnaissent les particularités expressives d'un technolecte déterminé » (p. 189).

¹⁷⁸ Notons que les deux conceptions de l'idiomaticité sont étroitement liées. Ainsi, González Hernández (2016) définit l'idiomaticité comme le « rasgo que incide sobre el carácter idiosincrásico de las unidades fraseológicas de una determinada lengua » (p. 160), ou Olza Moreno (2009) soutient que « la fraseología es uno de los terrenos de la lengua en que los hablantes depositan de manera más frecuente y variada sus intuiciones y apreciaciones sobre la realidad » (p. 94).

¹⁷⁹ Désormais, nous allons employer de préférence le terme *non-compositionnalité* pour éviter l'ambiguïté sémantique du terme *idiomaticité*.

¹⁸⁰ Par exemple, l'approche de Lamiroy (2010).

¹⁸¹ Force est de constater que le degré de compositionnalité sémantique peut être différent au plan de l'encodage ou du décodage. Ainsi, une collocation telle que *se laver les dents* est compositionnelle au niveau du décodage, mais au niveau de l'encodage, elle pourrait ne pas l'être ; par exemple, un locuteur allemand pourrait produire une construction comme *se nettoyer* ou *cirer les dents* (Schmale, 2013). Nous

graduelle. D'ailleurs, des auteurs comme Burger (2007, p. 91), García-Page (2008, p. 66, note 39) ou Legallois et Tutin (2013, p. 7) signalent que le critère de compositionnalité est trop subjectif. Il est peu fiable, parce qu'il peut y exister une importante variation individuelle. En effet, des études expérimentales ont mis en évidence le peu d'accord entre les locuteurs (Titone et Connine, 1994) ou la tendance des ceux-ci à remotiver ces expressions (Geeraerts, 2003 ; Moon, 1998).

La non-compositionnalité est étroitement liée, voire déterminée par l'opacité, puisque cette dernière est le résultat du sens non compositionnel des composants de la séquence figée¹⁸². La signification globale de l'unité phraséologique devient obscure par rapport au signifié de ses constituants. Selon Mejri (2003a) ces deux notions ne se trouvent pas dans la même perspective :

Même si elles couvrent partiellement les mêmes réalités, elles renvoient à deux types de considération : la première opposition [(transparence / opacité)] privilégie la manière dont la séquence est perçue par les locuteurs ; elle se situe au niveau du décodage. La seconde [(sens compositionnel / sens non compositionnel)] intéresse, à notre avis, l'opération d'encodage, parce qu'on cherche à dégager la relation qui existe entre les constituants de la séquence et le produit global. (p. 29)

Dans cette même veine, dans le cadre de la théorie Sens-Texte, on conçoit la compositionnalité dans la perspective du locuteur, c'est-à-dire la perspective de la synthèse linguistique. Ainsi, Polguère (2015b) propose la définition suivante : « Un segment linguistique *S* est sémantiquement compositionnel si son sens est la somme du sens des signes linguistiques qui le composent et du sens porté par le mode de combinaison de ces signes dans *S* » (p. 258). L'intérêt n'est pas de savoir si un énoncé est interprétable à partir de la somme des sens de ses constituants, mais de savoir comment un énoncé doit être modélisé comme le résultat d'un assemblage ; autrement dit il faut « déterminer quels signes linguistiques sont assemblés, au moyen de quelles règles, pour exprimer un Sens donné 'S' au moyen d'un Texte *T* » (Polguère, 2015b, p.

reviendrons sur la relation entre la notion de la compositionnalité du sens et l'encodage et le décodage quelques lignes plus tard.

¹⁸² Notons que certains linguistes assimilent l'idiomaticité et l'opacité (Čermak, 2001 ; Clas y Gross, 1998 ; Fellbaum, 1993 ; Gaatone, 1984 ; Mellado Blanco, 1999) (cité dans García-Page, 2008, p. 70, note 45) ; d'autres considèrent que l'opacité est l'un des traits de l'idiomaticité (Baránov et Dobrovol'skij, 1998 ; González Hernández, 2016).

259). Alors, dans cette perspective, le pragmatème *Chaud devant!* est non compositionnel, malgré sa structure lexicosémantique complètement évidente.

Mel'čuk (2013) aussi met l'accent sur cette imprécision : « Il est donc important de ne pas confondre l'opacité / la transparence psychologique d'une expression, qui est assez subjective et relève d'un phénomène continue, avec sa compositionnalité, qui est objective et discrète » (p. 133). La non-compositionnalité serait donc une propriété binaire, tandis que l'opacité / la transparence présenterait des degrés¹⁸³. Cet auteur cite l'exemple de la locution *prendre le taureau par les cornes*, considérée non compositionnelle, puisque le signifié d'aucun de ses composants n'est repris dans la signification de cette expression (à savoir 'traiter la difficulté en question immédiatement et directement'). Pourtant, il trouve que ce sens est assez transparent pour un locuteur, pouvant donc comprendre facilement la métaphore. Le linguiste russe souligne qu'il existe une corrélation entre la non-compositionnalité et l'opacité : « une expression compositionnelle est nécessairement transparente ; mais cette corrélation n'est ni directe ni linéaire » (p. 133).

À ce sujet, Náray-Szabó (2002) note que dans le cas de l'opacité sémantique, « le sens est non seulement différent par rapport à la somme des sens primaires de ses composantes (= sens global), mais il est impossible de l'interpréter à partir de ceux-ci » (p. 74). Ainsi, le collocatif *chaleureux* dans *applaudissements chaleureux* pourrait bien être interprété comme un intensificateur. Au contraire, la locution *casser sa pipe* est non compositionnelle et opaque.

Les unités phraséologiques ont été considérées par le passé comme des exceptions au principe de compositionnalité, en d'autres termes des « anomalies sémantiques » (Saussure, 1916/1978 ; Sechehaye, 1921). Cependant, les études de sémantique lexicale ont mis en évidence leur caractère largement régulier (Legallois et Tutin, 2013). En effet, des projets comme *Widespread Idioms in Europe and Beyond* (Piirainen, 2008 ; Schmale, 2013) ont relevé l'existence d'expressions équivalentes dans des langues géographiquement proches (p. ex., *être au septième ciel*, en français ; *estar en el séptimo cielo*, en espagnol ; *to be in seventh heaven*, en anglais ; *im siebten*

¹⁸³ Pour une étude plus approfondie des notions de compositionnalité et non-compositionnalité, voir Godart-Wendling et al. (1998), Mel'čuk (2004c) et Svensson (2008).

Himmel sein, en allemand), voire dans des langues éloignées du point de vue géographie et génétique (Legallois et Tutin, 2013)¹⁸⁴.

3.3. Classification des unités phraséologiques

Nous sommes d'accord avec Corpas Pastor (1998) lorsqu'elle affirme :

El estudio de los sistemas fraseológicos de las lenguas muestra, paradójicamente, una cierta “universalidad”. A pesar de que tradicionalmente ha sido considerada la parte más idiosincrásica y propia de cada lengua, la fraseología, como sistema, presenta más similitudes que diferencias (p. 170)

En effet, les propriétés communes entre les unités phraséologiques des différentes langues semblent s'imposer sur ce qui est spécifique, mais il existe toute une panoplie de classifications se heurtant à de nombreuses difficultés, souvent dues à la complexité du phénomène. D'une part, le problème de délimitation de l'objet d'étude de la discipline, c'est-à-dire le fait de considérer la phraséologie depuis une conception large ou restreinte détermine le type d'unités phraséologiques qui vont composer le domaine phraséologique. D'autre part, les unités phraséologiques forment un ensemble très hétérogène. Pour pouvoir les classer, il est nécessaire d'identifier les unités linguistiques présentant des traits communs (Pecman, 2004). La difficulté provient du fait que les paramètres qui caractérisent la nature et le fonctionnement des unités phraséologiques sont très nombreux. Ainsi, comme le signale Pecman (2004), « c'est la priorité donnée à l'un ou aux plusieurs de ces paramètres qui va modeler le classement » (p. 128). Les critères le plus souvent retenus sont la structure syntaxique et la non-compositionnalité sémantique, mais d'autres critères s'avèrent aussi très importants, comme la partie du discours, le figement, ou le fait que les unités phraséologiques constituent des énoncés complets. Ainsi, il existe de nombreuses classifications mélangeant des critères d'ordre sémantique, syntaxique, pragmatique et dénominatif.

Dans ce qui suit, pour faire un petit tour d'horizon, nous passerons brièvement en revue des typologies faites par des linguistes anglophones, francophones¹⁸⁵ et

¹⁸⁴ Pour la sémantique cognitive, l'existence de ces régularités confirme que ces expressions sont motivées par des métaphores productives et plus ou moins universelles (cf. Kövescses, 2000).

hispanophones. Nous ne présenterons que quelques classifications qui s'avèrent intéressantes pour notre travail du fait de leur conception large de la matière phraséologique et parce qu'elles ont influencé notre conception de cette discipline¹⁸⁶. Nous avons donc retenu les classifications proposées par Gläser, Mel'čuk, G. Gross et Corpas Pastor.

3.3.1. Typologie de Gläser

Gläser (1988) propose une classification globalisante de la phraséologie anglaise en trois sphères : la « sphère centrale » (*centre*), la « sphère transitoire » (*transitional area*) et la « sphère périphérique » (*periphery*).

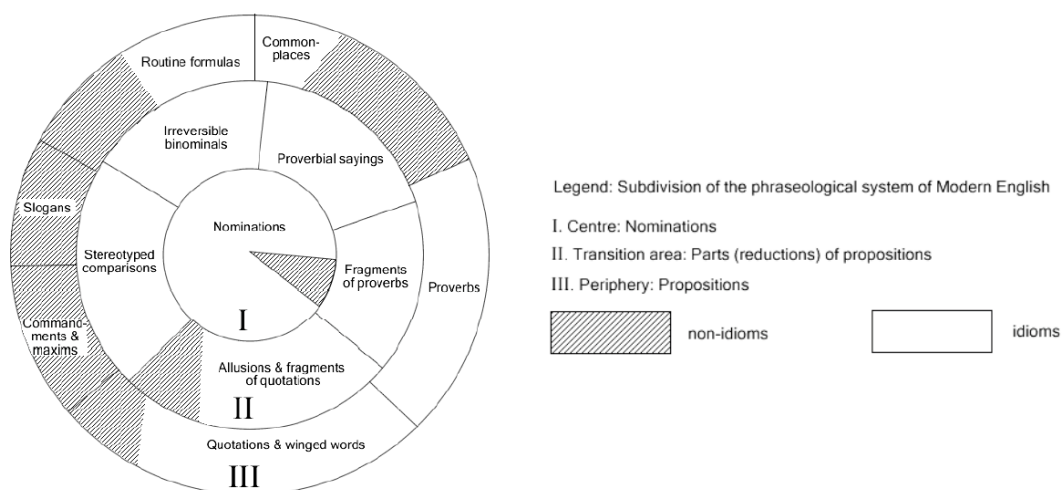


Figure 5 : Structure du système phraséologique selon Gläser (extrait de Pecman, 2004, p. 131)

La zone centrale contient les unités ayant une « fonction nominative », c'est-à-dire les unités qui ont la même fonction que des unités monolexicales, par exemple *bread and butter* (rôle nominal) ou *to have a swim* (rôle verbal). La zone transitoire inclut des unités qui se trouvent entre les monolèxèmes et les phrases. Cette linguiste les appelle des « idiomes sous forme de phrase ». Il s'agit de comparaisons stéréotypées (*as proud as a peacock*), de binômes irréversibles (*the ups and downs*), d'ellipses de proverbes (*a rolling stone*), etc. Finalement, la sphère périphérique englobe des unités

¹⁸⁵ Nous considérons Mel'čuk comme un linguiste francophone car, bien que sa langue maternelle soit le russe, il a beaucoup travaillé sur la description de la langue française.

¹⁸⁶ Rappelons que les collocations sont exclues de nombreuses classifications, car pour certains linguistes ayant une conception restreinte de la phraséologie, elles ne constituent pas des unités phraséologiques.

constituant des phrases : les proverbes (*An apple a day keeps the doctor away*), les clichés (*How do you do?*), les citations (*Nothing is certain except for death and taxes*), les truismes (*We only live once*), etc.

Cette typologie s'appuie également sur la fréquence d'usage ; les unités se trouvant dans la zone centrale sont plus utilisées que celles qui sont dans la zone périphérique.

3.3.2. Typologie de Mel'čuk

Mel'čuk (1993, 1995, 2011, 2013) construit une typologie basée sur les traits définitoires des phrasèmes¹⁸⁷. S'agissant d'une typologie déductive, elle a l'avantage d'obtenir des types universels de phrasèmes qui englobent toute la phraséologie de toutes les langues. Il s'agit d'un classement reposant sur deux axes : l'axe paradigmatique (les contraintes de sélection) et l'axe syntagmatique (les contraintes de combinaison ou leur compositionnalité).

Selon les contraintes de sélection, on distingue deux types de phrasèmes :

- le *phrasème lexical*, « dont le sens 'σ' est construit par le Locuteur librement pour n'importe quelle situation désignée, mais le choix des lexèmes pour exprimer 'σ' est contraint » (Mel'čuk, 2013, p. 131) ; par exemple, *porter son attention sur* ou *au bout du rouleau* ;
- Le *phrasème sémanticolexical*, « dont le sens 'σ' n'est pas construit par le Locuteur, mais sélectionné comme un tout de façon contrainte – en fonction du contenu conceptuel à verbaliser, donc en fonction de la situation désignée » (Mel'čuk, 2013, p. 131) ; la sélection des lexèmes pour exprimer ce sens est, le plus souvent, contrainte ; par exemple, *Vous dites ?* ou *Peinture fraîche*.

¹⁸⁷ Rappelons que cet auteur préfère le terme *phrasème*, au lieu d'*unité phraséologique*. Il explique que « ce terme s'appuie sur le terme établi de phraséologie – discipline linguistique qui étudie les phrasèmes » (Mel'čuk, 2011, p. 41). Pour faire bref, un phrasème est un syntagme qui ne peut pas être librement construit par le locuteur, car le phrasème est contraint par des contraintes de sélection (paradigmatiques), ou par des contraintes de combinaisons (syntagmatiques), ou bien par toutes les deux.

Selon leur compositionnalité sémantique, les phrasèmes peuvent être compositionnels (*porter son attention, Peinture fraîche*) et non compositionnels (*au bout du rouleau*).

Le croisement de ces deux dimensions (contraintes lexicales vs contraintes sémanticolexicales et sens compositionnel vs sens non-compositionnel) donne comme résultat trois classes majeures de phrasèmes logiquement possibles¹⁸⁸ :

Nature des contraintes \ Compositionnalité des phrasèmes	non compositionnels	compositionnels
	lexicales	LOCUTIONS
sémantico-lexicales	impossible	CLICHÉS

Tableau 2 : *Trois classes majeures de phrasèmes* (extrait de Mel'čuk, 2013, p. 132)

Nous présentons ci-dessous de façon succincte les notions de locution, collocation et cliché dans le cadre de la théorie Sens-Texte :

- Une locution est un phrasème lexical non compositionnel. On distingue trois types majeurs, selon le degré d'inclusion, dans le sens de la locution, du sens de ses composants : les locutions fortes (*manger la grenouille, les carottes sont cuites, au bout du rouleau*), des demi-locutions (*fruit de mer, casser les oreilles, bain de foule*) et les locutions faibles (*donner le sein, rouge à lèvres, attendre un enfant*)¹⁸⁹.
- Une collocation est un phrasème lexical compositionnel (*porter son attention sur, amour fou, apprécier grandement*). Elle est semi-contrainte : « une de ses composantes est sélectionnée par le Locuteur librement, juste pour son sens ; c'est l'autre qui doit être choisie en fonction du sens à exprimer et de de la première composante » (Mel'čuk, 2013, p. 137)¹⁹⁰.

¹⁸⁸ Une quatrième classe étant impossible, car le caractère non contraint et la non-compositionnalité sont incompatibles (Mel'čuk, 2011).

¹⁸⁹ En complément, sur les locutions, voir Mel'čuk (2006b).

¹⁹⁰ En complément, sur les collocations, voir Mel'čuk (1998, 2003a, 2004d, 2006a).

- Un cliché est un phrasème sémanticolexical compositionnel. Le locuteur produit le cliché comme un tout, car il est complètement contraint par son référent (Mel'čuk et Iordanskaja, 2017). Alors, en fonction du référent on peut distinguer quatre types de clichés : les pseudonominèmes ou sobriquets (*or noir, la Ville lumière*), les termèmes (*la maladie de Parkinson, forage rotatif*), les formulèmes¹⁹¹ (*Joyeux anniversaire !, Quel âge avez-vous ?*) et les sentencèmes (*Jamais deux sans trois, Qui sème le vent récolte la tempête*)¹⁹².

Ainsi, la typologie universelle des phrasèmes peut être représentée sous le schéma suivant¹⁹³ :

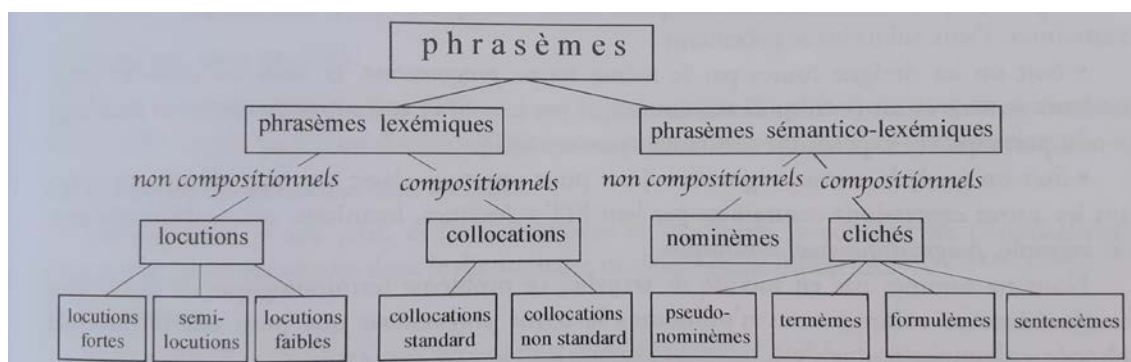


Figure 6 : Classification des phrasèmes (extrait de Mel'čuk et Iordanskaja, 2017, p. 103)

3.3.3. Typologie de G. Gross

Pour G. Gross (1996a) le trait principal des unités phraséologiques est le figement. D'ailleurs, il n'utilise pas le terme *unité phraséologique*, mais le terme *expression figée*. Étant donné que le figement est une propriété de nature scalaire, il se donne comme objectif de définir le degré de figement des unités phraséologiques à partir d'une série de transformations syntaxiques en suivant les règles de la grammaire

¹⁹¹ Notons qu'une sous-classe des formulèmes est particulièrement importante : les pragmatèmes (*Peinture fraîche, Défense de stationner*).

¹⁹² En complément, sur les clichés, voir Mel'čuk (2015).

¹⁹³ Mel'čuk et Iordanskaja (2017, p. 103-104) signalent que bien qu'il s'agisse d'une classification exhaustive, cette classification est simplifiée. Elle est commode pour un survol et une caractérisation générale, mais ces classes terminales peuvent se subdiviser d'avantage, en utilisant d'autres axes. En outre, les phrasèmes ne constituent pas une hiérarchie nette de classes.

générationnelle transformationnelle¹⁹⁴. Pour G. Gross les restrictions syntaxiques vont de pair avec l'opacité sémantique, car il considère que « le figement sémantique et le figement syntaxique sont deux aspects d'un même phénomène qu'il ne convient pas de séparer de façon artificielle » (G. Gross, 1996a, p. 8). Ainsi, le figement, l'opacité sémantique, la non-compositionnalité et les restrictions syntaxiques sont des phénomènes corrélés.

Pourtant, un classement des différents degrés de figement fondé sur des tests de transformations syntaxiques pourrait s'avérer une tâche pénible ayant des résultats peu exploitables, comme l'affirme l'auteur lui-même lorsqu'il analyse les locutions verbales :

Pour le [le classement des différents degrés de figement] mettre clairement en évidence, il faudrait examiner toutes les suites verbales non régulières au regard d'une batterie importante de critères et constituer une typologie de toutes les suites qui répondent aux mêmes critères. Il se pourrait qu'on obtienne plusieurs centaines de classes différentes. (G. Gross, 1996a, p. 88)

Par ailleurs, G. Gross (1996a) établit aussi une typologie complémentaire (préalable à celle du figement) basée sur le critère de la fonction. Ainsi, il distingue les noms composés, les déterminants composés, les verbes et locutions verbales, les locutions adjectivales, les locutions adverbiales et les locutions prépositives et conjonctives.

3.3.5. Typologie de Corpas Pastor

Corpas Pastor (1996, 1998) propose une classification fondée sur la combinaison de deux critères : l'énoncé et le figement. Selon cette linguiste, l'énoncé est « una unidad de comunicación mínima, producto de un acto de habla, que corresponde generalmente a una oración simple o compuesta, pero que también puede constar de un sintagma o una palabra » (Corpas Pastor, 1996, p. 51). Ce critère permet donc d'organiser les unités phraséologiques en deux grands groupes : celles qui sont des énoncés complets et celles qui ne le sont pas. En ce qui concerne le figement, Corpas

¹⁹⁴ Pour une explication plus en détail de ces transformations proposées par G. Gross, voir Chap. IV, section 3.6. *Un emploi est associé à des modifications.*

Pastor reprend la définition de Zuluaga¹⁹⁵ (1975) : « propiedad que tienen ciertas expresiones de ser reproducidas en el hablar como combinaciones previamente hechas » (p. 230). Corpas Pastor classifie les unités phraséologiques en fonction de leur figement dans la norme, dans le système ou dans la parole.

À partir de l'application de ces deux critères l'autrice établit un premier niveau de classification en trois sphères :

- *Sphère I.* Elle comprend les collocations, c'est-à-dire des unités phraséologiques seulement figées dans la norme et qui ne constituent pas des énoncés complets. Il s'agit de « sintagmas completamente libres a los que el uso les ha conferido cierto grado de restricción combinatoria» (Corpas Pastor, 1998, p. 172).
- *Sphère II.* Elle inclut des locutions. Ces unités phraséologiques sont figées dans le système et ne constituent pas des énoncés complets.
- *Sphère III.* Elle contient les énoncés phraséologiques. Selon Corpas Pastor (1998), « constituyen actos de habla y enunciados por sí mismos, además de estar fijados en el habla y formar parte del acervo sociocultural de la comunidad hablante » (p. 172).

Dans un second niveau de classification, chacune de ces sphères est divisée en sous-ensembles en fonction d'autres critères complémentaires. Voici un tableau schématique qui résume et exemplifie la classification de Corpas Pastor (1996) :

¹⁹⁵ Nous tenons à signaler que la classification de Corpas Pastor est fortement héréditaire de celles de Casares (1950/1992) et de Zuluaga (1980).

NO ENUNCIADO	ESFERA I (forman parte de la norma)	COLOCACIONES	<ol style="list-style-type: none"> 1. V + S (sujeto): <i>correr un rumor, declararse un incendio</i> 2. V + (prep. +) S (objeto): <i>asestar un golpe, poner en funcionamiento</i> 3. Adj./S + S: <i>momento crucial, visita relámpago</i> 4. S + prep. + S: <i>banco de peces</i> 5. V + Adv.: <i>negar rotundamente</i> 6. Adj. + Adv.: <i>opuesto diametralmente</i>
	ESFERA II (forman parte del sistema)	LOCUCIONES	<ol style="list-style-type: none"> 1. Nominales: <i>mosquita muerta, paño de lágrimas, el qué dirán</i> 2. Adjetivas: <i>corriente y moliente, más papista que el Papa, de rompe y rasga</i> 3. Adverbiales: <i>gota a gota, de tapadillo, a raudales</i> 4. Verbales: <i>llevar y traer, nadar y guardar la ropa, meterse en camisa de once varas</i> 5. Prepositivas: <i>gracias a, en lugar de</i> 6. Conjuntivas: <i>antes bien, como si</i> 7. Clausales: <i>salirle a alguien el tiro por la culata, como quien oye llover.</i>
ENUNCIADO	ESFERA III (forman parte del habla)	1. PAREMIAS	<ol style="list-style-type: none"> 1. Enunciados de valor específico: <i>Las paredes oyen, Ahí le duele</i> 2. Citas: <i>El hombre es un lobo para el hombre</i> 3. Refranes: <i>Por la boca muere el pez, Un día es un día.</i>
		2. FÓRMULAS RUTINARIAS	<ol style="list-style-type: none"> 1. Fórmulas discursivas de apertura y cierre y de transición: <i>¿Qué hay?, Hasta luego, A eso voy</i> 2. Fórmulas psicosociales, expresivas, comisivas, directivas, asertivas, rituales y miscelánea: <i>Lo siento</i> (de disculpa), <i>Ya lo creo</i> (de consentimiento), <i>Ni hablar</i> (de recusación), <i>Dios se lo pague</i> (de agradecimiento), <i>Y usted que lo vea</i> (desear suerte), <i>Qué se le va a hacer</i> (solidaridad), <i>¡A mí plin!</i> (insolidaridad); <i>Ya te apañaré</i> (de promesa y amenaza); <i>Largo de aquí</i> (de exhortación); <i>Tú dirás</i> (de información); <i>No es para tanto</i> (de ánimo); <i>Por mis muertos</i> (de aseveración); <i>No te digo</i> (emocionales); <i>¿Qué es de tu vida?</i> (de saludo); <i>Le saluda atentamente</i> (de despedida); <i>Pelillos a la mar</i> (miscelánea).

Tableau 3 : *Classification des unités phraséologiques de Corpas Pastor* (extrait de Saracho Arnáiz, 2015, p. 118)

4. Les collocations

Nous avons déjà introduit la notion de collocation comme un type d'unité phraséologique (conception large de la phraséologie). Dans ce qui suit, nous allons approfondir sur cette notion qu'on a souvent placée à mi-chemin entre la combinatoire libre et la combinatoire figée. Nous allons essayer de définir la notion de collocation ou

du moins, de donner les propriétés qui la caractérisent. Ensuite, nous reviendrons sur la polémique du statut phraséologique des collocations.

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

La notion de collocation est déjà présente chez Bally (1909/1951) sous le terme *série phraséologique*¹⁹⁶. Par contre, le terme *collocation* apparaît pour la première fois dans le domaine de la didactique du lexique (Palmer, 1933). Puis, il est adopté dans un sens plus large dans les modèles fonctionnels anglais, d'abord, par Firth (1957b), et ensuite ce terme est redéfini sur une base statistique par Sinclair (1991)¹⁹⁷.

Nous souscrivons aux propos d'Evert (2008) lorsqu'il affirme : « The concept of collocations is certainly one of the most controversial notions in linguistics » (p. 1212). Le concept de collocation, de par sa nature, est difficile à formaliser et, comme nous le démontrerons, aucune définition ne fait l'unanimité. De ce fait, on a souvent eu tendance à l'ignorer. La possibilité d'enquêter de grandes quantités de données (les corpus) a mis en relief l'omniprésence des collocations. En effet, des analyses de corpus basées sur la fréquence statistique des mots ont prouvé la fréquence très élevée des cooccurrences lexicales dans le discours (cf. Bolly, 2008 ; Hausmann et Blumenthal, 2006). La linguistique de corpus a bouleversé la lexicographie, et par la suite, les besoins de formalisation de cette dernière ont exigé de réexaminer la notion de collocation à la lumière d'approches linguistiques plus formelles. Actuellement, la phraséologie collocationnelle fait l'objet d'une littérature abondante.

Il existe deux grandes conceptions des collocations, correspondant à deux traditions linguistiques différentes. D'une part, la tradition contextualiste anglaise (cf. Firth, 1957b ; Halliday et Hasan, 1976 ; Sinclair, 1991 ; G. Williams, 2003) conçoit la notion de collocation dans une perspective quantitative, c'est-à-dire elle relève de la description statistique de corpus. Sous l'influence de Firth (1957b), la linguistique

¹⁹⁶ Nous avons déjà avancé la notion de *série phraséologique* dans la section 1. *Bref aperçu historique de la phraséologie*, dans ce même chapitre.

¹⁹⁷ Legallois (2012, p. 34-36) a trouvé des références plus anciennes de l'acception moderne de *collocation* dans une citation de 1750 pour illustrer l'entrée *collocation* dans l'*Oxford English Dictionary* (2^e éd.) et chez Jespersen (1917, 1922). Legallois attire l'attention sur le fait que ces emplois de *collocation* sont étroitement liés à la *colligation*.

contextualiste anglaise fait des collocations l'objet central de ses modèles linguistiques¹⁹⁸. Halliday (1961) propose une définition des collocations en termes de probabilité et de fréquence : « Collocation is the syntagmatic association of lexical items, quantifiable, textually, as the probability that there will occur, at n removes (a distance of n lexical items) from an item x, the items a, b, c... » (p. 276). Plus tard, Halliday et Hasan (1976) mettent l'accent sur l'importance de l'environnement textuel, car selon ces auteurs les collocations contribuent à la cohésion textuelle :

Without our being aware of it, each occurrence of a lexical item carries with it its own textual history, a particular collocational environment that has been built up in the course of the creation of the text and that will provide the context within which the item will be incarnated on this particular occasion. (p. 289)

Ainsi, l'unité lexicale tire sa signification du contexte linguistique et en même temps exerce une influence sur celui-ci pour créer l'environnement textuel.

Dans la même veine, Sinclair ne propose pas une définition différente de la notion de collocation, essentiellement définie, comme nous venons de le voir, par des critères statistiques : « the co-occurrence of words with no more than four intervening words » (Sinclair, 2004, p. 141). Il soutient l'idée d'un principe phraséologique de la langue (*idiom principle*), selon lequel les collocations sont des expressions récurrentes ayant une importance prépondérante dans le système de la langue (Sinclair, 1991).

Selon Evert (2008), « collocation in this Firthian sense can also be interpreted as empirical statements about the predictability of word combinations: they quantify the “mutual expectancy” (Firth 1957a, 181) between words and the statistical influence a word exerts on its neighbourhood » (Evert, 2008, p. 2113).

En somme, l'étude des collocations dans la tradition contextualiste anglaise est étroitement liée à la linguistique de corpus. Dans cette perspective, les collocations sont identifiées dans les corpus selon la puissance des liens mesurés statistiquement et elles

¹⁹⁸ Rappelons qu'il ne faut pas voir le contextualisme anglais comme une école unique, mais comme un courant pluriel (l'école de Londres, l'école de Birmingham...) fortement influencé par les idées de Firth.

sont donc conçues comme des combinaisons d'unités lexicales dont la coapparition est arbitraire mais statistiquement significative, et qui contribuent à la cohésion du texte¹⁹⁹.

Néanmoins, chaque méthode de mesure statistique adoptée pour relever les collocations ne semble appropriée que pour une application donnée et son choix dépend de nombreux paramètres tels que la taille du corpus, la fréquence des associations ciblées ou leur degré de figement (Kraif et Tutin, 2020). Ainsi, Clear (1993) note que :

The t-score statistic, by identifying frequent and very reliable collocations, offers the lexicographer a semantic profile of the node word and a set of particular fixed phrases, grammatical frames and typical stereotypical combinations. (p. 282)

tandis que

The MI [mutual information] is best consulted for information about pairs which, though not likely to be typical of the usage of the node word, will be strongly associated and tend to form idioms, proverbs and technical phrases. (p. 282)

Cela implique que la définition de collocation varie en fonction de la mesure statistique appliquée²⁰⁰.

D'autre part, l'approche européenne continentale des collocations, essentiellement qualitative, est héréditaire de la notion de *groupement usuel* (ou *série phraséologique*) de Bally (1909/1951) et prend naissance dans les études théoriques en lexicologie et sémantique lexicale (Cruse, 1986 ; Heid, 1994) et en lexicographie théorique (Hausmann, 1979, 1989, 1997 ; Mel'čuk et al., 1984, 1988, 1992, 1999). Cette notion de collocation est plus étroite et formelle que celle de l'école britannique :

¹⁹⁹ Les approches à la combinatoire de l'anglais sont beaucoup plus amples. La conception de l'école contextualiste, dans laquelle « collocation in its purest sense [...] recognises only the lexical co-occurrence of words » (Sinclair, 1991, p. 170), est le point de départ d'une tradition d'études de phraséologie collocationnelle qui va évoluer sous l'influence des théories sémantiques et cognitives. Nous tenons à mentionner, à titre d'exemple, Cowie (1981) dans une perspective sémantique, Benson (1985) dans une perspective sémantique et fonctionnelle, ou Hoey (1991, 2005) dans une perspective psycholinguistique. Pour une discussion sur la notion de collocation dans la tradition anglaise, voir Léon, 2007 ; G. Williams, 2003.

²⁰⁰ Il existe de nombreuses méthodes de mesure : MI (Church et Hanks, 1990), rapport de vraisemblance (Dunning, 1993), Chi2, t-score, z-score (Evert, 2008 ; Pecina et Schlesinger, 2006) ou le test de Fischer (Bestgen, 2017). Pour une révision assez complète mais non exhaustive des méthodes de mesure statistique z-score, t-score, information mutuelle, voir Daille (1994).

une collocation est une cooccurrence lexicale restreinte entre deux éléments linguistiques entretenant une relation syntaxique (cf. Hausmann, 1979, 1989, 1997 ; Grossmann et Tutin, 2003 ; Hausmann et Blumenthal, 2006 ; Anscombe et Mejri, 2011). Hausmann et Blumenthal (2006) proposent une explication très claire introduisant des caractéristiques essentielles :

Nous pouvons décrire la **collocation**²⁰¹ (exemples : *passer un examen, un célibataire endurci, grièvement blessé, une bouffée de colère*) comme la combinaison phraséologique (codée en langue) d'une **base** (*examen, célibataire, blessé, colère*) et d'un **collocatif** (*passer, endurci, grièvement, bouffée*). La base est un mot (plus précisément l'acception d'un mot, appelée aussi « lexie ») que le locuteur choisit librement parce qu'il est définissable, traduisible et apprenable sans le collocatif. Le collocatif est un mot (ou l'acception d'un mot) que le locuteur sélectionne en fonction de la base parce qu'il n'est pas définissable, traduisible ou apprenable sans la base. (p. 4)

Ainsi, dans la tradition européenne le centre d'intérêt n'est plus dans la fréquence, mais dans les contraintes sur l'utilisation du lexique, car il s'agit d'une propriété du lexique des langues. Par ailleurs, cette notion de collocation est axée sur la relation binaire (orientée et asymétrique) entre la base et le collocatif. Ce dernier, étant un élément non autonome et synsémantique, serait sélectionné par l'autre élément de cette combinaison syntagmatique (Tutin, 2013). C'est cette conception de la collocation que nous adoptons dans notre travail de recherche et que nous allons développer dans ce qui suit.

Néanmoins, dans les recherches actuelles, ces deux conceptions²⁰² sont plus complémentaires qu'antagonistes. D'une part, elles partagent l'approche descriptive du phénomène. D'autre part, les chercheurs de ces deux visions utilisent de vastes corpus informatisés. En effet, même les spécialistes de l'approche qualitative continentale appliquent le dépouillement automatique à leurs corpus soit pour se faire une première idée du contexte linguistique du mot de base, soit pour vérifier de manière empirique les

²⁰¹ Nous avons respecté les caractères gras du texte original.

²⁰² Pour une présentation et discussion autour de la notion de collocation dans ces deux approches, voir Alonso Ramos (1994-1995, p. 9-20).

résultats d'une recherche théorique sur la combinatoire restreinte²⁰³. Comme nous l'avons déjà signalé, le développement d'outils lexicométriques et de corpus numérisés a révolutionné la recherche linguistique (cf. Hausmann et Blumenthal, 2006).

Selon G. Williams (2001), « de par sa nature, la collocation demeure un concept difficilement formalisé, aucune définition ne satisfait tout le monde » (p. 9). Il existe de nombreuses définitions de la notion de collocation plutôt hétérogènes. Elles permettent de caractériser les collocations prototypiques, mais elles sont difficilement applicables à des cas « frontaliers » entre le préconstruit et le libre. D'ailleurs, Bosque (2001b) parle de la « inestabilidad del concepto de colocación » (p. 18). Nous préférons donc définir les collocations à partir de leurs propriétés.

4.2. Propriétés des collocations

Nous allons présenter les cinq critères retenus par Tutin et Grossmann (2002, p. 10-11) en se basant sur les définitions de la notion de collocation des textes fondateurs (Cruse, 1986 ; Hausmann, 1989 ; Mel'čuk, 1998). Nous les avons mis à jour à la lumière d'autres études plus récentes et que nous considérons importantes. Ces cinq paramètres qui nous aideront à mieux définir la notion de collocation sont : le caractère arbitraire (ou non prédictibilité), la compositionnalité et transparence sémantiques, le caractère binaire, la dissymétrie des composants de la collocation et la cooccurrence lexicale restreinte. Nous verrons aussi que ces caractéristiques ne concernent pas toutes les collocations ; par exemple, le sens de *peur bleue* n'est pas transparent, *l'âne braît* est complètement prédictible, ou *il pleut à verse* n'est pas composé de deux mots. Nous essayerons de démontrer que ces expressions qui apparemment transgressent les critères exposés ci-dessous sont des collocations. Toutefois, selon Tutin et Grossmann (2002) seulement le caractère binaire, la dissymétrie et la sélection lexicale contrainte sont des conditions nécessaires²⁰⁴.

²⁰³ D'ailleurs, une étude empirique menée par Kraif et Tutin (2020) a permis de constater qu'il existe une adéquation entre le repérage de collocations fait manuellement par un groupe de linguistes spécialistes et celui issu du comptage des fréquences d'occurrence et de cooccurrence purement quantitatif.

²⁰⁴ Dans une publication ultérieure, Tutin (2013, p. 49-50) redéfinit le caractère binaire et la dissymétrie. Ainsi, elle considère que la propriété de binarité recouvre deux aspects : la binarité catégorielle

4.2.1. Le caractère arbitraire de la collocation

L'une des premières caractéristiques qu'on attribue généralement aux collocations est la non prédictibilité du collocatif ; par exemple, le sens 'intense' est exprimé moyennant le collocatif *torrentielle* dans *pluie torrentielle*, mais cette cooccurrence semble bien moins naturelle dans *?précipitations torrentielles*²⁰⁵ ; dans le cadre de notre recherche, nous pouvons observer, par exemple le cas du verbe support²⁰⁶ *hacer* (*un préstamo / *una ayuda*). Une autre preuve du caractère imprévisible des collocations est la difficulté de faire une traduction littérale des collocatifs : *dar* (*un paso / un paseo*) ; *faire* (*un pas / une promenade*).

Néanmoins, certaines régularités syntaxiques et sémantiques ont été observées. En ce qui concerne la syntaxe, on a constaté l'existence d'une sorte de schémas syntaxiques (cf. M. Gross, 1988 ; Hausmann, 1989, 1998). Dans le cadre du lexique-grammaire on a mené de nombreuses recherches autour des constructions à verbe support (cf. Guillet, 1993 ; G. Gross, 1993b ; G. Gross et Vivès, 1986 ; M. Gross, 1981 ; 1993 ; Ibrahim, 1996 ; entre autres). M. Gross (1988) a démontré que le fonctionnement syntaxique de ces constructions est souvent régulier²⁰⁷. Par ailleurs, González Rey (2021, p. 137) a constaté la présence de structures qu'elle appelle des « moules reproducteurs », car ils permettent de faire apparaître de nouvelles collocations à partir d'expressions préexistantes dans la langue.

Sur le plan sémantique, de nombreux linguistes ont cherché à expliquer la cooccurrence entre le nom prédicatif et le verbe support par des critères sémantiques. Par exemple, Wierzbicka, (1982) dans son étude des expressions du type *to have a drink*, a trouvé une invariante sémantique du verbe *to have* qui fait que ce soit ce verbe qui se combine avec certains noms. Par ailleurs, Mendivil (1999) soutient que les noms

(association de deux constituants) et la binarité fonctionnelle (fonctionnement dissymétrique, antagonique entre les deux composants de la collocation).

²⁰⁵ Exemples empruntés à Tutin et Grossmann (2002, p. 10).

²⁰⁶ Nous considérons que les constructions à verbe support sont un type de collocation.

²⁰⁷ M. Gross n'isole pas les collocations des expressions figées.

prédicatifs sélectionnent leurs verbes supports en fonction de leur structure aspectuelle²⁰⁸.

Dans le cadre de la théorie Sens-Texte on a trouvé des régularités au niveau des relations sémantiques qui relient la base et le collocatif (cf. Kahane et Polguère, 2001 ; Mel'čuk et al. 1984, 1988, 1992, 1999 ; Mel'čuk, 1996, 2003a, 2003b ; Wanner, 1996). Certains sens de la langue ont tendance à s'exprimer de manière collocationnelle (Uzcanga Vivar, 2020a) ; par exemple, le sens 'intense' peut être exprimé par des collocatifs différents en fonction de la base nominale : *odio mortal, esfuerzo ímprobo, necesidad imperiosa*, etc. ; ou le sens 'causatif' peut apparaître sous la forme de verbes tels que *donner (donner faim)* ou *faire (faire peur)*. Ces contenus collocationnels universaux ont été identifiés et modélisés grâce au système formel des *fonctions lexicales*²⁰⁹. Les collocations s'inscrivent donc dans des relations sémantiques récurrentes, mais la sélection lexicale reste imprévisible.

Selon Tutin et Grossmann (2002), « les régularités sémantiques entre la base et le collocatif semblent relativement systématiques dans certains domaines de l'expérience, bien structurés, comme c'est souvent le cas en langue de spécialité » (p. 23). En effet, des études menées dans des domaines tels que l'informatique (cf. L'Homme, 1998 ; Meyer et Mackintosh, 1994 ;), l'aéronautique (L'Homme et Bertrand, 2000), la philosophie (L'Homme et Bertrand, 2000) ou le lexique académique spécifique des articles académiques (Cavalla et Grossmann, 2005) ont permis de démontrer que certaines associations lexicales semblent répondre à des patrons sémantiques réguliers.

²⁰⁸ L'auteur admet que cette solution de la structure aspectuelle ne fonctionne pas toujours ; par exemple, certains noms ne dénotant pas de dynamisme (comme *reposo*) sélectionnent le verbe *hacer* comme verbe support (Mendivil, 1999, p. 274-282).

²⁰⁹ La notion de fonction lexicale est une fonction au sens mathématique du terme : « Une fonction lexicale f appliquée à l'unité lexicale L – ce qui est noté $f(L)$ – fournit pour le sens ' σ^f ' associé à f un ensemble d'expressions alternatives de ce sens dont la sélection est contrainte par L » (Mel'čuk et Polguère, 2021, p. 77). Dans la linguistique Sens-Texte on soutient que les fonctions lexicales ne sont pas une simple création théorique, mais elles « existent » véritablement dans les langues. Pour approfondir sur la notion de fonction lexicale, voir Mel'čuk (1996).

Bosque (2001b) attire l'attention sur le fait que le regroupement des bases qui sélectionnent les mêmes collocatifs n'a rien de nouveau car Hausmann (1989, 1998), à la toute naissance de la phraséologie collocationnelle, avait déjà avancé la notion de *champ collocationnel*. De nombreux auteurs ont constaté l'existence de paradigmes lexicaux de bases de collocations, et d'une sorte de « synonymie restreinte » entre les bases appartenant au même paradigme. D'ailleurs, Mel'čuk et Wanner (1996) ont démontré qu'il existe une relation entre certains traits sémantiques des noms d'émotions et les collocatifs verbaux qu'ils sélectionnent. Dans le dictionnaire *Redes* (Bosque, 2004), les données de combinatoire de chaque prédicat constituant une entrée sont regroupées en classes lexicosémantiques.

Dans cette même veine, dans le cadre de la théorie des classes d'objets (G. Gross, 1996a, 2012), on cherche à établir des classes syntacticosémantiques de noms prédicatifs qui sont actualisés par les mêmes verbes supports²¹⁰. Alors, il serait possible de prédire le verbe support à partir de la classe sémantique du nom prédicatif²¹¹.

Si nous ne refusons pas l'aspect arbitraire de la cooccurrence lexicale, nous ne négligeons pas pour autant l'existence de certaines voies de généralisation syntaxique et sémantique qui limiteraient le caractère imprévisible des collocations et dont la description systématique peut faciliter, par exemple, leur traitement automatique.

4.2.2. Le sens compositionnel ou la transparence de la collocation

Il est généralement admis que le sens des collocations est compositionnel ou semi-compositionnel (cf. entre autres, Cruse, 1986 ; Hausmann, 1989 ; Koike, 1996 ; Zuluaga, 2002). La base de la collocation est librement sélectionnée par le locuteur en fonction de son sens ; par contre, la sélection du collocatif est contrainte par la base à laquelle il devrait se combiner et il possède un sens souvent très vague ou presque vide.

De nombreux linguistes ont observé une certaine spécialisation du sens du collocatif en contexte, c'est-à-dire le sens de celui-ci s'adapte en fonction de la base qu'il accompagne (cf. Koike, 2002 ; Buvet, 2013b). On parle aussi d'une valeur figurée,

²¹⁰ Dans la théorie des classes d'objets on n'isole pas a priori les collocations des expressions figées.

²¹¹ Nous présenterons ce modèle de description linguistique et les notions de construction à verbe support et de classe de prédicats dans le chapitre III.

voire métaphorique du collocatif, comme dans *briser des chaussures neuves* ou *domar unos zapatos nuevos* (González Rey, 2021²¹²).

Selon Polguère (2015b), toutes les collocations sont à 100 % sémantiquement compositionnelles. Ce linguiste soutient que les collocations ne sont pas constituées seulement de deux lexèmes, mais elles possèdent aussi une structure syntaxique, puisqu'elles constituent des syntagmes. Nous savons que le sens d'une collocation résulte du sens de la base auquel s'ajoute le sens du collocatif. Cependant, en ce qui concerne le sens de ce dernier,

Le sens 'S_{colloc}'²¹³ n'est pas le sens d'un lexème donné de la langue, mais le sens d'un sous-arbre syntaxique donné, en tant qu'il s'arrime à un lexème donné dans le cadre de l'activation d'une règle de combinatoire propre à ce dernier, c'est-à-dire d'une règle lexicale du français. (Polguère, 2015b, p. 264)

Par exemple, dans le cas de la collocation *verre blanc* ('verre de qualité ordinaire')²¹⁴, le collocatif *blanc* porte le sens 'de qualité ordinaire' auprès de la lexie *verre*. Ainsi, dans ce cas, « le sens de l'adjonction du syntagme vient strictement (i) du sens de la lexie BLANC_{Adj} et (ii) du contenu exprimé par la dépendance syntaxique épithète » (Polguère, 1990, p. 126-127, cité dans Polguère, 2015b, p. 265). Dans cette perspective, on refuse la thèse selon laquelle le collocatif prendrait un sens particulier en contexte. On considère que c'est le sous-arbre syntaxique du collocatif²¹⁵ qui porte un sens particulier du fait de l'intervention d'une règle lexicale du français associée à une base donnée. De cette façon, on peut affirmer que l'équation représentant la compositionnalité

$$'S' = 'S_{base}' + 'S_{colloc}'$$

²¹² Cette linguiste évoque même une composante stylistique mise au service d'une expressivité plus efficace. Gréciano (1997) aussi, elle avait déjà étudié l'utilisation de la métaphore dans la construction de collocations terminologiques.

²¹³ 'S_{colloc}' est le sens du collocatif selon les conventions d'écriture propres à la théorie linguistique Sens-Texte.

²¹⁴ Exemple emprunté à Polguère (2015b, p. 263-265).

²¹⁵ Selon Polguère (2015b), « les signes linguistiques se combinent en structures et non en chaînes, par simple concaténation » (p. 265).

est respectée dans le cas de la collocation *verre blanc*, comme d'ailleurs c'est le cas de toutes les collocations (Polguère, 2015b).

En ce qui concerne la transparence, les collocations sont facilement comprises par le locuteur (même non natif) car leur sens peut être inféré. Par contre, un locuteur non natif serait incapable de produire des collocations qu'il n'a pas mémorisées au préalable.

Le critère sémantique est souvent difficile à manier (cf. Nunberg et al., 1994 ; Svensson, 2004) ; citons les célèbres exemples *peur bleue* ou *colère noire* qui font preuve d'opacité sémantique. Ainsi, selon Tutin et Grossmann (2002) trois cas de figure sont à envisager²¹⁶:

1. Le sens du collocatif en cooccurrence avec L est différent du sens qu'il a en dehors de cette association (*peur bleue, colère noire*). Dans ce cas (**collocations opaques**²¹⁷), l'association paraît arbitraire et non transparente sur le plan sémantique.
2. Le collocatif peut avoir un sens interprétable (la collocation est « transparente ») mais apparaître difficilement prédictible (*faim de loup, brouillard à couper au couteau, grièvement blessé, avoir faim*). Le collocatif n'a pas de statut lexical (c'est un syntagme comme *de loup* ou *à couper au couteau*) ou bien a un sens décodable en cooccurrence avec la base (*grièvement, avoir*). On parlera de **collocations transparentes**.
3. Le collocatif inclut le sens de la base ou a un sens très générique (*nez aquilin, l'âne brait, grande tristesse*). L'association est motivée et transparente. Le collocatif est à base unique (on définirait le collocatif à l'aide de la base, comme dans le cas de *aquilin*) ou bien au contraire porte sur un paradigme qui semble pouvoir être défini à l'aide des traits sémantiques. On parlera de **collocations régulières**. (p. 12-13)

Alors, les collocations transparentes constitueraient le cas prototypique de collocation : elles sont à priori peu prédictibles mais décodables par un locuteur non natif. Les collocations opaques resteraient proches des expressions perçues comme figées : le collocatif n'est pas prédictible et le sens est difficilement décodable. Enfin, les collocations régulières sont proches des expressions libres : décodables et plutôt prédictibles, car elles semblent présenter des affinités sémantiques entre les composants.

²¹⁶ Les auteurs signalent que cette esquisse de typologie est une reformulation de la proposition de Mel'čuk (1998).

²¹⁷ Nous avons respecté les caractères gras du texte original.

La transparence ne caractérise donc pas l'ensemble des collocations. Néanmoins, une partie du sens dénoté peut être bien comprise car, comme nous venons de le signaler, la base est librement sélectionnée par le locuteur en fonction de son sens. Autrement dit, dans *peur bleue*, peut-être un locuteur non natif ne comprendra-t-il comment elle est cette peur, mais il aura compris *peur*. Finalement, nous pourrions considérer que l'idiomaticité présente dans une grande partie des collocations est graduelle.

4.2.3. Le caractère binaire de la collocation

Dans la tradition européenne continentale on tend à décrire les collocations comme des expressions binaires. Nous pouvons l'observer dans les exemples suivants :

- « on appellera collocation la combinaison de deux mots » (Hausmann, 1989, p. 1010) ;
- « collocations are combinations of exactly two lexemes » (Heid, 1994, p. 228) ;
- « una colocación es una expresión semifraseológica formada por dos unidades léxicas » (Alonso Ramos, 2004b, p. 20) ;
- « une collocation est un syntagme *AB* (ou *BA*), formé des lexies L_A et L_B » (Polguère, 2008a, p. 165).

Pourtant, depuis une vingtaine d'années, le caractère binaire des collocations a été remis en question (cf. Bartsch, 2004 ; Siepmann, 2006). En effet, la collocation peut bien dépasser les deux termes (*avoir une peur bleue*, *dormir comme un loir*, *avoir un geste déplacé*). Ces collocations *n-aires* (ternaires, dans le cas des exemples précédents) sont généralement le résultat de combinaisons de collocations, d'ailleurs assez courantes dans la langue²¹⁸. Par exemple, *avoir une peur bleue* comporte une superposition deux collocations indépendantes *avoir peur* et *peur bleue* (Tutin, 2013, p. 52-53). Quant à *dormir comme un loir*, il s'agit d'une collocation dont le collocatif (*comme un loir*) est

²¹⁸ Il s'agit également d'un phénomène très étudié ; par exemple, Hausmann (2007) les appelle « chaînes collocationnelles », et Alonso Ramos et Wanner (2007), « collocations chaînes ».

une locution²¹⁹. On propose alors une extension de la notion de collocation vers des suites plus longues (cf. Tutin, 2008a).

Plutôt qu'une combinaison de deux mots²²⁰, aujourd'hui on s'accorde à dire que les collocations ont un fonctionnement fondamentalement binaire ou une structure binaire : « las colocaciones son grupos binarios compuestos por una base y el colocativo, pero en donde cada uno de ellos puede ser a su vez binario » (Alonso Ramos, 2004b, p. 55-56). Ces constructions *n-aires* ne remettent donc pas en question le caractère binaire des collocations.

4.2.4. La dissymétrie des composants de la collocation

Cette caractéristique reprend les autres caractéristiques décrites ci-dessus et, en même temps, en est la conséquence. Nous venons d'évoquer la structure binaire des collocations. Ainsi, les collocations sont constituées de deux parties dont le statut est inégal. Nous avons déjà évoqué cette idée : seule la base est choisie librement par le locuteur, le collocatif étant sélectionné par la base ; seule la base garde son sens habituel, le sens du collocatif est général ou vague et il est modulé, en partie, en fonction de la base avec laquelle il se combine (p. ex., l'unité lexicale *bleu* prend le sens intensif en cooccurrence *peur*, mais pas avec son synonyme *crainte* : *une peur bleue* ; **une crainte bleue*). C'est donc la base qui contrôle la collocation : « The meaning of the base is always the semantic pivot of the collocation » (Mel'čuk, 2012, p. 39).

Cependant, parfois il s'avère difficile de distinguer l'élément qui fait office de base de celui qui serait le collocatif (cf. Bartsch, 2004 ; González Rey, 2002, 2021 ; Hausmann, 2007 ; Siepmann, 2006 ; Tutin, 2013). Tutin (2013) souligne qu'il s'agit notamment des structures V prep N (*rougir de honte, trembler de peur*) et N prep N (*meute de loups, crise d'angoisse*). Elle soutient que ces structures sont ambiguës du

²¹⁹ D'ailleurs, le verbe dormir sélectionne de nombreuses locutions pour exprimer l'intensité : *à poings fermés, comme un bébé, comme un bienheureux, comme une marmotte, comme un sabot, comme une souche, sur ses deux oreilles*. Parfois, ce genre de constructions ont été décrites comme des locutions verbales (cf. Penadés, 1999, p. 47). Nous ne sommes pas d'accord avec cette dernière analyse.

²²⁰ Quant à la terminologie utilisée dans les bouts de définitions cités plus haut (dans cette section), il est nécessaire de souligner que le terme *mot* ne désigne pas nécessairement les mêmes éléments linguistiques qu'*unité lexicale, lexie* ou *lexème*. En effet, la tradition n'a jamais vraiment réussi à fournir une définition précise du mot (cf. Martinet, 1966).

point de vue sémantique et syntaxique, car elles se prêtent à deux interprétations : le premier élément (le verbe dans le cas *rougir de honte*, et le premier nom dans le cas *meute de loups*) peut être interprété comme un modifieur, parfois désémantisé, ou il peut être interprété comme un mot plein, contrôlant la construction²²¹. Malgré ces difficultés la structure dissymétrique des collocations n'est pas véritablement remise en question, puisqu'il y a toujours un pivot qui contrôle la collocation.

4.2.5. La cooccurrence lexicale restreinte de la collocation

Souvent, pour distinguer la collocation de la simple cooccurrence on s'appuie sur la dichotomie langue/parole : la collocation est inscrite dans le système de la langue, en revanche la cooccurrence appartient au discours, et, par conséquence, elle relève de la combinatoire libre²²². Selon Buvet (2013b, p. 172-173), ce raisonnement est insuffisant, car il est nécessaire qu'une construction syntagmatique soit identifiée comme cooccurrence pour avoir le statut de collocation ; pourtant toutes les cooccurrences ne sont pas linguistiquement pertinentes. Seulement les cooccurrences linguistiquement fondées peuvent constituer des collocations. Ainsi, Neveu (2011) définit la cooccurrence comme « la coprésence dans un énoncé de deux ou plusieurs unités linguistiques (morphèmes, mots, syntagmes, propositions), mais également leur compatibilité » (p. 100). D'autre part, il définit la collocation comme :

²²¹ Tutin (2013, p. 56) argüe aussi que les propriétés de binarité et de dissymétrie semblent être prototypiques pour la fabrication d'un modèle d'expressions. Rappelons que González Rey (2002, 2021) évoque aussi l'existence de « moules reproducteurs » notamment prolifiques en langues de spécialité.

²²² Il existe un certain dissensus parmi la communauté linguistique sur l'appartenance des collocations au domaine de la langue ou de la parole. Par exemple, selon Mejri (1997, 2008a) la collocation est un phénomène de nature fondamentalement discursive, puisque ses contraintes se donnent dans l'emploi. Pour cet auteur, la collocation n'est pas un préconstruit linguistique. Ainsi, il soutient que « la collocation renvoie à toutes les combinaisons syntagmatiques de la combinatoire libre dont les éléments sont appropriés les uns aux autres. Plus l'appropriation est grande, plus les restrictions sont importantes, plus on s'approche des séquences figées » (Mejri, 2008a, p. 197). Cette perception s'inscrit donc dans un continuum allant du plus libre au plus figé (cf. Mejri, 2008a). Nous avons déjà abordé le caractère gradable du figement. Nous reviendrons ultérieurement sur ce caractère gradable et son rapport à la notion de collocation.

une cooccurrence conventionnelle, résultant d'une forte contrainte sémantique de sélection qui se manifeste dans la valence d'une unité lexicale, et qui a pour effet de restreindre la compatibilité des mots avec l'unité en question. (Neveu, 2011, p. 85)

Dans le cas de la cooccurrence, le linguiste ne fait pas allusion à la restriction de sélection, puisqu'il est juste question de coapparition dans le discours ; tandis que pour la collocation, la solidarité syntagmatique entre les constituants s'avère nécessaire. Autrement dit, une collocation est une cooccurrence, mais toute cooccurrence n'est pas pour autant une collocation.

Il faut donc comprendre la cooccurrence restreinte comme une sélection lexicale. Pour Mel'čuk (2013), une collocation est un phrasème lexical semi-contraint : « une de ses composantes est sélectionnée librement, juste pour son sens ; c'est l'autre qui doit être choisie en fonction du sens à exprimer et de la première composante » (p. 138). Rappelons l'exemple déjà cité des (*pluies, ?précipitations*) *torrentielles*, ; ou les célèbres *gravement (malade, *blessé), grièvement (*malade / blessé)*. Nous pouvons observer que pour exprimer le sens 'intense', la base de ces collocations (les noms) opère des restrictions de cooccurrence sur les collocatifs qui l'accompagnent. Alonso Ramos (2004b, p. 56-57) souligne que même dans le cas d'adjectifs très productifs comme l'espagnol *gran/grande*, il existe des restrictions de cooccurrence lexicale : *gran (*hambre, *fiebre, *hipótesis, *objeción)*. Pour l'autrice, ces impossibilités sont la preuve qu'une expression telle que *gran dolor*, qui est prédictible et complètement transparente, est une collocation, puisque le sens 'intense' est systématiquement exprimé de manière phraséologique²²³.

Néanmoins, des fois il s'avère difficile de définir la limite entre la véritable collocation et la simple cooccurrence de mots. En effet, selon Hausmann (1985) il existe un continuum entre les cooccurrences banales et les collocations. C'est le cas, par exemple, des routines sémanticorhétoriques surreprésentés dans un discours de spécialité donné (p. ex. *il est frappant de constater que*). Selon Kraif et Tutin (2020), « aucun test ne permet vraiment de rattacher ces expressions aux catégories du figement

²²³ Rappelons que la linguiste Alonso Ramos mène ses recherches dans le cadre de la théorie Sens-Texte. Ce sens 'intense' constitue l'une des fonctions lexicales standards (Magn) décrites dans cette approche linguistique (cf. Kahane, 2003a ; Kahane et Polguère, 2001 ; Mel'čuk, 1996, 2003a, 2003b ; Wanner, 1996 ; Žolkovskij et Mel'čuk, 1970).

syntaxique ou de l'opacité sémantique. Simplement, elles sont préfabriquées dans certains discours » (p. 99). De plus, si nous regardons les langues comme des processus, et pas comme des produits, nous pouvons bien comprendre que ces cooccurrences lexicales habituelles puissent changer au cours du temps (G. Williams, 2001). La collocation devrait donc être conçue comme un phénomène dynamique (cf. Mejri, 2011a ; G. Williams, 2001). La propriété de la cooccurrence lexicale, bien qu'elle soit nécessaire, elle reste un critère flou.

4.3. Le statut phraséologique des collocations

Nous savons que traditionnellement la recherche en phraséologie s'est organisée en trois grands blocs : dans le bloc oriental (URSS et d'autres pays de l'est) s'impose une conception large de la phraséologie ; dans le bloc occidental (structuralisme européen, notamment l'École de Genève, et ses développements ultérieurs) on privilégie une conception restreinte de la phraséologie, dont le centre est occupé par des séquences figées et idiomatiques, et les collocations seraient reléguées à la périphérie ; finalement, les linguistes anglo-américains ont pratiquement banni les collocations (cf. Corpas Pastor, 2003, p. 68). L'échange scientifique entre les spécialistes et le développement de la linguistique de corpus ont fait que ces différentes postures se rapprochent. Néanmoins, le statut phraséologique des collocations continue à soulever beaucoup d'interrogations.

Certains linguistes considèrent que les collocations relèvent de la combinatoire libre ; pour d'autres, c'est de la combinatoire figée. Dans le premier cas, la liberté de sélection de la base par son sémantisme, l'existence de variations comme l'insertion ou la commutation (*avoir faim, avoir très faim, avoir une faim de loup, avoir une de ces faims*)²²⁴, c'est-à-dire des libertés de combinatoire, etc. sont la preuve d'une combinatoire libre. Par contre, dans la perspective de la combinatoire figée, on focalise sur l'idée de contrainte (syntaxique et sémantique). Les collocations admettent des modifications formelles, mais moins que les syntagmes libres (Polguère, 2011b). D'ailleurs, nous avons montré que les unités phraséologiques et les collocations partagent certaines propriétés comme la polylexicalité, la spécialisation sémantique, un certain degré de figement (ou de contrainte), la fréquence de cooccurrence ou

²²⁴ Exemples empruntés à Mejri (2008a, p. 89).

l'institutionnalisation (Corpas Pastor, 1998, p. 167). Les collocations pourraient donc se trouver au croisement entre la combinatoire libre et la combinatoire figée. D'ailleurs, Mejri (2011a) affirme : « à voir de plus près dans les données relatives aux unités phraséologiques et aux collocations, on se rend compte qu'il y a une très grande interpénétration entre la combinatoire libre et la combinatoire figée dans les deux cas » (p. 91).

Mejri (2011a) introduit la notion d'*hybridité*. Pour cet auteur, le figement est un processus, et les collocations sont des cas particuliers d'unités phraséologiques en cours de figement : « l'appropriation²²⁵ [...] se situe à la proximité du figement, dans ce sens qu'elle occupe une zone de pré-figement » (Mejri, 2011b, p. 69). Ces sortes de combinaisons seraient comme du « préfigement » ou le premier pas vers le figement car elles réduisent au maximum le paradigme des mots qui lui sont syntagmatiquement associés. Ainsi, elles présentent des contraintes sémantiques, des blocages syntaxiques ou des limitations d'emplois pragmatiques²²⁶. Le fait que les unités phraséologiques présentent une certaine régularité syntaxique et sémantique, c'est-à-dire conformément aux règles de la combinatoire libre, ne nuit pas au statut phraséologique de ces constructions. Cet auteur considère donc que « grâce à ce parallélisme, on pourrait attribuer aux collocations, tout comme on l'a fait pour les séquences figées, le statut d'unités lexicales mais de nature un peu spéciale » (Mejri, 2011a, p. 91). Ainsi, pour ce linguiste, les collocations sont des combinaisons libres appropriées, appartenant donc à la combinatoire libre, mais faisant partie de la phraséologie, où elles représentent un degré limité de figement (Mejri, 2008a, 2011b)²²⁷. L'avantage de la notion d'unité

²²⁵ Dans cet article, Mejri utilise « appropriation » ou « emploi approprié » comme synonyme ou quasi-synonyme de collocation.

²²⁶ Mejri (2011a, p. 93) met en relief la part de figement pragmatique qu'on peut attribuer aux collocations. Il soutient que ces constructions sont étroitement liées au contexte d'emploi ; il cite comme exemple les collocations du type *à consommer avant* et *à consommer de préférence avant* utilisées pour des produits de consommation périssables, tandis qu'*à consommer avec modération* est utilisé dans un tout autre contexte, pour des produits dont la consommation excessive nuit la santé.

²²⁷ Mejri (2011b) distingue trois niveaux de combinatoire libre. Le premier niveau regroupe un petit nombre de prédicats dont les paradigmes d'arguments constituent des séries très ouvertes (p. ex. : *petit, regarder...*). Ces séquences donnent donc une impression de très grande liberté. Le deuxième niveau correspond à celui des classes restreintes. Ces restrictions produisent des interdépendances entre les

lexicale hybride réside dans le fait qu'elle permet de rendre compte de leur double appartenance au discours et à la langue, à la combinatoire libre et figée.

Dans cette même veine, pour Polguère (2008a, 2011b) les collocations ne sont pas des lexies (contrairement aux locutions), mais il s'agit tout de même de syntagmes préfabriqués par la langue. Cet auteur perçoit aussi la phraséologie comme une notion graduelle. Pour lui, la phraséologie est entière dans le cas des locutions et partielle dans le cas des collocations. Les collocations sont donc un type de phrasème ou un *semi-phrasème*²²⁸.

5. Conclusions

L'incursion historique dans le domaine de la phraséologie a mis en évidence l'importance progressive que cette discipline a prise au cours du XX^e et du XXI^e siècles et son essor croissant depuis l'avènement de l'informatique et le développement de la linguistique de corpus. Ces deux facteurs ont bouleversé les études en phraséologie, entraînant une extension du domaine dans son objet d'étude (les limites du phénomène phraséologique sont constamment redéfinies et élargies), ses approches (au croisement de nombreuses disciplines des sciences du langage telles que la psycholinguistique ou l'enseignement/apprentissage de langues) et sur ses méthodes (fortement déterminées par ces limites floues et ces approches plurielles).

La profusion d'approches et d'études a largement enrichi la discipline de la phraséologie, mais en même temps y a apporté un certain flou (terminologique et

prédicats et les paradigmes d'arguments sélectionnés, les uns faisant appel aux autres (p. ex. : N0<Hum> *ressemeler* N1<chaussures>). Plus un paradigme est réduit, plus on est proche du troisième niveau de combinatoire libre : celui de l'appropriation. Dans ce cas, à la restriction sémantique s'ajoutent des limitations plus restrictives sur le paradigme sélectionné (p. ex. : N0<Hum> *brasser* N1<bière> ; N0<Hum> *sabler* N1<champagne>). Selon Mejri, dans ce troisième niveau on n'est pas loin du figement. Notons que la notion de *construction libre* chez Mejri (2011) diffère de la notion chez Mel'čuk et Iordanskaja (2017) que nous avons présentée de manière succincte dans la section 3.2.4. *Le figement*.

²²⁸ Dans le cadre de la théorie Sens-Texte, souvent, au lieu d'utiliser le terme *collocation*, on préfère le terme *semi-phrasème*. D'une part, le préfixe *semi-* suggère que le figement n'est pas complet ; et d'autre part, *phrasème* indique qu'il s'agit bien d'un phénomène phraséologique.

conceptuel). Nous avons donc considéré nécessaire d'esquisser un portrait de notre conception des unités phraséologiques. L'analyse de leurs propriétés essentielles a fait la lumière sur le fonctionnement sémantique, syntaxique et pragmatique des diverses expressions phraséologiques. Nous tenons à souligner spécialement deux propriétés définitoires et transversales : la polyfactorialité et la gradualité. De plus, l'exploration de certaines classifications des unités phraséologiques nous a permis de retracer un parcours sur l'hétérogénéité du phénomène phraséologique pour ensuite pouvoir cerner (autant que possible) l'objet de notre étude : les collocations.

S'inscrivant dans un continuum entre la combinatoire libre et la combinatoire figée, la notion de collocation ne fait pas l'objet de consensus parmi la communauté de linguistes. L'examen des propriétés des collocations nous permet de les considérer des objets phraséologiques, ayant des particularités telles que la relation orientée et asymétrique où la base sélectionne le collocatif.

Nous pouvons maintenant exposer le cadre théorique dans lequel s'inscrit notre description de la combinatoire des prédicats nominaux des <aides financières>.

CHAPITRE IV : LE MODÈLE DES CLASSES D'OBJETS

Parmi les théories qui privilégient le lexique, il y en a qui accordent une place centrale à la sémantique ; d'autres qui abordent la description des unités lexicales du point de vue de la syntaxe (notamment le lexique-grammaire) ; et d'autres qui adoptent une position intermédiaire d'équilibre entre la sémantique et la syntaxe, comme c'est le cas du modèle des classes d'objets (cf. G. Gross, 1992, 1994a, 1995, 2012 ; Le Pesant et Mathieu-Colas, 1998). Il s'agit d'une méthode de description de la langue destinée au traitement automatique : elle permet de rendre compte d'une manière exhaustive et systématique des faits de langue, s'adaptant ainsi aux spécificités de l'outil informatique. C'est sur les bases théoriques de ce modèle que nous avons entrepris la description des noms prédicatifs des <aides financières>²²⁹.

Cette méthode de description de la langue est fondée sur l'observation des constructions syntaxiques et des combinaisons lexicales possibles. Le principe de base sous-jacent à cette approche est la grammaire transformationnelle de Harris (1971, 1976) dont le postulat veut que l'unité minimale d'analyse soit la phrase et non pas le signe linguistique (dans le sens saussurien)²³⁰. En effet, les mots isolés sont presque tous polysémiques, ils ne prennent leurs sens que dans le contexte de la phrase. Ainsi, c'est dans le cadre de la phrase que l'on va comprendre la différence entre :

(1) a. *Luc regarde un match de tennis.* ('voir')

b. *L'équipe de direction a déjà regardé tous les aspects négatifs de la situation.* ('considérer')

²²⁹ Par convention, dans le modèle des classes d'objets, le nom représentatif d'une classe d'objets est encadré de chevrons <>.

²³⁰ Selon Harris, les propriétés syntaxiques d'une unité lexicale doivent être analysées dans le cadre de la phrase simple, celle-ci étant constituée d'un opérateur (prédicat) et de ses objets (arguments : sujet grammatical et différents compléments). C'est l'opérateur qui sélectionne le nombre et la distribution de ses objets. Pour affiner le classement des opérateurs et distinguer les différents emplois d'une même forme, Harris les fait passer au tamis d'une série de transformations, observables toujours dans le cadre de la phrase.

c. *Cette décision ne te regarde pas.* ('concerner')

Ce point de départ offre de nombreux avantages pour le traitement automatique de la langue, notamment pour la désambiguïsation de la polysémie. Ce qui nous mène à affirmer que le mot prend son sens dans le cadre de la phrase. Il n'est donc pas possible d'étudier le lexique en dehors de la syntaxe²³¹.

Le modèle des classes d'objets est inspiré de la grammaire transformationnelle de Harris (1971, 1976) et des travaux en lexique-grammaire de M. Gross (1975, 1981)²³² développés dans le laboratoire LADL (Laboratoire d'Automatique Documentaire et Linguistique), et il a été mis au point par G. Gross au sein du laboratoire LLI (Laboratoire de Linguistique Informatique), ensuite appelé LDI (Lexiques, Dictionnaires et Informatique)²³³. Les classes d'objets permettent d'accorder à la sémantique une place plus importante qu'elle ne l'avait ni chez Harris ni chez M. Gross. En effet, le modèle des classes d'objets est une réponse au manque de précision des traits syntacticosémantiques avec lesquels on travaillait depuis des années, et une réponse aussi au caractère intuitif d'autres taxonomies existantes.

Pour que l'ordinateur soit capable de reconnaître et de générer des textes, il a besoin d'une description complète de la langue. Nous ne pouvons pas nous contenter des règles générales de grammaire, de morphologie et de syntaxe. Il est nécessaire de faire une fine description de toutes les propriétés linguistiques de chaque mot. Ces descriptions seront formatées en lexiques, d'où l'importance accordée à l'élaboration des dictionnaires électroniques. Il s'avère que deux des grands problèmes du traitement automatique sont des phénomènes linguistiques largement répandus dans les langues

²³¹ Le terme *syntaxe* est employé dans le sens d'ensemble de règles qui régissent la combinaison des unités lexicales dans le cadre de la phrase. La différence avec le terme *combinatoire*, tel que nous l'utilisons, est que ce dernier comporte, en plus de la syntaxe, des règles sémantiques, morphologiques, pragmatiques, etc., qui déterminent toutes l'emploi des unités lexicales dans le cadre de la phrase.

²³² Pour comprendre les principes généraux du lexique-grammaire, voir également Boons et al., 1976a, 1976b ; Leclère, 1989 ; et le numéro 59 de la revue *L'information grammaticale*, entièrement consacré au lexique-grammaire.

²³³ Actuellement, le laboratoire LLI est devenu le TTN (Texte, Théories numériques) où l'on trouve encore l'empreinte des recherches du modèle des classes d'objets.

naturelles : la polysémie et le figement. Nous allons montrer que les classes d'objets sont extrêmement utiles pour résoudre ces problèmes.

Dans ce chapitre nous tenterons d'expliquer les fondements théoriques du modèle des classes d'objets, sur lequel est fondée notre étude. Le postulat de ce modèle de description de la langue est le suivant : la phrase élémentaire est constituée d'un prédicat de premier ordre et de ses éventuels arguments (Buvet et al., 2005)²³⁴. Ainsi, les unités linguistiques peuvent avoir trois fonctions dans la phrase : prédicat, argument ou actualisateur. D'abord, nous nous attarderons sur la notion centrale de ce modèle théorique : le prédicat. Avant d'entamer la description de ces trois fonctions primaires, nous avons considéré nécessaire de définir la notion de phrase élémentaire car elle constitue notre unité minimale d'analyse. Ensuite, nous aborderons successivement la description des prédicats, des arguments et des actualisateurs. Enfin, nous survolerons la notion de phrase complexe, car les textes ne sont pas constitués que de phrases simples.

1. La notion de prédicat dans la théorie des classes d'objets

1.1. Les origines harrissiennes de la notion de prédicat

Pour comprendre la notion de *prédicat* et son utilité pour la description de la langue et pour le traitement automatique, nous considérons nécessaire de faire le point sur la notion de *phrase*, puisque c'est dans la phrase que le prédicat s'inscrit et prend un sens donné.

La littérature linguistique compte de nombreuses définitions de la phrase élémentaire. Traditionnellement, il existe deux grands types de représentation de la phrase. D'une part, les héritiers de la philosophie antique, notamment de la logique classique, décomposent la phrase élémentaire en *thème* ou *sujet* (ce dont on parle) et *rhème* ou *prédicat* (ce que l'on dit). Cette conception a été reprise par la grammaire générative en divisant la phrase en groupe nominal (GN) et groupe verbal (GV). D'autre part, l'autre versant provient de la philosophie analytique. Les études du mathématicien-linguiste Gottlob Frege (1984/1970) avaient conduit à remplacer la paire *sujet/prédicat*

²³⁴ Rappelons que ce postulat est hérité de la position de Harris.

par les notions de *fonction/variable*, considérant la phrase élémentaire comme un atome dont le noyau est un prédicat insaturé (*fonction*) qui sera complété par un ou plusieurs arguments nominaux (*variables*). On est donc très proche de la fonction mathématique, qui a besoin de l'attribution de valeurs à des arguments que la fonction introduira dans sa dépendance. On retrouve cette représentation atomique avec des variantes chez Tesnière (1959)²³⁵ sous la forme *verbe/actant*, chez Fillmore (1968) et son modèle casuel ou chez Harris (1976) avec le découpage *opérateur/arguments*²³⁶ (cf. Le Pesant et Mathieu-Collas, 1998).

Comme G. Gross le signale (2012, 2013), le découpage binaire (GN+GV) n'est pas justifié d'un point de vue syntaxique. D'une part, autant le sujet que l'objet ou les objets profitent d'une autonomie distributionnelle dans le cadre de la phrase, comme le prouve le nombre de transformations²³⁷ dont tous les deux font l'objet :

- a) La possibilité d'effacer des compléments est la preuve de la liberté syntaxique du complément par rapport au prédicat :

(2) *Léa mange (une pomme).*

- b) Des mises en évidence comme le détachement ou l'extraction :

(3) a. *Mon enfant a perdu le ballon.*

b. *Mon enfant, il a perdu le ballon.*

²³⁵ À la différence de Fillmore et de Harris, Tesnière restreint la nature grammaticale du prédicat au seul verbe. Lorsqu'il est nécessaire d'avoir un terme plus général, il utilise *nœud*, qui est le sommet d'un stemma prédicatif.

²³⁶ En ce qui concerne la terminologie utilisée, Harris n'emploie à peine le terme *prédicat*, préférant celui d'*opérateur*. Pourtant, M. Gross (disciple de Harris, et qui a développé ce travail pour la langue française) utilise *prédicat*, le distinguant parfois de son pendant sémantique (le *prédicat sémantique*). Quant aux éléments dépendants, ils sont parfois appelés *actants*, d'autres fois, *arguments*.

²³⁷ Étant donné que le terme *transformation* a été repris par différentes écoles linguistiques avec des sens sensiblement divergents, nous tenons à préciser que la notion de transformation telle que G. Gross la conçoit est héritière de la grammaire transformationnelle de Harris, c'est-à-dire des règles de réécriture et de substitution, et de la conception empirique de M. Gross (« Les transformations pourraient n'être qu'un dispositif expérimental qui permet de découvrir et de localiser les contraintes syntaxiques (et sémantiques) qui lient les éléments des phrases », M. Gross, 1975, p. 9). La validité de ces opérations repose sur le maintien du même sens de base de A en B.

c. *C'est mon enfant qui a perdu le ballon.*

d. *Le ballon, mon enfant l'a perdu.*

e. *C'est le ballon que mon enfant a perdu.*

c) Des insertions entre le sujet et le prédicat, ainsi qu'entre l'objet et le prédicat :

(4) a. *Comme vous le savez, la France a gagné la Coupe du monde de football.*

b. *La France, comme vous le savez, a gagné la Coupe du monde de football.*

c. *La France a gagné, comme vous le savez, la Coupe du monde de football.*

D'autre part, cette distribution en GN et GV place le sujet en position de supériorité vis-à-vis des objets. Néanmoins, souvent les informations apportées par les objets sont plus importantes que celles du sujet. Citons quelques cas de figure où nous pourrions observer que ce sont les objets qui déterminent :

a) le sens en cas de polysémie du verbe :

(5) a. *Paul aspire à de longues vacances.* ('ambitionner')

b. *Paul aspire une bouffée d'air.* ('inspirer')²³⁸

b) le figement d'une séquence²³⁹ :

(6) a. *Léa a pris le taureau par les cornes.*

b. **Léa a pris un taureau par les cornes.*

c. **Le taureau que Léa a pris par les cornes.*

c) la possibilité d'effacement d'un objet avec ou sans modification de l'interprétation du prédicat :

(7) a. *Max a appris les tables de multiplication aux enfants.* ('faire connaître')

b. *Max a appris les tables de multiplication.* ('acquérir des connaissances')

d) la possibilité de passivation :

²³⁸ Ces deux exemples sont empruntés à Leclère (1998, p. 156).

²³⁹ Dans le modèle des classes d'objets on utilise de préférence le terme *séquence figée* pour se référer aux unités phraséologiques. Désormais, nous allons incorporer ce terme à notre discours.

(8) a. *Léa a adressé cette lettre au DRH de son entreprise.*

→ *Cette lettre a été adressée par Léa au DRH de son entreprise.*

b. *Léa s'est adressée au DRH de son entreprise.*

→ **Le DRH de l'entreprise a été adressé par Léa.*

En revanche, aucun moyen syntaxique ne permet de justifier la notion de *groupe verbal*²⁴⁰ (G. Gross, 2012). Ainsi, il s'est avéré que le modèle de la phrase élémentaire de Harris, justifié de manière strictement linguistique comme nous venons de l'observer et possédant une base mathématique, est plus opératoire d'un point de vue lexical pour le traitement automatique des langues naturelles que le découpage binaire (*sujet/prédicat ; thème/rhème*). C'est pourquoi dans le modèle des classes d'objets, on comprend la phrase élémentaire comme étant constituée d'un prédicat²⁴¹ et de ses arguments²⁴².

1.2. Le prédicat comme relation

En ce qui concerne le terme *prédicat*, il peut aussi correspondre à des réalités différentes en fonction de la position théorique adoptée quant à la conception de la phrase élémentaire. Dans le premier modèle, le prédicat désigne le verbe et ses

²⁴⁰ G. Gross (2012) indique que la notion de groupe verbal pourrait avoir de l'intérêt pour désigner les séquences figées, plutôt que pour les constructions libres. En effet, dans le cas des séquences figées, les substantifs à droite ne fonctionnent pas vraiment comme des objets et toute la séquence (V + N) forme un ensemble avec un seul sens global qui n'est pas analysable en sous-séquences syntaxiques ou sémantiques.

²⁴¹ Dans le lexique-grammaire (M. Gross, 1975), dont la méthode des classes d'objets est héritière, les termes *opérateur* et *prédicat* sont considérées équivalents.

²⁴² Nous tenons à faire une remarque sur la différence entre la notion de *phrase élémentaire* et la notion de *phrase simple*. On entend par *phrase élémentaire* la phrase qui ne comporte qu'un prédicat ; par exemple dans *Luc appelle Max*, le verbe *appeler* est le seul prédicat. Par contre, la *phrase simple* est celle qui comporte un seul verbe conjugué, mais elle peut comporter plusieurs prédicats ; ainsi, dans *Luc appelle Max dans le but d'avoir de ses nouvelles*, il y a trois prédicats : le verbe *appeler*, la conjonction *dans le but de*, et le nom *nouvelles*. Sur les formes morphologiques du prédicat, voir la section 1.3. *Les formes morphologiques du prédicat* (dans ce chapitre).

compléments. Dans le second, les compléments ne font pas partie du prédicat, mais ils sont, avec le sujet, des arguments du prédicat.

Pour préciser la notion de prédicat tel que nous le concevons, nous reprendrons la définition de G. Gross (2012) :

À la suite de Z. Harris (1976), nous définissons un prédicat comme un mot qui opère une sélection déterminée parmi les mots du lexique pour établir avec eux un schéma formant la base d'une assertion. Les arguments sont les éléments lexicaux ainsi sélectionnés par les prédicats dans le cadre de la phrase simple. Ils n'exercent eux-mêmes aucune contrainte sur d'autres mots ; Harris les appelle des arguments *élémentaires*²⁴³ (p. 13).

Nous employons donc le terme *prédicat* au sens harrisien d'*opérateur* : dans la phrase, celui qui sélectionne les arguments. À la base, le prédicat est donc une relation entre plusieurs arguments. Dans la phrase élémentaire *Léa boit un thé*, *Léa* renvoie à un référent ; *thé* renvoie lui aussi à un référent ; et *boit* exprime la relation qui existe entre *Léa* et *thé*. Autrement dit, le prédicat verbal *boit* met en relation deux éléments : celui qui boit et ce que l'on boit. Le prédicat a les propriétés d'un opérateur dont les variables sont les arguments. Ainsi, la représentation fonctionnelle du contenu propositionnel serait la suivante : $F(x)$; F étant la fonction ou relation (le prédicat) et x étant la variable (argument).

1.3. Les formes morphologiques du prédicat

Le verbe n'a pas forcément la place centrale qui était la sienne dans la tradition linguistique héritée par la grammaire générative. Le prédicat étant le noyau de la phrase, il est susceptible d'être représenté par d'autres parties du discours outre le verbe²⁴⁴, et il peut correspondre autant à des formes simples que complexes :

- un verbe :

(9) a. *Léa boit un thé.*

²⁴³ L'argument élémentaire est celui qui ne peut jamais être prédicatif ; par exemple, des noms des <outils> comme *fourchette*, *table*, *engin*. En effet, il y a des arguments qui peuvent fonctionner comme des prédicats dans d'autres constructions (cf. section 6. *La phrase complexe*, dans ce chapitre).

²⁴⁴ Nous observons ici une autre innovation de Harris.

b. *Max a posé un lapin à ses amis.*

- un nom :

(10)a. *Léa a fait un cadeau à Paul.*

b. *Léa donne un coup de main à Luc.*

- un adjectif :

(11) a. *Léa est heureuse.*

b. *Léa est bête comme ses pieds.*

- une préposition :

(12) a. *Les clés sont dans le tiroir.*

b. *Le thermomètre est au-dessous de zéro.*

La morphologie est donc reléguée à une place secondaire car elle ne permet pas de rendre compte du rôle syntaxique des unités lexicales dans la phrase. En ce qui concerne l'analyse automatique, un logiciel qui catégorise de manière adéquate les unités lexicales d'un texte, ne serait pas capable d'en faire l'analyse.

1.4. Les prédicats polymorphiques

Par ailleurs, d'autres prédicats sont polymorphiques, c'est-à-dire un même contenu prédicatif peut développer plusieurs formes grammaticales différentes simultanément, sans que le contenu sémantique ni le schéma d'arguments soient modifiés. Par exemple, le contenu propositionnel $PI_{(1)} = \text{aimer}$ (prédicat (*Luc* argument1 ; *Léa* argument2)) peut être réalisé ayant comme prédicat un verbe, un substantif ou un adjectif (Grezka, 2006, p. 30-31) :

(13) a. *Luc aime Léa.* → $PI_{(1)} = \text{aimer}$ (N0 : *hum* ; N1 = *hum*)

b. *Luc éprouve de l'amour pour Léa.* → $PI_{(1)} = \text{amour}$ (N0 : *hum* ; N1 = *hum*)

c. *Luc est amoureux de Léa.* → $PI_{(1)} = \text{amoureux}$ (N0 : *hum* ; N1 = *hum*)

Les prédicats polymorphiques doivent avoir une racine prédicative identique, le même schéma d'arguments et la même signification.

1.5. Les racines prédicatives

La notion de racine prédicative permet de regrouper des emplois prédicatifs²⁴⁵ associés aux niveaux morphologique et sémantique, mais qui n'ont pas nécessairement les mêmes propriétés syntaxiques (cf. Buvet, 2008b). Ainsi, des dérivés morphologiques, qui partagent la même racine prédicative, peuvent avoir des schémas prédicatifs différents, et donc, ils correspondent à des classes sémantiques différentes (G. Gross, 2013, p. 27) :

(14) a. *Paul a peur de descendre à la cave.*

b. **Paul est peureux de descendre à la cave.*

Si *peur* et *peureux* ne subissent pas les mêmes modifications, c'est parce que le nom prédicatif *peur* appartient à la classe <sentiment> ; tandis que *peureux* est un adjectif prédicatif de <propriété>. Il en va de même, par exemple, pour *sympathie* et *sympathique* (G. Gross, 2013, p. 27) :

(15) a. *J'ai de la sympathie pour Jean.*

b. *Jean m'est sympathique.*

Le nom prédicatif est un actif, et l'adjectif prédicatif est un passif. Ainsi, souvent on a surestimé ces régularités morphologiques qui existent dans la langue naturelle (G. Gross, 2013), mais elles demandent une analyse de toutes leurs propriétés au cas par cas.

Dans ce sens, l'analyse en racines prédicatives offre de nombreux avantages car elle peut résoudre des problèmes d'analyse contrastive de plusieurs langues. Le passage d'une catégorie grammaticale dans une langue à une autre n'est pas automatique. Alors la description de la langue en racines prédicatives pourrait aider à surmonter cette difficulté ; par exemple, parmi les prédicats des <aides>, le verbe français *emprunter* n'a pas de verbe équivalent verbal en espagnol, il faudrait dire *tomar/coger prestado*. En espagnol cette signification est donc exprimée à partir de l'adjectif prédicatif

²⁴⁵ Pour une explication de la notion d'emploi tel que nous le concevons, voir la section 3. *La description des prédicats : la notion d'emploi* (dans ce chapitre).

prestado, qui est actualisé par un verbe support (*tomar* o *coger*). Néanmoins, on a pu montrer que la constitution des racines et la description de tous les emplois qui en font partie est bien plus complexe qu'il ne paraît²⁴⁶.

1.6. Le schéma d'arguments vs la polysémie

Les prédicats se caractérisent d'abord par le nombre et la nature sémantique de leurs arguments. La plupart des prédicats sont polysémiques, le changement de sens du prédicat implique nécessairement un changement du schéma d'arguments. Observons les phrases suivantes (Buvet, 2008b, p. 26) :

(16) a. *Cette étoile est particulièrement brillante.*

b. *Léa est particulièrement brillante.*

L'unité lexicale *brillant* correspond à deux prédicats différents du fait que leurs schémas d'arguments sont différents. En effet, l'argument 1 (celui qui correspond au sujet) de la phrase (16a) a le trait sémantique < inanimé concret > tandis que l'argument 1 de la phrase (16b) doit être caractérisé par le trait < humain >.

Ainsi, un prédicat est d'abord défini par la suite la plus longue des arguments qu'il sélectionne²⁴⁷. C'est ce qu'on appelle le *schéma d'arguments* ou le *schéma prédictif*. Au niveau du traitement automatique des textes, la reconnaissance d'un schéma prédictif dans un texte donné permet d'identifier la sélection de l'ensemble de mots qui le composent et permet aussi d'exclure les autres lectures possibles de ces mots. Il devient donc évident que la première étape de l'analyse automatique d'un texte doit être la reconnaissance de l'ensemble des schémas prédictifs.

Cela implique que l'on soit capable de distinguer un argument d'un circonstant. La différence réside dans le fait que l'argument est induit par le sens même du prédicat. Ainsi dans :

²⁴⁶ À ce sujet, Sfar (2007) a mené une remarquable étude dans le cadre de sa thèse de doctorat. Dans notre dictionnaire nous n'avons travaillé que sur les noms prédictifs des <aides>. La recherche de racines prédictives pourrait faire l'objet d'une autre étude postérieure.

²⁴⁷ Ensuite, il sera défini aussi par son actualisation et les transformations qu'il accepte.

(17) *Luc a acheté cette moto à Max pour trois mille euros la semaine dernière*

le prédicat *acheter* implique un acheteur (*Paul* = argument 1), un produit ou un bien acheté (*moto* = argument 2), un vendeur (*Max* = argument 3) et le prix (*trois mille euros* = argument 4). Ces quatre notions sont impliquées dans le sens même du verbe *acheter*. C'est pourquoi il s'agit d'arguments. Cependant, la spécification temporelle (*la semaine dernière*) n'est pas induite par le sens du prédicat. Ce dernier complément est un circonstant.

On notera que certains arguments peuvent être effacés lors de la réalisation de la phrase. Par exemple, dans *Paul a acheté cette moto la semaine dernière*, on observe que seulement deux des arguments ont été saturés, mais on comprend bien que Paul a dû acheter la moto à un vendeur pour un certain prix.

2. La phrase élémentaire

2.1. La notion de phrase élémentaire

À la lumière des notions que nous avons déjà révisées, nous pouvons donc définir la phrase élémentaire de la manière suivante : une phrase élémentaire est composée d'un prédicat de premier ordre et de ses arguments élémentaires (Boons et al., 1976a, 1976b ; M. Gross, 1975 ; Harris, 1971, 1976 ; Leclère, 1998)²⁴⁸. En conséquence, la représentation de la phrase est :

Pred (arg₁, arg₂, arg_n...)

Si nous reprenons la dernière phrase qui nous a servi comme exemple, la représentation serait la suivante : *acheter (Paul, voiture, Jean, mille euros)*. Le passage de cette notion abstraite de relation entre des éléments à la réalisation d'une phrase exige deux opérations : la *linéarisation* et l'*actualisation*.

²⁴⁸ Nous ne citons que l'influence la plus directe (Harris) et des auteurs du modèle du lexique-grammaire, qui est à la base du modèle des classes d'objets, mais nous n'oublions pas que des linguistes tels que Tesnière et Fillmore avaient déjà une interprétation similaire de la phrase simple (cf. section 1.1. *Les origines harrissiennes de la notion de prédicat*, dans ce chapitre).

Les langues analytiques, comme le sont le français et l'espagnol, ont recours à la linéarisation (la position des arguments par rapport au prédicat) pour spécifier le rôle sémantique des arguments : le sujet est à gauche et les compléments à droite. Ainsi dans :

(18) *Max aime Léa*

la position des arguments est déterminante (*Max* est l'expérimenteur et *Léa* la bénéficiaire) car l'inversion des arguments modifierait le sens de la phrase. On notera qu'en français l'ordre des éléments de la phrase suit des règles plus strictes qu'en espagnol. Néanmoins, dans les deux langues on peut altérer l'ordre des éléments dans la phrase à des fins communicatives telles que la focalisation ou la topicalisation. Par contre, dans les langues flexionnelles, la structure de la phrase profite d'une grande souplesse. Dans des langues flexionnelles comme le latin et le grec, la linéarisation se traduit de manière différente. Ce sont les cas (désinences) qui assument les fonctions argumentales. Encore, il existe d'autres langues, comme l'allemand qui participent des deux systèmes : linéarisation et cas.

D'autre part, l'actualisation sert à spécifier les différentes informations relatives à l'expression du temps, de l'aspect et de la personne dans le cas des prédicats, et la détermination dans le cas des arguments et des prédicats. Elle se situe donc à la croisée du lexique, de la morphologie, de la sémantique, de la syntaxe et de la pragmatique. Comme le résume Mejri (2011b) :

[l'actualisation] n'est en fait que l'opération par laquelle s'effectue l'ancrage de tout énoncé dans la situation d'énonciation dans laquelle il prend naissance : un locuteur (le moi), un espace dans lequel il se situe (l'ici) et une inscription dans le temps (le maintenant) dont la division en époques se fait par la création de l'énoncé lui-même. (p. 65)

L'actualisation prédicative dépend uniquement des occurrences des prédicats, car leur forme varie en fonction de la forme morphologique du prédicat. Si les prédicats verbaux portent en eux-mêmes leurs propres marques (désinence verbale ou verbe auxiliaire), les prédicats nominaux et les adjectivaux doivent avoir recours à des actualisateurs externes tels que les verbes supports (pour les noms prédicatifs) et le verbe *être* (pour les adjectifs prédicatifs). Par ailleurs, l'actualisation argumentale est subordonnée aux relations entre les prédicats et leurs arguments. L'outil d'actualisation propre aux arguments est la préposition. À part ces deux sortes d'actualisation

différentes, autant les prédicats que les arguments sont actualisés par une détermination spécifique (Buvet, 2008a, 2009a).

2.2. L'organisation du lexique en trois fonctions primaires

En conséquence, dans le cadre du modèle des classes d'objets, le lexique s'organise en trois sortes d'unités linguistiques : les prédicats, les arguments et les actualisateurs. Pour reprendre des concepts déjà abordés et bien marquer les différences entre ces trois entités, nous pouvons définir les prédicats comme des unités linguistiques qui acquièrent leur sens à partir de leurs constructions syntagmatiques (le schéma d'arguments) ; les arguments par contre sont des unités linguistiques de nature phrastique ou non phrastique, à valeur référentielle, dotés intrinsèquement de sens ; et les actualisateurs sont les éléments (unités lexicales ou morphèmes flexionnels) qui permettent d'inscrire le prédicat saturé de ses arguments dans le discours.

En effet, l'étiquetage des unités lexicales en catégories grammaticales (encore appelées *parties du discours*) ne permet pas de rendre compte de manière automatique des relations syntaxiques reliant les éléments de la phrase pour créer les structures qui sous-tendent les textes (G. Gross, 2012). On notera aussi que les trois fonctions du lexique n'ont pas de relation exclusive avec une seule catégorie grammaticale. Ainsi, la même unité linguistique peut assurer les trois fonctions dans des phrases différentes :

- (19) a. *Luc regarde les nuages.* (argument)
- b. *Il y a des nuages.* (prédicat)
- c. *Je voudrais un nuage de lait.* (actualisateur²⁴⁹)

D'ailleurs, comme Buvet (2002) l'indique, « la partition des unités linguistiques relative à leur statut de prédicat, d'argument élémentaire ou d'actualisateur ne recoupe pas celle qui a trait aux parties du discours du point de vue de l'opposition mot lexical / mot grammatical » (p. 189). C'est pourquoi dans ce modèle, les unités linguistiques sont catégorisées en termes de prédicats, d'arguments et d'actualisateurs et elles font l'objet de descriptions différentes selon qu'elles relèvent de la signification lexicale, en permettant notamment de rendre compte de la polysémie, ou de la signification

²⁴⁹ Dans cette phrase le mot *nuage* est un déterminant quantifieur.

grammaticale, qui prend en charge l'actualisation et permet aussi de traiter la polysémie (Buvet 2008a ; Blanco et Buvet, 2004).

Cette division en trois fonctions est fondée sur une hiérarchisation syntacticosémantique. Les prédicats prévalent sur les arguments. En effet, le prédicat est le noyau de la phrase et par conséquent, il est l'élément le plus important. C'est le prédicat qui crée la phrase et qui détermine sa structure car c'est lui qui sélectionne le nombre et le type d'arguments. D'ailleurs, « c'est pourquoi la structuration d'un discours donné est rapporté en premier lieu aux divers prédicats qu'il comporte et en deuxième lieu, aux arguments qui leur sont rattachés » (Buvet, 2002, p. 189). Les actualisateurs occupent la partie inférieure de cette hiérarchisation. Rappelons qu'ils sont même exclus de la représentation de la phrase ; ils n'interviennent qu'au dernier moment, quand on passe de la phrase à l'énoncé. Ils sont subordonnés soit directement aux prédicats, soit aux relations entre ceux-ci et leurs arguments. Ainsi, les unités lexicales sont définies par leur combinatoire et par leurs relations hiérarchiques de dépendance entre les différents constituants de la phrase.

2.3. La phrase élémentaire vs la polysémie

Nous avons signalé plus haut que l'un des postulats du modèle des classes d'objets est que l'unité minimale d'analyse est la phrase élémentaire. À l'origine de ce postulat se trouve le *principe de contextualité* de Frege (1884/1970), ultérieurement repris par Putnam : « Les mots n'ont ni signification ni référence en dehors du contexte de l'énoncé où ils figurent. Autrement dit, la signification d'un énoncé n'est pas la somme des significations des mots qui le composent. C'est le contraire : un mot tire son sens de l'énoncé dans lequel il se trouve »²⁵⁰ (Putnam, 1995, cité par Le Pesant et Mathieu-Colas, 1998, p. 7). En somme, le sens est véhiculé par l'environnement syntacticodistributionnel du mot, ce qui justifie que la phrase élémentaire soit considérée le niveau d'analyse minimale et non pas le mot.

²⁵⁰ Propos de Putnam dans une interview publiée dans *Le Monde*, 22-23/10/1995.

En effet, la plupart des mots sont polysémiques, ce qui constitue l'un des problèmes majeurs du traitement automatique des langues naturelles. Voyons quelques exemples du verbe *descendre* :

(20)a. *Luc descend les escaliers.* ('aller du haut en bas')

b. *Luc descend le criminel.* ('tuer')

c. *Luc descend deux bouteilles de vin.* ('boire')

Le sens de *descendre* varie en fonction de la nature sémantique de son second argument : un <locatif>, un <humain> ou une <boisson>. C'est le contexte linguistique qui détermine le sens. Donc, les mots ne peuvent être décrits que dans une structure phrastique²⁵¹. En outre, les recherches en traitement automatique de la langue ont montré que la séparation de significations au moyen des propriétés syntacticosémantiques est le seul outil de désambiguïsation dont disposent les automates, car la polysémie est quasi inexistante dans le cadre de la phrase.

Si on nous demandait de définir le sens du verbe *descendre*, ce qui semble bien banal, peut-être aurait-on répondu 'aller du haut en bas'. Nous aurions sélectionné l'un des sens de ce verbe, celui que nous considérons le plus fréquent ou tout simplement, le premier qui nous vient à l'esprit. Or, nous aurions bien pu répondre 'tuer', 'boire' ou encore 'diminuer' (*La marée descend*), 'avoir comme ancêtres' (*Descendre d'une famille ouvrière*), etc²⁵². Pour comprendre ou pour définir un mot, il faut prendre en considération le contexte lexical.

²⁵¹ M. Gross (1975) trouvait aussi que les structures syntaxiques sont les réceptacles de la matière sémantique. Il a ainsi soumis des masses d'entrées lexicales à des grilles syntaxiques pour établir des ensembles syntactico-distributionnels dans lesquels le lexique est classifié. Or, le fait de partager le même moule syntaxique n'est pas suffisant pour discriminer les différentes significations d'une unité lexicale, comme nous avons pu vérifier dans les exemples du verbe *descendre* (NVN). En effet, ces schémas n'apportent aucune information sémantique permettant d'affiner la description.

²⁵² Force est de constater que dans les cours de langue vivante, quand un élève demande la traduction d'un mot, très souvent le professeur donne le contexte linguistique pour proposer les différentes traductions de ce mot en L1. Une pratique bien moins courante dans le cours qui se tiennent en langue maternelle. Pourrait-on en déduire que lorsqu'on est dans des situations de communication bilingues ou plurilingues on est plus conscient de la polysémie des mots ?

2.4. La notion de contexte

Dans cette section, nous allons examiner ce que nous entendons par *contexte* (aussi appelé *environnement de gauche* et *environnement de droite* d'un mot) et la complexité de cette notion, à la lumière des réflexions de G. Gross (2008, 2010a).

2.4.1. Les contextes aléatoires

Comme les unités lexicales ne peuvent pas être définies hors de leur contexte, il est capital de savoir comment déterminer ce contexte définitoire. On pourrait penser que toutes les informations nécessaires à la description d'un mot peuvent être déterminées par l'ensemble des environnements que l'on pourrait trouver de manière automatique dans de grands corpus. Pourtant une requête consistant à sélectionner les cinquante caractères qui précèdent et qui suivent immédiatement notre unité lexicale ne nous mènerait pas directement à des résultats exploitables pour la définition de l'emploi de l'unité lexicale, car les éléments qui la définissent ne sont pas forcément contigus. La recherche des cooccurrents ne peut donc pas être mécanique. En effet, les insertions sont assez fréquentes entre un prédicat et ses arguments. Ces insertions sont de nature diverse :

a) des incisives :

(21) *La CAF, nous le savons tous, n'accorde pas des allocations à n'importe qui.*

b) des négations :

(22) *Mon entreprise n'octroie plus de congés de formation.*

c) des adverbes :

(23) *Il a participé grandement au mécénat de cet artiste émergent.*

d) des circonstanciels :

(24) *Luc bénéficie, depuis quelques mois, d'une allocation familiale.*

e) des déterminants nominaux :

(25) *Cette grande entreprise bénéficie d'un tas d'avantages fiscaux.*

Alors, si on définit les prédicats par la nature de ses arguments, il faut que l'automate soit capable d'ignorer ces insertions pour retrouver la distribution de l'unité

lexicale en question. Cette notion de distribution doit reposer sur une analyse qui repère les éléments dont l'interdépendance constitue des ensembles significatifs. Pour que la machine soit capable de faire abstraction des insertions et récupérer le sens du prédicat, elle a besoin d'un logiciel qui contienne une description préalable de tous les emplois des mots en question (prédicat + schéma d'arguments).

2.4.2. Les contextes contraints

Parfois l'environnement d'une unité lexicale n'aide pas à reconnaître son sens. Les travaux de Maurice Gross et du LADL ont montré que les constructions verbales figées ont la même structure interne que les constructions régulières libres. Il s'agit de « faux » environnements puisque le sens du prédicat ne provient pas de la nature de ces « faux » arguments. Toute la séquence figée constitue une unité sémantique dont le sens est non compositionnel. Ainsi dans :

(26) *Jean a cassé sa montre.*

(27) *Jean a cassé sa pipe.*

il y a apparemment une structure : *casser (Jean, montre)* et *casser (Jean, pipe)*. Pourtant le second exemple peut avoir une seconde interprétation, en tant que séquence figée : *casser sa pipe (Jean)*. À syntaxe différente, sens différent : 'rompre' / 'mourir'.

Il s'avère nécessaire de faire une description exhaustive des séquences figées pour que la machine soit capable de reconnaître s'il s'agit d'une seule unité sémantique²⁵³ et, par exemple, la traduire en espagnol par *estirar la pata* et non pas par **romper su pipa*. Ces constructions doivent être repérées les premières dans la traduction automatique, pour éviter une traduction compositionnelle.

2.4.3. Les contextes significatifs

Comme G. Gross (2010b) l'indique, « la détermination du sens d'un prédicat à l'aide du contexte n'est donc pas une opération simple » (p. 190). Nous venons de

²⁵³ Ces séquences peuvent avoir le même comportement distributionnel que les séquences libres, en acceptant des insertions à l'intérieur même de la séquence figée : *Paul prend toujours les vessies pour des lanternes ; La moutarde lui a, semble-il, monté au nez* (G. Gross, 2010b, p. 189). Pour revenir sur les caractéristiques du figement, on voit que la possibilité d'insertion n'est pas un critère valable pour distinguer les séquences libres des séquences figées.

démontrer que le contexte n'est pas nécessairement une donnée immédiate. Alors, la première tâche de tout traitement automatique consiste à repérer les prédicats dans un texte donné, et à partir de là, reconnaître les phrases élémentaires dans lesquelles ils déploient leurs arguments²⁵⁴. Ce travail de reconnaissance des structures phrastiques ne peut se faire que sur la base d'un dictionnaire électronique qui contienne une description exhaustive des prédicats.

3. La description des prédicats : la notion d'emploi

Nous avons vu que le sens est le résultat de l'interaction entre les composantes syntaxique et lexicale, et que la syntaxe fait partie de la définition des unités lexicales. Ces trois niveaux (lexical, syntaxique et sémantique) sont donc interdépendants. Il ne convient donc pas de les étudier isolément, car les résultats seraient des descriptions qui ne proposeraient que des points de vue fragmentaires sur le message linguistique, qui intègre tous les niveaux. D'ailleurs, il est extrêmement difficile de tracer la frontière entre un niveau et l'autre. De surcroît, la polysémie et le figement étant deux des propriétés fondamentales des langues naturelles et deux des principaux problèmes du traitement automatique des langues naturelles, il s'avère nécessaire de faire une description de la langue où le lexique, la syntaxe (la combinaison des éléments lexicaux) et la sémantique (le résultat de la combinatoire des éléments lexicaux) ne soient pas cloisonnés. Justement ces trois niveaux d'analyse sont regroupés sous la notion d'*emploi*, autrement dit l'emploi d'une unité lexicale dans une phrase.

Dans le modèle des classes d'objets on reprend la notion d'emploi développée par M. Gross (1975) dans le cadre du lexique-grammaire²⁵⁵. Il avait déjà annoncé que le traitement automatique du lexique exigeait une description qui accorde à chaque mot un comportement prévisible dans la phrase. Cette description doit intégrer tous les niveaux

²⁵⁴ L'étape suivante serait de repérer les liens qui relient les phrases entre elles (les connecteurs).

²⁵⁵ À l'origine, on retrouve Gottlob Frege, qui avait déjà signalé que les unités lexicales doivent être définies en fonction de leur emploi. Cette idée est à la base de la conception de la phrase simple (opérateur + arguments) de Zellig S. Harris ; celui-ci, rappelons-le, est le mentor en matière linguistique de M. Gross.

d'analyse pour que l'automate soit capable de reconnaître les emplois des mots et de résoudre les cas d'ambiguïté.

Buvet et Grezka (2009) signalent que :

Un emploi prédicatif est de nature linguistique, il est directement observable. Un prédicat est de nature métalinguistique, on le met en évidence par l'observation des emplois prédicatifs. Autrement dit, un emploi prédicatif est une occurrence d'un prédicat sous forme d'adjectif, de nom ou de verbe (p. 72).

G. Gross (2012) définit l'emploi prédicatif ou emploi de prédicat comme « un schéma prédicatif et l'ensemble des propriétés qui lui sont rattachées » (p. 33). Ainsi, deux emplois différents d'un prédicat n'ont jamais les mêmes propriétés. Ces propriétés se basent sur six catégories générales de propriétés linguistiques : des propriétés configurationnelles comme la structure argumentale du prédicat, qui sert à spécifier le nombre et la nature des arguments, et leur mode de structuration ; des propriétés sémantiques (le sens du prédicat et le domaine auquel il appartient) ; des propriétés combinatoires, servant à décrire l'actualisation des prédicats et les verbes supports aspectuels qu'ils sélectionnent ; des propriétés paraphrastiques, servant à spécifier les possibles restructurations du schéma prédicatif ; des propriétés morphologiques telles que la spécification de la partie du discours, si c'est une forme simple (*aide*) ou une forme complexe (*aide personnalisée au logement*) et l'actualisation des prédicats lorsqu'il s'agit de morphèmes flexionnels ; et finalement, des propriétés sociolinguistiques, telles que le niveau de langue (Buvet, 2008b). Ces paramètres nécessaires à la description des emplois ont été définis à maintes reprises dans la littérature du lexique-grammaire et notamment dans celle du modèle des classes d'objets (Buvet, 2008b ; G. Gross, 2010a, 2012, 2014 ; Mathieu-Colas, 1994). Dans ce qui suit, nous allons les décrire au détail.

3.1. Un emploi est associé à un schéma d'arguments

Il s'agit de l'information la plus importante, car les autres propriétés du prédicat dépendent de ce schéma d'arguments. Un emploi prédicatif est d'abord défini par son schéma d'arguments, et inversement le prédicat délimite le champ de ses arguments.

Ceux-ci doivent être préalablement décrits en termes de classes d'objets²⁵⁶. En guise d'exemple, nous présentons les schémas d'arguments de certains emplois du verbe *abattre* (G. Gross, 2012, p. 35) :

- (28) a. *Abattre* /N0: *hum* /N1: <*arbre*>
- b. *Abattre* /N0: *hum* /N1: <*aéronef*>
- c. *Abattre* /N0: *hum* /N1: <*construction*>
- d. *Abattre* /N0: *hum* /N1: *hum*
- e. *Abattre* /N0: *évènement* /N1: *hum*
- f. *Abattre* /N0: *hum* /N1: <*animal de boucherie*>

Nous pouvons donc constater qu'un prédicat a autant d'emplois qu'il a de classes d'objets différentes en position argumentale.

Comme certains arguments peuvent être effacés dans le discours et que ce fait peut provoquer des confusions ou des ambiguïtés dans l'interprétation du prédicat, il faudra aussi spécifier dans le dictionnaire électronique les possibles effacements d'arguments et leur incidence dans l'interprétation de la phrase²⁵⁷.

3.2. Un emploi est associé à un sens

Rappelons que le sens est déterminé formellement par le schéma d'arguments. Autrement dit, « les positions argumentales, définies en compréhension à l'aide de ces classes sémantiques, déterminent le sens du prédicat » (G. Gross, 2012, p. 34). Les unités lexicales prennent donc leur sens dans le cadre des emplois. Dès lors que l'emploi est finement défini, il est possible de lui attribuer un synonyme (noté *Sy*)²⁵⁸ et une traduction dans une langue étrangère (notée *Es*). D'ailleurs, la synonymie est un outil très performant pour mettre en évidence les différentes significations d'un prédicat

²⁵⁶ Une classe d'objets est un ensemble de substantifs sémantiquement homogènes qui partagent les mêmes propriétés syntacticosémantiques. À propos de cette notion, voir la section 4.1. *La notion de classe d'objets* (dans ce chapitre).

²⁵⁷ Par exemple, lorsque le verbe *sentir* apparaît avec un argument 1 (sujet) <concret> ou <humain> et sans adverbe après le verbe, il signifie 'sentir mauvais'.

²⁵⁸ Et éventuellement un antonyme.

polysémique, car chaque emploi de prédicat différent aura un synonyme différent. Observons quelques exemples du verbe *abattre* (G. Gross, 2012, p. 35) :

- (29) a. *Abattre* /NO: hum /N1: <arbre> /Sy :couper /Es : talar
b. *Abattre* /NO: hum /N1: <aéronef> /Sy :descendre /Es : abatir
c. *Abattre* /NO: hum /N1: <construction> /Sy : démolir /Es : derriber
d. *Abattre* /NO: hum /N1: hum/Sy : exécuter /Es : matar
e. *Abattre* /NO: évènement /N1: hum/Sy : démoraliser /Es : abatir
f. *Abattre* /NO: hum /N1: <animal de boucherie> /Sy : tuer /Es : sacrifier

Nous pouvons donc observer que le schéma d'arguments lève toute ambiguïté. Les emplois polysémiques d'un prédicat ne peuvent pas avoir les mêmes classes sémantiques dans leurs positions argumentales. Le changement de classe implique un changement de sens du prédicat.

3.3. Un emploi est associé à une forme morphologique

La fonction de prédicat peut être assumée par des verbes, des noms, des adjectifs et éventuellement certaines prépositions. Rappelons que les racines prédicatives²⁵⁹ peuvent regrouper trois formes morphologiques ; par exemple, la racine *respect-* (G. Gross, 2012, p. 36) :

- (30) a. *Luc respecte les lois.*
b. *Luc a le respect des lois.*
c. *Luc est respectueux des lois.*

D'autres racines prédicatives n'ont que les formes du verbe et du nom, comme *emprunt-* :

- (31) a. *Léa a emprunté de l'argent à cette banque.*
b. *Léa a fait un emprunt à cette banque.*

Force est de constater que chaque variante morphologique correspond à un emploi différent. Enfin, d'autres prédicats n'ont qu'une forme morphologique :

²⁵⁹ Voir la section 1.5. *Les racines prédicatives* (dans ce chapitre).

(32) *Max donne un coup de main à Paul.*

Ces derniers sont appelés des *prédicats autonomes*. Ces informations sont intéressantes du fait qu'il n'existe pas toujours des équivalences morphologiques dans deux langues, c'est-à-dire qu'il existe une sorte de « trous » morphologiques.

3.4. Un emploi est associé à une actualisation

Rappelons que pour pouvoir inscrire un prédicat et son schéma d'arguments dans le discours, il doit être actualisé. L'actualisation des prédicats varie en fonction de leur nature morphologique. Ainsi, les prédicats verbaux sont actualisés moyennant la conjugaison ; tandis que dans le cas des prédicats nominaux, l'actualisation est prise en compte notamment par des verbes supports et la détermination²⁶⁰. C'est le prédicat nominal qui sélectionne le verbe support qui l'accompagne.

(33) a. *Luc soutient Paul.*

b. *Luc apporte son soutien à Paul.*

Revenant à la complexité de la notion de prédicat polymorphique, l'actualisation peut s'avérer utile pour distinguer des emplois différents de prédicats ayant la même forme morphologique. Par exemple, l'actualisation du verbe *abattre* est la même quel que soit son emploi, c'est-à-dire la conjugaison numéro 55 du *Bescherelle* (2012), tandis que les différentes nominalisations correspondant aux différents emplois de ce verbe polysémique sélectionnent des verbes supports différents pour leur actualisation : *procéder à l'abattage d'un arbre ; être dans un grand abattement*, etc. (G. Gross, 2014, p. 45). Encore, un verbe peut être défectif dans un emploi donné et ne pas l'être dans d'autres emplois : *Il pleut ; Les bombes pleuvaient sur le pays.*

3.5. Un emploi est associé à un système aspectuel

La catégorie grammaticale de l'aspect peut être définie comme : « une catégorie grammaticale qui exprime la représentation que se fait le sujet parlant du procès exprimé par le verbe (ou par le nom d'action), c'est-à-dire la représentation de sa durée, de son déroulement ou de son achèvement » (Dubois et al., 2012, p. 53).

²⁶⁰ Dans la section 5. *La description des actualisateurs*, nous expliquerons que les prédicats nominaux sont susceptibles d'être actualisés aussi par certains adjectifs et certains adverbes.

Le système aspectuel d'un emploi prédicatif décrit l'aspect interne du prédicat et la compatibilité aspectuelle avec les différents éléments de la phrase. L'aspect contribue donc à la définition de la classe des prédicats. Il existe deux grands types d'aspects : l'aspect *inhérent* ou *interne* des prédicats et l'aspect *contextuel*. L'aspect inhérent (aspect lexical) est constitutif de la valence sémantique des prédicats et il détermine s'il s'agit d'un prédicat ponctuel ou duratif. Le fait qu'il s'agisse d'un prédicat à aspect ponctuel ou duratif a une grande influence sur la sélection des déterminants et de nombreux adverbiaux. Par exemple, le prédicat *sottise*, lorsqu'il sélectionne le verbe support *faire*, il a un aspect ponctuel, au singulier, il accepte un déterminant indéfini, et, au pluriel, il ne semble pas avoir des restrictions sur les déterminants : *Paul a fait une sottise, Paul a fait (des, quelques, trois, d'autres) sottises* ; par contre, si le verbe support a un aspect itératif, c'est le déterminant défini qui s'impose : *Paul multiplie (les, *des, *ces, *quelques) sottises* (G. Gross, 2010a, p. 109 ; G. Gross, 2012, p. 39). Cette détermination dépend donc de l'aspect du prédicat.

D'autre part, l'*aspect contextuel* (aspect grammatical) relève de la combinatoire des formes prédicatives avec différents types de marqueurs. Il exprime l'itérativité ou informe sur les différentes phases du procès (inchoativité, progressivité, terminativité). Ces marqueurs d'aspect sont de nature très diverse en français et en espagnol, car il peut être pris en charge par des procédés grammaticaux tels que le préfixe *re-* (*Luc doit réécrire sa dictée / Luc tiene que reescribir el dictado*), des verbes auxiliaires (*Mon bébé commence à marcher / Mi bebé comienza a caminar*), des verbes supports (*L'État supprimera les bourses d'études d'élèves absentéistes / El Estado suprimirá las becas de estudios de alumnos absentistas*), des adverbes (*Léa lit souvent / Léa lee a menudo*), etc. C'est pourquoi la description de la compatibilité aspectuelle pour le traitement automatique de la langue est une tâche très complexe. Cependant, ces actualisateurs aspectuels contextuels sont contraints par l'aspect inhérent du prédicat en question.

Toujours au sujet des racines prédicatives, G. Gross (2012) fait remarquer que « l'aspect ne doit [...] pas être considéré comme une propriété d'une racine prédicative donnée, mais de chacun de ses emplois particuliers » (p. 38). Ainsi, l'emploi de *abattre* /NO: hum /NI: hum/ correspond à un prédicat ponctuel, car il n'accepte pas les marqueurs aspectuels suivants : **Le soldat (a commencé à, continue à, finit de) abattre l'ennemi* ; pourtant, l'emploi *abattre* /NO: hum /NI: <arbre> les accepte bien : *Le*

bucheron a commencé à abattre cet arbre ; Le bucheron a mis trois heures pour abattre cet arbre ; Le bucheron est en train d'abattre cet arbre.

3.6. Un emploi est associé à des modifications

Il est évident que les textes ne sont pas constitués que des phrases élémentaires où le prédicat est accompagné de la suite la plus longue de ses arguments (d'ailleurs, il est plutôt rare). Les schémas prédicatifs peuvent subir un grand nombre de modifications. Alors, il est important que l'automate soit capable d'identifier les changements de forme des phrases, et par la suite, de relever les emplois en question. Ces modifications sont très utiles pour dissocier les différents emplois prédicatifs. Il s'avère donc nécessaire de recenser de manière systématique les différentes constructions des emplois prédicatifs pour contribuer à l'identification des emplois dans les textes.

On conçoit ces modifications comme un ensemble de règles de réécriture qui permettent d'opérer des transformations dans le but de découvrir des faits de langue, des mécanismes et des données linguistiques qui ne se manifestent pas directement à l'observateur. Nous allons proposer ici une vue d'ensemble de la typologie de modifications²⁶¹ avec laquelle on travaille souvent dans le cadre du modèle des classes d'objets (cf. G. Gross, 2012).

3.6.1. Les transformations

Elles portent sur :

- la forme morphologique des prédicats (la nominalisation, l'adjectivation, etc.) ;

²⁶¹ Dans cette section, nous avons suivi la terminologie et la typologie de G. Gross (2012), selon laquelle les *transformations* seraient un type de *modification*. Cependant, nous avons constaté que dans d'autres publications antérieures et postérieures (G. Gross, 1996a, 2017, entre autres) cet auteur utilise les termes *modification* et *transformation* en tant que des synonymes. En effet, G. Gross avoue que « la terminologie est flottante dans ce domaine, elle change en fonction du temps et des écoles linguistiques » (G. Gross, 2012, p. 37). D'autant plus que la notion de transformation est à la base des grammaires transformationnelles et d'autres modèles de description de la langue plus ou moins influencés par leurs thèses.

- la modalité des phrases, par exemple, l'interrogation et la négation ;
- la récursivité, par exemple l'enchâssement, qui permet l'insertion d'une phrase verbale ou nominale en position argumentale : *On a appris que le président avait donné sa démission ; On avait appris la démission du président* (G. Gross, 2012, p. 40) ;
- les reprises anaphoriques, par exemple les pronominalisations (*il, le, leur, en, etc.*) et les relativations (*Luc a eu peur ; La peur que Luc a eue l'a empêché d'agir*).

3.6.2. Les mises en évidence

Elles servent à mettre l'accent sur un élément de la phrase en le plaçant en position initiale. Elles peuvent concerner :

- le verbe : *La neige tombe → Il tombe de la neige* (extraposition du verbe) ; *On a dit beaucoup de choses à ce sujet → Il a été dit beaucoup de choses à ce sujet* (extraposition passive) ; *Sa passivité m'exaspère → Ce qui m'exaspère, c'est sa passivité* (dislocation du verbe) ;
- les arguments : *J'ai aidé Luc → C'est moi qui ai aidé Luc* (extraction de l'argument) ; *Luc, je l'ai aidé* (dislocation à gauche) ; *Je l'ai aidé, Luc* (dislocation à droite) ; *Ce que j'ai vu, je ne pourrai jamais l'oublier* (dislocation d'un argument phrastique) ;
- des déterminants : *Des milliers de gens ont quitté le pays → Par milliers les gens ont quitté le pays* (G. Gross, 2012, p. 40).

3.6.3. Les effacements

Certains éléments de la phrase peuvent être omis dans différentes conditions textuelles, pour, par exemple, éviter des répétitions inutiles ou parce que l'information est considérée non pertinente dans la situation de communication. Cette modification peut concerner :

- le(s) argument(s) : *Luc est en train de lire (un roman) ;*
- le prédicat : *Ta voiture est plus propre que la mienne ;*
- l'actualisation du prédicat : *Pierre espère gagner.*

3.6.4. Les restructurations

Il s'agit de changements dans l'ordre canonique des arguments provoqués par différents besoins d'expressivité. Elles peuvent entraîner la modification de la forme du prédicat ou pas. Voici quelques exemples :

- les constructions croisées : *Les abeilles grouillent dans le jardin* → *Le jardin grouille d'abeilles* (G. Gross, 2012, p. 41) ;
- les reconstructions d'un complément indirect : *Luc m'exaspère par sa passivité* → *La passivité de Luc m'exaspère* ;
- les restructurations contrastives : *À Paul j'ai offert une chemise, à Luc un pull* ;
- les permutations de longueur : *J'ai demandé à la professeure l'aide dont j'avais besoin* ;
- les inversions rhétoriques : *Vint un gardien qui nous interdit l'entrée* (G. Gross, 2012, p. 41) ;
- les prolepses : *Cette prestation, j'en ai absolument besoin* ;
- les constructions à arguments réciproques : *Tu compares A avec B, Tu compares B avec A, Tu compares A et B* (G. Gross, 2012, p. 41) ;
- les constructions à sujets réciproques : *Luc aime Léa, Léa aime Luc, Luc et Léa s'aiment* ;
- le passif : *Un architecte basque a dessiné ce plan* → *Ce plan a été de dessiné par un architecte basque* ;
- la thématization de compléments circonstanciels: certains compléments circonstanciels peuvent être déplacés en début de phrase : *Il y a un grave accident à Paris* → *Paris a été le théâtre d'un grave accident* (G. Gross, 2012, p. 42).

3.7. Un emploi est associé à un domaine spécifique

Les marques d'un domaine spécifique (droit, géométrie, linguistique) aident à déterminer le sens d'un emploi. Cette information a été traditionnellement utilisée dans les dictionnaires du fait de son efficacité et est ainsi reprise dans les dictionnaires automatiques et les systèmes de traduction automatique. La spécification du domaine s'avère très utile en traduction automatique pour la levée d'ambiguïtés²⁶². Reprenant l'exemple du verbe *abattre*, dont nous avons présenté seulement 6 emplois, observons l'utilité de cette indication (G. Gross, 2014, p. 46) :

- (34) a. *Abattre* /N0: *hum* /N1: <*arbre*> : sylviculture
- b. *Abattre* /N0: *hum* /N1: <*aéronef*> : militaire
- c. *Abattre* /N0: *hum* /N1: <*construction*> : BTP
- d. *Abattre* /N0: *hum* /N1: *hum* : lg
- e. *Abattre* /N0: *évènement* /N1: *hum* : psychologie
- f. *Abattre* /N0: *hum* /N1: <*animal de boucherie*> : boucherie

3.8. Un emploi est associé à un registre de langue²⁶³

Étant de nature sociolinguistique, cette information relève d'un ordre différent des paramètres précédents. Il s'agit d'une propriété peu pratique pour les descriptions de la langue de spécialité, mais très utile lorsque le lexique relève de la langue générale. Ainsi, on peut bien saisir la différence, notamment au niveau du contexte de communication, entre *administrer un soufflet* ou *filer une mandale* (G. Gross, 2010a, p. 110).

²⁶² Un travail mené par Buvet (1996) a permis de constater que les indications de domaines d'emploi figurant dans les dictionnaires électroniques peuvent lever des ambiguïtés lexicales. Toutefois, les contre-exemples consignés montrent que les informations syntacticosémantiques et distributionnelles fines sont tout autant nécessaires pour réduire automatiquement la polysémie.

²⁶³ Cette propriété ne fait pas l'unanimité dans toutes les descriptions des emplois prédicatifs et travaux publiés dans le cadre du modèle des classes d'objets. Nous la considérons très utile, notamment pour la traduction automatique.

En résumé, la notion d'emploi prédicatif montre que tous les niveaux de la description linguistique sont interconnectés : lexicale, syntaxe et sémantique. Chaque emploi correspond à un schéma prédicatif et à une signification spécifiques, corrélés à un ensemble de propriétés sémantiques, syntaxiques, morphologiques et pragmatiques nécessaires pour le traitement automatique des langues naturelles. Dans le modèle des classes d'objets, un dictionnaire électronique est donc le recensement ordonné de l'ensemble d'emplois. Nous avons pu montrer également l'utilité de cette notion pour résoudre les problèmes d'ambiguïté dans la reconnaissance et d'interprétation du figement et de la polysémie. D'ailleurs, la description des emplois prédicatifs implique l'identification et la description de la polysémie d'un prédicat. L'emploi est donc l'unité structurelle de base que l'automate doit être capable de reconnaître, pour finalement « comprendre » et générer toutes les phrases possibles.

4. La description des arguments : les classes d'objets

4.1. La notion de classe d'objets

G. Gross (2012) fait une remarquable constatation au sujet de la levée de l'ambiguïté sémantique des prédicats :

Quand un prédicat a, dans une position argumentale donnée, des substantifs d'une même classe sémantique, son sens reste constant lors de la substitution d'un élément du paradigme. [...] Inversement, des arguments qui relèvent de classes sémantiques différentes déterminent des interprétations différentes du prédicat (p. 71).

Il s'avère donc nécessaire de classer les arguments potentiels d'un prédicat, en vue de discriminer les différentes significations que ce prédicat peut avoir. Ainsi, on sera en mesure de reconnaître de manière automatique un emploi donné et, par la suite, reconnaître aussi la phrase, qui, rappelons-le, est notre unité minimale d'analyse. La description de tous les emplois des prédicats implique donc la définition des arguments qui sont appropriés à chacun de ces emplois. Ces restrictions de sélection nous mènent à considérer quatre types de prédicats (G. Gross, 2008, 2012 ; Le Pesant et Mathieu-Colas, 1998) :

a) Des prédicats sans restriction argumentale : le prédicat accepte en position argumentale n'importe quel substantif. Ces prédicats sont assez rares. Il s'agit

généralement de prédicats monosémiques, alors, aucun problème d'interprétation ne se pose. Par exemple, les verbes *penser à*, *parler de*, *se souvenir de*, n'exercent aucune contrainte sur le substantif qui figure en position du premier complément ; les adjectifs *important*, *intéressant* (ainsi que d'autres formes morphologiques de ces prédicats telles qu'*intéresser*, *intérêt*) n'imposent non plus aucune restriction sur le sémantisme du substantif en fonction de sujet.

(35) *Je pense à (mes prochaines vacances, mon vélo cassé, mon voisin bruyant).*

(36) *(Ce problème, mon frère, ce livre) est important pour moi.*

b) Des prédicats à argument unique (les *singletons*) : ce sont des prédicats dont le spectre argumental est réduit à un seul substantif. C'est le cas, par exemple, du verbe *extraire* dans *extraire la racine (carrée, cubique) d'un nombre*²⁶⁴.

c) Des prédicats à large spectre argumental : ces prédicats sélectionnent de grandes classes sémantiques en position argumentale. Celles-ci sont appelées des *hyperclasses* ou des *traits syntaxiques*²⁶⁵. Les traits sont en nombre de six : humain non prédicatif, animal, végétal, inanimé concret, locatif, nom de temps. Voici quelques exemples d'emploi :

- humain non prédicatif (*Paul adore ce film*) ;
- animal (*L'ours hiberne une partie de l'année*) ;
- végétal (*Les fleurs sont fanées*) ;
- inanimé concret (*Max a acheté un vélo*) ;
- locatif (*Luc habite à Paris*) ;

²⁶⁴ Bien que la position d'objet ne se prête à aucun paradigme, *extraire la racine carrée* ne constitue pas une unité phraséologique. Le fait que cette expression accepte, par exemple, le passif (*La racine carrée a été extraite par tous les élèves*) ou la mise en évidence (*La racine carrée, tous les élèves doivent l'extraire*) prouve qu'il n'y a pas de figement, mais qu'il s'agit d'une restriction de sélection très sévère.

²⁶⁵ L'utilisation des traits syntaxiques s'est généralisée avec la grammaire générative. En ce qui concerne le modèle de description de la langue dans lequel s'inscrit notre recherche, M. Gross s'en servait aussi dans ses tables de verbes, ce qui prouve que la sémantique n'est pas totalement absente de ses travaux. De nos jours, les dictionnaires en font usage de manière très générale et assez arbitraire pour organiser la microstructure des entrées.

- temps (*La coupure d'électricité a duré une demi-journée*).

d) Des prédicats à spectre argumental réduit : des prédicats que la notion d'hyperclasse ne permet pas de décrire. En effet, les traits ont deux inconvénients majeurs (G. Gross, 1994a, 1995, 2017). D'une part, leur nombre est très restreint et leur caractère est trop vague. D'autre part, souvent, ces traits ne suffisent pas à distinguer les différents emplois d'une forme prédicative²⁶⁶. Prenons comme exemple quelques réalisations du verbe *prendre* et leur schéma d'arguments, défini avec des traits sémantiques :

(37) a. *Max prend un café ; N0 <humain> prendre N1 <inanimé concret>*

b. *Luc prend le train ; N0 <humain> prendre N1 <inanimé concret>*

c. *Max prend l'autoroute ; <humain> prendre N1 <locatif>*

d. *Les Grecs prennent Troie ; <humain> prendre N1 <locatif>*

Les phrases (37a) et (37b) ont le même schéma d'arguments, et les phrases (37c) et (37d) en ont un autre. Or aucun des verbes *prendre* de ces phrases n'a le même sens. La description en traits est insuffisante puisqu'elle ne permet pas de distinguer ces quatre emplois différents. Dans (37a) et (37b), le complément du verbe *prendre* n'est pas un nom inanimé quelconque mais une boisson (exemple 37a) et un moyen de transport (exemple 37b). De même, les compléments appropriés de *prendre* dans (37c) et (37d) relèvent de sous-classes de locatifs comme celle des <voies> (exemple 37c) et celle des <places fortes> (exemple 37d). On a donc besoin de classes plus fines pour décrire les prédicats avec plus de rigueur et que cette description permette de prédire de manière automatique les substantifs qui peuvent se trouver en position d'argument pour un emploi prédicatif donné. Ces classes sémantiques sont les *classes d'objets*²⁶⁷. Ainsi,

²⁶⁶ Ces failles peuvent être dues au fait que les traits ont été définis de manière intuitive, à partir des connaissances du monde. Or la description de la langue devrait se faire de manière interne à la langue, avec des outils que la langue fournit elle-même.

²⁶⁷ Mathieu-Colas a proposé cette appellation au début des années 90. L'expression existait déjà en informatique (en programmation orientée objet) pour désigner une notion qui a une certaine relation avec nos classes d'objet. La programmation orientée objet (POO) consiste à modéliser un ensemble d'éléments d'une partie du monde réel (appelés *domaine*) en un ensemble d'entités informatiques. Une *classe* est un ensemble d'informations et de mécanismes reliés à un sujet. Un *objet* est un produit (un conteneur

« une classe d'objets est un ensemble de substantifs, sémantiquement homogènes, qui détermine une rupture d'interprétation d'un prédicat donné, en délimitant un emploi spécifique. Un prédicat a autant d'emplois différents qu'il a de classes d'objets différentes en position argumentale » (G. Gross, 2015, p. 20). Cette définition implique, d'une part, que les classes d'objets sont construites sur des bases syntaxiques, et d'autre part, qu'elles sont induites par la signification du prédicat. Les classes d'objets constituent des sous-ensembles des hyperclasses.

4.2. L'élaboration des classes d'arguments²⁶⁸ : les opérateurs appropriés

Le Pesant et Mathieu-Colas (1998) signalent que « les classes d'objets, considérées en tant que classes d'arguments, se définissent *par relation avec*²⁶⁹ les prédicats qui leur sont spécifiques » (p. 12). On considère que les arguments ne sont pas fondamentalement définis par leurs traits sémantiques inhérents, mais par les prédicats qui leur sont appropriés. Les classes d'objets sont donc fondées sur la combinatoire des prédicats et leurs arguments. Pour établir les classes d'arguments il est nécessaire de repérer les prédicats qui leur sont spécifiques. On les appelle des *opérateurs appropriés*²⁷⁰. Il s'agit d'une sorte de « faisceau définitionnel » qui sert à définir une classe d'objets donnée. Ce groupe d'opérateurs appropriés doit être économique et pertinent. Dans le meilleur des cas, un seul prédicat suffira à définir une classe, par exemple *épeler* <mot>. Seulement les substantifs qui peuvent apparaître comme

d'information) qui sort d'un « moule » (la classe), et duquel il hérite les comportements et les caractéristiques. On notera que cette notion est proche de notre « version » linguistique de la classe d'objets. D'ailleurs, ce terme existait déjà en lexicologie, pour la définition des concepts. La notion de classe d'objets proposée au LLI est différente, d'une part, parce qu'elle désigne des classes d'unités lexicales, et non des entités extralinguistiques, et d'autre part, parce que la définition de ces classes d'objets prend en compte des propriétés syntagmatiques (Mathieu-Colas, 2007).

²⁶⁸ Devrait-on dire « classes d'emplois argumentaux » ? Pour simplifier, on utilise le terme argument, mais on fait référence à chaque emploi.

²⁶⁹ En italiques dans le texte d'origine.

²⁷⁰ On les appelle des *opérateurs appropriés*, par opposition aux *opérateurs généraux*, qui caractérisent tous les substantifs qui partagent le même trait syntacticosémantique.

complément de ce verbe appartiennent à cette classe²⁷¹. Cependant, souvent, ces opérateurs ne sont appropriés à une classe que par l'un de leurs emplois. Donc la polysémie des prédicats rend nécessaire le recours à au moins deux ou trois verbes pour constituer une classe d'arguments. Prenons l'exemple des noms de vêtements (Le Pesant et Mathieu-Colas, 1998, p. 13) :

- (38) a. *Luc a mis (son pantalon, son costume, ses sandales...).*
b. *Luc était en (pantalon, costume, sandales...).*
c. *(Ce pantalon, ce costume, ces bermuda) lui va bien.*

Aucun des verbes utilisés dans les phrases (38) n'est spécifique à la classe des <vêtements> :

- (39) a. *Luc a mis le chauffage ; Luc a mis la télé ; Luc a mis deux heures pour...*
b. *Luc était en forme ; Luc était en vacances ; Luc était en Angleterre.*
c. *Les cheveux courts lui vont bien ; La colère lui va bien ; Cette décision lui va bien.*

C'est l'association des trois verbes prédicatifs qui permet de délimiter la classe des <vêtements>²⁷². Autrement dit, cette classe est fondée sur des critères strictement syntaxiques.

Nous tenons à mettre l'accent sur le fait que les opérateurs appropriés à une classe donnée ne sont pas forcément les plus fréquents en cooccurrences. Il ne s'agit donc pas d'une question de fréquence ou de statistique, mais de syntaxe. Par exemple, parmi les occurrences verbales les plus fréquentes des substantifs appartenant à la classe <arbres> se trouvent (nous les classons par ordre de fréquence) : *planter, abattre, tomber, couper, arracher, tailler, déraciner, émonder, ébrancher*. On peut bien comprendre qu'aucun des verbes les plus fréquents n'est approprié à la classe des <arbres>, s'agissant de verbes à fort caractère polysémique. C'est justement les deux

²⁷¹ D'autres exemples sont *ressemeler* <chaussure> ou *rédigier* <texte>.

²⁷² Le Pesant et Mathieu-Colas (1998, p. 13) attirent l'attention sur d'autres noms appartenant au voisinage de la classe des <vêtements>, tels que *lunettes, montre, bague*, etc., Ils se combinent bien avec le verbe *mettre* ; plus ou moins bien avec le verbe *aller* ; mais pas avec *être en* (**être en lunettes ; *être en montre...*).

derniers verbes de cette liste qui définissent la classe des <arbres> de manière plus exacte (G. Gross, 2012).

Force est de constater que la description de la combinatoire ne doit pas se limiter à ce nombre minimum de prédicats qui ont permis de délimiter une classe d'objets donnée. Le traitement automatique de la langue exige des descriptions exhaustives. Donc, il s'avère nécessaire de recenser en extension tous les prédicats qui sélectionnent cette classe d'objets.

Ainsi, les classes d'objets servent à factoriser les arguments de manière strictement linguistique et non conceptuelle. Ensuite, la définition en extension²⁷³ des classes d'objets permettra de reconnaître et de générer toutes les phrases possibles relevant du même emploi.

4.3. La hiérarchisation des classes d'objets

Une description fine des ensembles lexicaux nécessite souvent une subdivision des classes en sous-classes, c'est-à-dire une hiérarchisation en hypéronymes et hyponymes. La notion de classe d'objets redéfinit la relation d'hypéronymie/hyponymie. Cette hiérarchisation du lexique n'est pas fondée sur une conception ontologique, comme c'est le cas des arbres sémantiques²⁷⁴, qui, selon G. Gross (2008) « n'ont aucune base syntaxique et constituent des ensembles vagues sans valeur opérationnelle sur le plan linguistique » (p.121). En effet, les arbres sémantiques

²⁷³ Le terme *extension* désigne l'ensemble des entités lexicales qui constituent une classe d'objets donnée ; par exemple, l'extension d'*autobus* est définie par l'ensemble d'unités lexicales qui appartiennent à la même classe d'objets qu'*autobus* : *métro*, *tramway*, *car*, etc. L'extension permet donc de définir une classe d'objets par l'énumération des unités lexicales qui la constituent (Neveu, 2011, p. 154). En somme, l'extension est le paradigme d'une classe d'objets.

²⁷⁴ L'un des arbres sémantiques le plus sophistiqués a été développé dans le dictionnaire électronique du système de traduction automatique SYSTRAN. Ce dernier utilise des classes sémantiques (*semcat*) pour la résolution de certaines ambiguïtés. SYSTRAN est un système de traduction automatique développée dans les années 70. Il connaît un grand succès grâce à sa richesse (plus de 40 langues), sa remarquable performance et à sa gratuité. Il est utilisé par Google, AOL et AltaVista, entre autres. Il est disponible sur <http://www.systransoft.com/>.

ou d'autres types de réseaux lexicaux (p. ex., WordNet²⁷⁵) fondent la hiérarchisation du lexique sur des observations psycholinguistiques, c'est-à-dire à partir du découpage du monde en catégories fondamentales. Cependant, notre conception de l'univers n'est pas forcément corrélée avec l'expression langagière. Soit la classe d'objets <vin>. Elle est définie par des prédicats verbaux appropriés comme *madériser* ou des prédicats adjectivaux comme *gouleyant* ou *charnu*. Or, la sous-classification en types de vin (*vin rouge*, *vin blanc*, *vin rosé*, *vin vert*, *vin jaune*) est référentielle et non linguistique, parce que chacun de ces types de vin n'a pas ses propres opérateurs appropriés différents aux opérateurs de la classe <vin> (G. Gross, 2004b).

L'approche des classes d'objets est rigoureusement linguistique : la hiérarchisation des classes d'objets est fondée sur l'observation dans des corpus des constructions syntaxiques et des combinaisons lexicales possibles. Les classes d'objets ne sont pas fondées sur des relations sémantiques internes (synonymie ou hypéronymie) ni sur des classements ontologiques, mais sur la combinatoire des éléments dans le cadre de la phrase. Selon G. Gross et Clas (1997), « la démarche part de la langue et non d'arbres prédéterminés qui découpent l'univers selon la représentation mentale du concepteur » (p.148).

Les méthodes d'élaboration de classements hiérarchiques étant différentes, les résultats et applications sont différents aussi. Ainsi, les arbres sémantiques sont moins performants que les classes d'objets pour le traitement automatique des langues naturelles. D'une part, ils ne prennent pas en considération les informations syntagmatiques ; alors, ils restent essentiellement paradigmatiques, contrairement aux classes d'objets (Le Pesant et Mathieu-Colas, 1998). D'autre part, une représentation arborescente du lexique ne reflète pas la complexité de celui-ci, car certaines unités lexicales peuvent dériver de différentes tranches de l'arborescence ; par exemple l'unité polylexicale *cotte de maille* est à la fois un <vêtement>, puisqu'on peut la *mettre* ou on peut *être en cotte de maille*, et un <objet métallique>, puisqu'elle peut *rouiller* (G. Gross, 1994b ; G. Gross et Clas, 1997). Comme le propose Le Pesant (1994, P. 39), il

²⁷⁵ WordNet est une énorme base de données lexicale qui contient une description rigoureuse des relations sémantiques entre les mots (synonymie, antonymie, hypéronymie, hyponymie, troponymie, méronymie, etc.). C'est l'un des réseaux sémantiques les plus riches et l'un des plus réussis. Nous reviendrons sur cette base de données dans la section 3.2. *Ressources pour l'étiquetage sémantique* du Chap. V.

serait donc plus convenable de concevoir les relations lexicales d'hypéronymie/hyponymie comme un treillis.

4.4. La notion d'héritage

La représentation conceptuelle du monde se compose d'un ensemble de concepts hiérarchisés. Il est évident que certaines relations conceptuelles d'hypéronymie et d'hyponymie sont inscrites dans la langue et parfois les représentations mentales coïncident avec la classification fondée sur des propriétés linguistiques. C'est le cas, par exemple, de la définition notionnelle : *L'Airbus est un avion et un avion est un moyen de transport* (Le Pesant et Mathieu-Colas, 1998, p. 15). L'un des emplois du verbe *prendre* est approprié à la classe des <moyens de transport> (*prendre un train, un bateau, un avion*). Ce verbe est transmis par héritage à ses hyponymes (*prendre un TGV, un paquebot, un Airbus*). D'autres opérateurs appropriés de cette classe sont : *monter à, descendre de, voyager en, aller en*. Dans un niveau immédiatement inférieur (sous-classe), d'autres prédicats appropriés caractérisent linguistiquement l'*Airbus* comme appartenant à la classe <*avion*> : *piloter, décoller, atterrir*. Les sous-classes héritent des propriétés syntaxiques et sémantiques des classes supérieures. Le regroupement qu'on peut faire grâce à la mise en place des relations d'héritage correspond à des factorisations syntaxiques, et non pas à des arborescences ontologiques, qui, d'ailleurs, sont très difficiles à contrôler par l'automate (Massoussi et Sfar, 2009). Cette notion d'héritage est donc une conséquence directe de la notion de hiérarchie en hypéronymes et hyponymes ci-dessus. Pour mettre au point ces héritages, il faut partir du niveau le plus élevé (dans le cas de notre exemple, de la notion de <concret>). Ce travail descriptif est surtout un problème de linguistique, plus que de formalisation informatique. L'avantage de la notion d'héritage est que les éléments sont remplacés par leur classe, ce qui représente une grande simplification de la description (G. Gross, 2004b).

Encore, il existe des classes sémantiques complexes, rattachées à plusieurs arborescences. Ces classes héritent les propriétés de toutes les classes supérieures. Elles reçoivent donc des *héritages multiples*. Par exemple, le mot *cheval* appartenant à la

classe des <équidés> et, par la suite, à la classe des <mammifères>²⁷⁶, il a des opérateurs tels que *pouliner*, *mettre bas*. Mais il peut appartenir aussi à la classe des <animaux de trait> et, dans ce cas, il est cooccurrent des verbes tels qu'*atteler* ou *harnacher*. De même, il peut appartenir à la classe des <animaux de course>, accompagné par les opérateurs *monter*, *entraîner*, *jouer sur* ou *miser sur*. Comme l'argument est défini par ses prédicats appropriés, un même mot peut appartenir à autant de classes qu'il est défini par des ensembles prédictifs différents (G. Gross, 2004b).

En somme, la création de classes d'arguments est exclusivement fondée sur des critères linguistiques, c'est-à-dire tous les substantifs appartenant à une classe doivent avoir les mêmes propriétés sémantiques et syntaxiques. En ce qui concerne le traitement automatique, l'avantage des classes d'objets (et des sous-classes et des hyper-classes) est de pouvoir factoriser et décrire de manière compacte tous les substantifs d'une classe donnée. Finalement, si l'on associe à chaque substantif le code de la classe à laquelle il appartient et que l'on décrit les schémas d'arguments des prédicats moyennant ces classes, l'automate sera capable de reconnaître et de générer toutes les phrases possibles d'un emploi donné.

5. La description des actualisateurs

Les actualisateurs sont les autres constituants de la phrase, avec les prédicats et les arguments. Rappelons que l'actualisation n'est pas un procédé référentiel, mais un simple procédé grammatical²⁷⁷. Selon Sfar (2010) :

Les actualisateurs sont des éléments qui permettent d'inscrire les prédicats et les arguments dans l'énonciation en véhiculant les différentes catégories grammaticales de la personne, du genre, du nombre, du temps, de la diathèse, de la modalité et de l'aspect. Ils peuvent être de nature grammaticale (préposition, déterminant, flexion verbale, nominale ou adjectivale, etc.) ou lexicale (adjectifs modificateurs, adverbes, morphèmes dérivationnels, etc. (p. 101).

²⁷⁶ Nous tenons à rappeler que cette sous-classification est fondée sur des critères linguistiques (prédicats appropriés), non référentiels.

²⁷⁷ La notion d'actualisation a déjà été abordée dans ce chapitre, dans les sections 2.1. *La notion de phrase élémentaire* et 2.2. *L'organisation du lexique en trois fonctions primaires*.

La fonction essentielle de l'actualisation est donc d'assurer la bonne formation grammaticale et syntaxique de la phrase.

Dans cette section, nous survolerons les différences entre le mode d'actualisation des arguments et celui des prédicats. Ensuite, nous nous attarderons sur les deux éléments principaux d'actualisation des prédicats nominaux : les verbes supports et la détermination. La tradition grammaticale a toujours défini les verbes supports par leur caractère arbitraire et imprédictible. Nous tenterons de définir ces éléments linguistiques par leurs propriétés syntaxiques et nous en proposerons une typologie. D'autre part, nous montrerons que, malgré les difficultés que la détermination pose, du fait de sa richesse et variété, les classes d'objets se sont avérées un outil efficace pour décrire le phénomène d'actualisation.

5.1. L'actualisation prédicative vs l'actualisation argumentale

L'actualisation des prédicats diffère de celle des arguments. Dans le premier cas, les occurrences d'actualisation dépendent de la nature des emplois prédicatifs. Par exemple, parmi les noms prédicatifs des <aides financières>, les prédicats *soutien financier* et *allègement fiscal* sélectionnent le verbe support *accorder*, alors que le verbe support *pratiquer* n'est sélectionné que par *allègement* :

(40) a. *L'État (accorde + *pratique) son soutien financier aux agriculteurs.*

b. *L'État (accorde + pratique) un allègement fiscal aux agriculteurs*

Dans le cas de l'actualisation des arguments, elle dépend des relations entre les prédicats et les arguments. Ainsi, dans la phrase (40a) ci-dessus le prédicat nominal est actualisé par un adjectif possessif, et il pourrait être actualisé aussi par un déterminant indéfini. Par contre, dans la phrase (40b) le possessif ne serait pas accepté :

(41) a. *L'État accorde (un + son) soutien financier aux agriculteurs.*

b. *L'État pratique (un + *son) allègement fiscal aux agriculteurs.*

L'actualisation des arguments est prise en charge par la détermination et les prépositions. Par contre, l'actualisation des prédicats est bien plus complexe car elle sollicite des moyens linguistiques différents en fonction de la nature morphologique de celui-ci (qu'il s'agisse d'un verbe, d'un nom ou d'un adjectif). D'autant plus que l'actualisation des prédicats comprend à la fois des informations de nature temporelle

(la conjugaison) et des informations sur les propriétés sémantiques internes ou externes aux prédicats (l'aspect). La conjugaison sert à actualiser les prédicats verbaux. Elle porte les marques de la personne, le nombre, le temps, l'aspect, la modalité, etc. Les prédicats nominaux sont actualisés par la détermination et par les verbes supports, et éventuellement par certains adjectifs et certains adverbes (cf. G. Gross, 1996b et 2017). Ceux derniers permettent aux prédicats nominaux de porter toutes les marques propres à la conjugaison. Ils servent à « conjuguer » les prédicats nominaux. Enfin, les prédicats adjectivaux sont actualisés par le verbe *être*.

5.2. L'actualisation des prédicats nominaux

L'actualisation des prédicats nominaux est plus complexe que celle des prédicats verbaux. Le fait qu'elle soit de nature lexicale la rend plus difficile à décrire. De plus, l'actualisation des prédicats nominaux est assurée par des éléments appartenant à différentes parties du discours : des verbes, des déterminants, des adjectifs et des adverbes. Malgré les difficultés que cette actualisation pose du fait de sa richesse et hétérogénéité, les classes d'objets se sont avérées un outil efficace pour décrire le phénomène de l'actualisation.

À continuation, nous décrirons ces modes d'actualisation nominale ayant joué un rôle essentiel dans la description des prédicats nominaux des <aides>. D'abord, nous nous attarderons sur la description des verbes supports : nous définirons les principales propriétés syntaxiques qui caractérisent les verbes supports et qui permettent aussi de les différencier des verbes prédicatifs, et nous présenterons les différents types et les fonctionnalités des verbes supports, en accordant une spéciale attention aux verbes supports appropriés et aux verbes supports aspectuels, car leur utilisation a été spécialement utile dans l'élaboration des sous-classes syntacticosémantiques des prédicats nominaux des <aides> et eds <aides financières>. Enfin, nous survolerons les particularités de la détermination et la manière de les décrire moyennant les classes s'objets.

5.2.1. L'actualisation moyennant le verbe support

La notion de verbe support a été introduite par Harris (1964/1970) afin de traiter les relations transformationnelles de nominalisation²⁷⁸. Il décrivait ces verbes comme des verbes opérant sur des nominalisations (*verbs operators* ou *nominalized verbs*)²⁷⁹. Par la suite, la notion de verbe support a été développée par l'école de M. Gross, le lexique-grammaire. Cette notion est désormais bien établie et sa littérature est très riche²⁸⁰.

Vivès (1993) les définit comme des « verbes vides ou vidés de leur sens lexical d'origine (c'est le sens lexical qui opère la sélection des arguments), et n'ont pour rôle que d'actualiser, dans une phrase élémentaire, un terme prédicatif n'appartenant pas à la catégorie du verbe » (p.10). Selon cet auteur, un verbe donné n'est pas support par définition, mais il a un emploi en tant que support lorsqu'il entre en relation avec un prédicat nominal (Vivès, 1983). G. Gross (2017) a observé qu'il existe certains verbes dont la seule fonction est d'être supports, c'est-à-dire ils n'ont pas d'autres emplois. Tel est le cas des verbes *effectuer, intimer, procéder, perpétrer, advenir, avoir lieu* et *se produire*. Néanmoins, la plupart des verbes supports correspondent morphologiquement à des emplois d'un verbe prédicatif (G. Gross, 2017), comme *faire (du pain/un rêve)* ou *avoir (une fortune/des défauts)*²⁸¹. Dans ces exemples, le premier nom forme, avec le verbe, une phrase verbale (verbe prédicatif) et le second, une construction à prédicat nominal (verbe support).

²⁷⁸ Les relations transformationnelles sont des relations d'équivalence entre deux phrases. Elles ont été largement développées par Harris. En français, c'est M. Gross (1977) qui a décrit et développé cette notion. Selon Harris, la nominalisation (transformation morphologique) n'est pas la transformation d'une phrase en un groupe nominal, mais la transformation d'une phrase en une autre phrase. La nominalisation est donc une phrase élémentaire.

²⁷⁹ Ensuite Daladier (1978) a forgé le terme *verbes supports*.

²⁸⁰ Dans le cadre du lexique-grammaire, voir Daladier (1978) ; Giry-Schneider (1987) ; G. Gross (1993b, 1998, 2017) ; G. Gross et Vivès (1986) ; M. Gross (1981, 1993) ; Ibrahim (1996) ; Vivès (1993). Pour consulter quelques travaux concrets sur un verbe support donné, voir Giry-Schneider (1978b) ; G. Gross (1989, 1993b) ; G. Gross et Vivès (1986) ; M. Gross (1981, 1995) ; Vivès (1983). Pour un autre type de formalisation de l'étude des constructions à verbe support, voir Mel'čuk (2004d) ; Mel'čuk et al. (1984, 1988, 1992, 1999). Sur l'espagnol, voir Alonso Ramos (1998, 2001a, 2001b, 2004b) ; Blanco (2000), Bosque (2001a), Herrero Ingelmo (2004) ou Mendivil (1999), entre autres.

²⁸¹ Exemples empruntés à G. Gross (2017, p. 135).

Les verbes supports sont donc de verbes vides ou vidés de sens qui, par conséquent, n'ont pas de schéma d'arguments et dont la seule fonction est d'actualiser les prédicats nominaux, c'est-à-dire de leur permettre de véhiculer des informations de type temporel, aspectuel, modal, etc. Ainsi, dans les phrases :

(42) a. *Luc a apporté de l'aide à Max.*

b. *Luc a aidé Max.*

ce n'est pas le verbe *apporter* qui sélectionne les arguments dans (42a), mais le substantif *aide*. Il sélectionne deux arguments : *Luc* et *Max*, et il sélectionne aussi le verbe support qui l'actualise. Dans cette phrase, la fonction du verbe est d'inscrire dans le temps le substantif *aide*. La particularité de l'actualisation du verbe support est qu'il véhicule « certaines informations grammaticales de par son choix lexical et, qu'en même temps, il en véhicule d'autres à travers ses désinences flexionnelles » (Blanco et Buvet, 2004, p. 337). Dans la phrase (42b) le verbe est prédicatif. On l'appelle aussi *verbe distributionnel* ou *verbe plein*. On notera que la représentation des deux phrases est la même :

(43) a. *aide (Luc, Max)* (représentation de la phrase 41a)

b. *aider (Luc, Max)* (représentation de la phrase 41b)

Rappelons que l'actualisation ne fait pas partie du schéma prédicatif. Alors, le verbe *accorder*, n'apparaissant pas dans le schéma prédicatif (43a), il fait partie de l'actualisation.

Étant donné la dépendance du verbe support par rapport au nom qu'il actualise, il ne peut pas être traduit directement en langue d'arrivée. Le verbe support doit être calculé à partir de la combinatoire lexicale restreinte du prédicat nominal. Autrement dit, la traduction d'une construction à verbe support doit se faire à partir du prédicat nominal. Ensuite, il faudra chercher en langue d'arrivée le verbe support qui exprime l'information grammaticale visée (voire l'information sémantique) pour le prédicat nominal en question. Soit la phrase :

(44) a. *L'astronaute fit quelques pas sur la Lune.*

La traduction littérale du verbe support donnerait en espagnol une phrase peu naturelle :

b. **El astronauta hizo unos pasos sobre la Luna.*

Pour obtenir une bonne traduction, il faut donc partir du prédicat nominal (*paso*) et chercher le verbe support sélectionné par celui-ci en langue d'arrivée :

c. *El astronauta dio unos pasos sobre la Luna.*

5.2.1.1. Propriétés syntaxiques du verbe support

Les propriétés syntaxiques des verbes supports nous permettent de les distinguer des verbes prédicatifs. G. Gross et Vivès (1986) signalent que ces propriétés sont au nombre de trois : l'effacement du verbe support, la double analyse du complément *N Prép N* et les contraintes sur la détermination du nom supporté. Selon ces auteurs, « aucune [de ces propriétés] n'est, isolément, à la fois nécessaire ni suffisante » (G. Gross et Vivès, 1986, p. 14). Nous y ajouterons une autre propriété syntaxique très communément utilisée pour distinguer les verbes supports des verbes prédicatifs : l'absence de nominalisation du verbe support²⁸².

5.2.1.1.1. Effacement du verbe support²⁸³

Le verbe support n'ayant pas de fonction prédicative, il peut être supprimé de la phrase sans que celle-ci perde son statut de phrase et sans que le sens change. La phrase prendra donc la forme d'un groupe nominal, sans actualisation. Cet effacement se fait par l'intermédiaire d'une thématization avec relatif. Ainsi :

(45) a. *Luc a apporté de l'aide à Max.*

b. *L'aide que Luc a apporté à Max.*

c. *L'aide de Luc à Max.*

Dans (45c) le prédicat *aide* n'est pas inscrit dans le temps. Observons maintenant quel serait le résultat de cette suite de transformations avec un verbe prédicatif :

²⁸² Il existe d'autres propriétés utilisées comme test pour distinguer les verbes supports des verbes prédicatifs, par exemple, la pronominalisation du nom supporté, l'interrogation par *Que*, la paraphrase par un verbe morphologiquement associé au nom prédicatif, etc. Nous ne les décrivons pas ici parce que souvent dans le cas des noms prédicatifs des <aides> ces propriétés ne permettent pas de distinguer les emplois prédicatifs des emplois en tant que verbe support. Ces tests sont donc peu pertinents dans le cas de notre étude.

²⁸³ Dans le cadre du lexique-grammaire, cette transformation est connue aussi sur l'appellation *réduction* du verbe support ([Réd Vsup]).

- (46)a. *Luc a apporté un livre à Max.*
 b. *Le livre que Luc a apporté à Max.*
 c. *?Le livre de Luc à Max*²⁸⁴.

Dans ce cas, l'effacement du verbe *apporter* a impliqué la suppression de la prédication et, par la suite, la suppression de la phrase. Dans (46c) il n'y a qu'une succession de syntagmes, mais rien ne les relie²⁸⁵. Par contre, dans (45c), le prédicat nominal *aide* met en relation *Luc* et *Max*. On notera aussi que dans (46c) on a perdu de l'information.

5.2.1.1.2. Double analyse du complément *N Prep N*

M. Gross (1977) a montré que les groupes nominaux complexes *N Prep N* peuvent être analysés de deux façons : en tant qu'un seul constituant ou comme deux constituants distincts. Giry-Schneider (1978b) a également constaté que cette double analyse est applicable à un grand nombre de constructions à verbe support et à un nombre très important de verbes constituant des extensions aspectuelles inchoatives et duratives. D'ailleurs c'est elle qui propose que la double analyse constitue une propriété définitoire du verbe support. Observons les phrases suivantes²⁸⁶ :

- (47)*Luc commet une agression contre Max.*
 (48)*Luc raconte une agression contre Max.*

Les phrases (47) et (48) partagent la même structure en termes de relations syntaxiques et de restrictions de sélection, et leur contenu lexical est presque identique (à l'exception du verbe). Pourtant, si on y applique des transformations qui servent à déplacer le syntagme nominal (l'extraction par *C'est...que*, la relativation, la passivation, etc.), les résultats sont différents. La phrase (47) accepte deux possibilités d'extraction, c'est-à

²⁸⁴ Notons que la plupart des phrases portant la marque ? pourraient être acceptables, selon les interprétations. G. Gross (2012) signale que dans le discours, certains facteurs comme, par exemple, le fait de vouloir éviter la redondance dans la coordination ou la comparaison, pourraient entraîner la suppression du prédicat. Ce prédicat peut être rétabli grâce au contexte linguistique. Notre propos n'est pas de signaler ce qui est grammatical et ce qui ne l'est pas, mais de mettre en relief les différences syntaxiques entre les verbes supports et les verbes prédicatifs.

²⁸⁵ Rappelons qu'un prédicat est, à la base, une relation.

²⁸⁶ Nous avons emprunté les exemples à G. Gross et Vivès (1986, p. 13).

dire l'extraction de la construction *une agression contre Max* (N Prép N) comme un seul constituant ou la dislocation en deux constituants distincts :

(47) a. *C'est une agression contre Max que commet Luc.*

(47) b. *C'est une agression que commet Luc contre Max.*

Cette double analyse prouve que *contre Max* est à la fois complément du nom *agression* dans le groupe nominal de (47a) et complément de la construction à verbe support *commettre Dé_t agression* dans (47b). Par contre, la phrase (48) n'accepte pas la dislocation parce que *raconter une agression* n'est pas une expression à verbe support :

(48) a. *C'est une agression contre Max que raconte Luc.*

(48) b. **C'est une agression que raconte Luc contre Max.*

5.2.1.1.3. Contraintes sur la détermination du nom supporté²⁸⁷

Il n'est pas possible d'adjoindre un complément de nom *de Nhum* ou un possessif au nom prédicatif :

(49)a. *Luc fait un câlin à un chien.*

b. **Luc fait le câlin de Paul à un chien.*

Il existe une contrainte de coréférence : le premier argument (agent) de *câlin* est nécessairement le sujet du verbe support. Cette contrainte est donc à la fois sémantique et syntaxique. Si nous faisons passer le test aux phrases précédentes on observera que les phrases qui comportent des constructions à verbe support n'acceptent pas le complément du nom *de Nhum* :

(50) a. *Luc commet une agression contre Max.*

b. **Luc commet l'agression de Paul contre Max.*

(51) a. *Luc raconte une agression contre Max.*

b. *Luc raconte l'agression de Paul contre Max.*

(52) a. *Luc a apporté de l'aide à Max.*

²⁸⁷ Sur la notion de détermination, voir la section 5.2.2. *L'actualisation moyennant la détermination* (dans ce chapitre).

b. *Luc a apporté l'aide de Paul à Max.*

(53) a. *Luc a apporté un livre à Max.*

b. *Luc a apporté le livre de Paul à Max.*

On notera que si la phrase (52b) est acceptable, c'est parce que le verbe n'est plus interprété comme un verbe support, il s'agit d'un emploi prédicatif du verbe *apporter*. Dans cette phrase, le verbe a une structure d'arguments (*Nhum apporte Nconcr à Nhum*) et il est sémantiquement plein (sens 'porter'). En effet, pour un grand nombre de verbes on peut distinguer un emploi comme actualisateur (ils entreraient dans des constructions à verbe support) et un emploi prédicatif (*faire un bisou / faire une maison ; dar un beso / dar un libro*).

5.2.1.1.4. Pas de nominalisation du verbe support

Le verbe support ne peut pas faire l'objet d'un changement de catégorie. Étant donné que la fonction du verbe support est de porter les marques de personne, temps, mode, aspect, etc., s'il est nominalisé, il perd la fonction d'actualisation. Observons les phrases suivantes :

(54) a. *Luc a donné un livre à Léa.*

b. *Luc lui a fait don d'un livre.*

(55) a. *Luc a donné une conférence.*

b. **Luc a fait don d'une conférence.*

Dans la phrase (54a) *donner* est prédicatif et il peut être nominalisé (54b). Pourtant, la nominalisation du verbe support de (55a) donne lieu à une phrase non acceptable (55b). En effet, quand la forme verbale a un emploi prédicatif et un emploi en tant que verbe support, elle peut être nominalisée lorsqu'elle remplit la fonction de prédicat, mais pas lorsqu'elle remplit la fonction d'actualisateur (G. Gross, 1998).

La nominalisation est l'un des tests le plus utilisés pour distinguer les verbes supports des verbes prédicatifs. Pourtant, Blanco et Buvet (2004) ont constaté que ce test ne marche pas avec des verbes supports qui véhiculent d'autres types de signification grammaticale, par exemple, des informations aspectuelles : *el retorno de la pasión* (aspect itératif) (Blanco et Buvet, 2004, p. 338).

Nous avons relevé quelques occurrences de nominalisation du verbe support *apporter* n'ayant pas de valeur aspectuelle spécifique. Nous n'en citons que quelques exemples :

- (56) a. *Le service territorial éducatif de milieu ouvert Paris Centre exerce les missions suivantes : [...] la mise en œuvre, dans l'environnement familial et social des mineurs et des jeunes majeurs, des décisions civiles et pénales, autres que les mesures de placement, le cas échéant, **apport d'aide** et de conseil à la famille du mineur suivi* (14/06/2010 ; <https://www.legifrance.gouv.fr/loda/id/LEGIARTI000039349088/2020-01-01/>).
- b. [...] *l'interface besancon.soyons-solidaires.fr permet de répondre aux besoins de chacun en favorisant les actions de solidarité par l'**apport d'aide matérielle** ou **morale*** (02/04/2020 ; <https://www.francebleu.fr/infos/societe/coronavirus-la-ville-de-besancon-lance-une-plate-forme-d-entraide-sur-internet-1585820733>) ;
- c. *La Maison-Blanche a pour sa part fait savoir que les dirigeants des Etats du G20, dont le président américain Joe Biden, restaient « extrêmement concentrés » sur la lutte contre le terrorisme ainsi que sur l'**apport d'aide humanitaire** à l'Afghanistan* (12/10/2021 ; <https://www.nouvelobs.com/afghanistan/20211012.OBS49782/l-union-europeenne-debloque-un-milliard-d-euros-d-aide-pour-l-afghanistan.html>) ;

Dans l'exemple (56c) le substantif *apport* ne remplirait pas une fonction d'actualisateur, mais de prédicat. En revanche, l'emploi prédicatif est bien moins évident dans les phrases (56a) et (56b), notamment du fait de la présence des prédicats nominaux *conseil* (56a) et [*aide*] *morale* (56b). Néanmoins, l'application du test de nominalisation demeure efficace pour l'immense majorité des verbes supports analysés dans notre dictionnaire.

Après avoir survolé les caractéristiques les plus saillantes des verbes qui servent à actualiser des prédicats nominaux, nous allons présenter les différents types de verbes supports et leur rôle dans la description des langues.

5.2.1.2. Types et fonctionnalités des verbes supports

C'est le nom prédicatif qui sélectionne les verbes supports compatibles avec le type sémantique auquel il appartient (G. Gross, 1996b). Ainsi, comme le signale G. Gross (1999) : « la conjugaison lexicale des prédicats nominaux dépend de leur signification, en quoi ils se distinguent fondamentalement des prédicats verbaux, pour lesquels il est impossible d'établir une corrélation entre le sens et la conjugaison » (p. 82). Ainsi, les verbes supports assument la conjugaison lexicale des noms prédicatifs et ils sont prévisibles à partir du nom prédicatif avec lequel ils se combinent. Selon G. Gross (1998), si l'on veut automatiser le traitement de la conjugaison des noms, au même titre que celui de la conjugaison des verbes, il faudra donner pour chaque prédicat nominal son verbe support (de même que l'on rattache à un verbe un type de conjugaison donnée).

Or, il n'y pas une correspondance biunivoque entre supports et prédicats et il n'existe pas non plus un verbe support spécifique pour chaque signification grammaticale. Au lieu de faire une description au cas par cas, l'établissement de classes syntacticosémantiques de prédicats nominaux qui sélectionnent les mêmes verbes supports permet de compacter au maximum la description de la combinatoire lexicale restreinte. Blanco et Buvet (2004) manifestent la nécessité de préciser pour chaque classe syntacticosémantique de prédicats nominaux quels sont les verbes supports susceptibles d'exprimer les différentes significations grammaticales. Parfois, comme le signalent ces auteurs, la spécificité du verbe support correspondra plus à une marque diatechnique qu'à une signification grammaticale, par exemple, *elevar una propuesta*²⁸⁸, dans le domaine de l'administration. Une catégorisation des prédicats nominaux à partir des verbes supports qu'ils sélectionnent permettra à l'automate de prédire les collocations verbales. Cette catégorisation s'avère aussi très utile pour résoudre la polysémie des prédicats nominaux car la sélection des verbes supports permet de distinguer les différents emplois des noms prédicatifs.

Les classes syntacticosémantiques de prédicats nominaux sont donc établies à partir de la combinatoire lexicale restreinte. Ainsi, le choix du verbe support dépend de la classe syntacticosémantique du prédicat nominal à deux niveaux (G. Gross, 2009) :

²⁸⁸ Exemple emprunté à Blanco et Buvet, 2004, p. 340.

- hyperclasses : actions (*faire, effectuer*), états (*être, avoir*), évènements (*avoir lieu, se produire*) ;
- sous-classes ayant des supports spécifiques : les <méfais> prennent *commettre* ou *perpétrer* ; les prédicats des <coups> prennent *donner* (et une dizaine de variantes).

Il existe donc deux grands types de verbes supports standards (G. Gross, 1993) : les verbes supports généraux (ils se combinent avec l'ensemble d'une hyperclasse de noms prédicatifs) et les verbes supports appropriés (ils ne se combinent qu'avec des classes ou des sous-classes de noms prédicatifs).

Observons maintenant les phrases suivantes²⁸⁹ :

(57) a. *Luc entretient une amitié avec Paul.*

b. *Luc noue une amitié avec Paul.*

Ces deux phrases ont le même schéma d'arguments : *amitié (Luc, Paul)*. Néanmoins, on notera qu'il existe une différence entre les deux phrases. Cette différence ne relève pas de la sémantique lexicale, mais des significations grammaticales (Blanco et Buvet, 2004). Le verbe support de la phrase (57b) apporte un plus d'information grammaticale par rapport au verbe support de la phrase (57a). En effet, le verbe *nouer*, à part les informations de personne et de temps, il est porteur aussi d'information sur le déroulement de l'action, c'est-à-dire il fournit de l'information aspectuelle. Il marque l'aspect inchoatif de l'action dénotée par le prédicat nominal. À ce sujet, Danlos (1994) affirme que la nature très diverse des informations grammaticales (personne, temps, aspect, voix, etc.) rend les verbes supports particulièrement instables dans la traduction.

À part ces deux grands groupes de verbes supports standards (généraux et appropriés), il existe des verbes supports qui apportent des informations aspectuelles (*entamer, réitérer, atteindre*), d'autres qui constituent des variantes diastratiques (*choper la grippe*), d'autres basés sur des images métaphoriques qui sont lexicalisées dans la langue (*caresser un rêve*) et d'autres qui impliquent une conversion des arguments (*accorder une aide / bénéficier d'une aide*). À continuation, nous

²⁸⁹ Exemples adaptés de Blanco et Buvet, 2004, p. 328.

présenterons ces différents types de verbes supports et leurs fonctionnalités. Nous allons notamment nous attarder sur les verbes supports appropriés et sur les verbes supports aspectuels, car leur rôle a été essentiel pour définir des sous-classes syntacticosémantiques des prédicats nominaux des <aides financières>. Nous n'aborderons pas les significations grammaticales de temps et d'espace, parce qu'elles ne sont pas pertinentes pour la description des prédicats nominaux.

5.2.1.2.1. Verbes supports généraux

Ces verbes se combinent avec un très grand nombre de noms prédicatifs, comme c'est le cas du verbe *donner*, qui a un large spectre lexical :

(58) *Luc donne (un conseil + un ordre + une gifle + un pourboire) à Max.*

Compte tenu des emplois multiples de certains verbes supports, dans le modèle des classes d'objets on considère que pour décrire les constructions à verbe support, le plus rentable est de partir d'une classification sémantique des noms prédicatifs. L'actualisation d'un nom prédicatif moyennant un verbe support donné est la preuve des restrictions de sélection qui dépendent de la classe sémantique du prédicat (Blanco, 2000). Un premier classement sépare les prédicats nominaux en trois hyperclasses : actions, états et évènements. Comme nous l'avons signalé dans le chapitre II, chacune de ces classes est rattachée à un ou deux verbes supports généraux qui les définissent : *faire* pour les actions (*faire une promenade*) ; *avoir* ou *être* pour les états (*avoir une maladie* ; *être malade*) ; et *avoir lieu* ou *se produire* pour les évènements (*le concert a eu lieu* ; *un tremblement de terre s'est produit*).

5.2.1.2.2. Verbes supports appropriés

Le caractère très général de la classification des noms prédicatifs en trois hyperclasses donne lieu à des phrases de grammaticalité douteuse. Par exemple, quant aux noms d'actions, G. Gross (2004a) signale que les prédicats nominaux de <coups> sélectionnent le verbe support *donner* ou *asséner* : (**faire, donner*) *une gifle* ; les prédicats de <méfais> prennent plus naturellement *commettre* : (**faire, commettre*) *un assassinat* ; les <cris> sont actualisés avec *pousser* : (**faire, pousser*) *un cri* ; les prédicats nominaux de <décision> sélectionnent *prendre* : (**faire, prendre*) *une décision*. En ce qui concerne l'objet de notre étude, les prédicats nominaux des <aides financières> sont des noms d'action, mais la plupart des noms qui composent la

nomenclature de notre dictionnaire ne peuvent pas être actualisés moyennant le verbe support *faire* : **faire* (*de l'aide, son soutien, une subvention, une prime*).

Alors, cette classification en hyperclasses est trop simple pour que la description du fonctionnement des collocations verbales soit utilisable pour le traitement automatique des langues naturelles. Une classification plus fine en sous-classes syntacticosémantiques permettra d'assigner aux prédicats nominaux le support adéquat. Ainsi, les verbes supports appropriés *donner* ou *asséner* pour les noms de <coups>, *commettre* pour les <crimes>, *prendre* pour les <décisions>, *pousser* pour les <cris> et *accorder* ou *apporter* pour les <aides> permettent de créer des sous-classes syntacticosémantiques de prédicats nominaux plus précises.

Les verbes supports ont souvent plusieurs emplois ; par exemple, le verbe support *commettre* est sélectionné par les prédicats nominaux de <crimes>, comme nous venons de le voir, mais aussi par d'autres classes comme les <erreurs> (*erreur, ânerie*), les <accidents> (*accident, carambolage*), les <bassesses> (*infidélité, trahison*), les <bêtises> (*stupidité, confusion*), etc. (G. Gross, 2012). On notera que tous ces noms prédictifs appartiennent à des classes sémantiquement proches, regroupant des prédicats à interprétation négative, mais ce n'est pas toujours le cas. Ainsi, le verbe *mener* est approprié à plusieurs classes très distinctes (G. Gross, 2012, p. 261) :

<combats> : *assaut, bataille, charge, guerre*

<action> : *activité, entreprise, tâche*

<bruit> : *tapage, vacarme*

<coopération> : *collaboration, partenariat*

<discussion> : *dialogue, entretien, pourparler*

<recherche> : *enquête, inspection, investigation, quête*

<politique> : *stratégie, politique, intrigue*

<représailles> : *arrestation, vengeance*

<révolte> : *rébellion, révolte, révolution*

Un seul verbe support ne permet donc pas de délimiter des sous-classes syntacticosémantiques. Les classes et sous-classes de prédicats sont constituées à partir d'une combinaison de plusieurs verbes supports (appropriés ou autres) qui n'est spécifique qu'à une classe ou sous-classe syntacticosémantique de prédicats

nominaux²⁹⁰. Par exemple, le croisement des verbes supports *mener* et *livrer* ne s'applique qu'à la classe <combats>, ou l'ensemble *émettre*, *jeter*, *lancer* et *pousser* est sélectionné par des prédicats nominaux de <cris> (G. Gross, 2012). Dans le cas des prédicats nominaux des <aides financières>, ce sont les verbes supports *accorder* et *consentir* qui définissent la classe.

Nous avons observé que de nombreux noms de cette classe se combinent avec le verbe prédicatif approprié *verser* et qu'il pourrait faire partie de ce faisceau définitionnel. Pourtant, l'étude sur corpus nous a permis d'attester la sélection lexicale de ce verbe :

(59) (*Une association, L'État*) a versé (*une allocation, bourse, une donation*
**une exonération, *un allègement fiscal*) à Max.

5.2.1.2.3. Verbes supports aspectuels

Rappelons que nous dissociions l'aspect inhérent d'un prédicat de l'aspect contextuel. Le premier correspond à l'aspect intrinsèque du prédicat et le second provient de la combinatoire des formes prédicatives avec d'autres éléments de la phrase. D'une manière générale, l'aspect interne fait partie du sémantisme du prédicat et il conditionne la syntaxe du prédicat. Par conséquent, l'aspect contextuel d'un prédicat sera déterminé par son aspect inhérent²⁹¹.

L'aspect contextuel n'est pas seulement pris en charge par le verbe support. Il y a d'autres éléments dans la phrase qui apportent une valeur aspectuelle, tels que les déterminants (p. ex., *une amorce de* ou *un début de* marquent l'aspect inchoatif), les adjectifs (p. ex. dans *Luc pousse de continuel soupirs*²⁹² l'adjectif marque l'aspect

²⁹⁰ Rappelons que nous avons déjà évoqué la notion de « faisceau définitionnel » pour l'élaboration de classes d'objets d'arguments (section 4.2. *L'élaboration des classes d'arguments : les opérateurs appropriés*, dans ce chapitre). Dans le cas des arguments, c'est à partir de leur combinatoire avec un ensemble de prédicats qu'on va permettre de les classer en classes d'objets.

²⁹¹ Grezka et Buvet (2007) mettent en relief la propriété de certains marqueurs aspectuels de modifier l'aspect inhérent du prédicat ; par exemple, le prédicat nominal *génie* dénote une propriété intrinsèquement durative (*Léa a du génie*), ce qui n'est pas le cas lorsqu'il est accompagné du déterminant nominal *un éclair de*, qui est un marqueur d'instantanéité (*Léa a un éclair de génie*).

²⁹² Exemple emprunté à G. Gross, 2004a, p. 350.

itératif) et les adverbes (p. ex., *de nouveau* ou *plusieurs fois* marquent l'aspect itératif aussi), d'où la complexité de la description de l'aspect dans les langues romanes²⁹³. Souvent, on a appliqué à ces langues des descriptions qui avaient été faites pour des langues slaves, dans lesquelles l'aspect a une base morphologique (G. Gross, 1996b). Par contre, dans les langues analytiques, comme l'espagnol et le français, plusieurs éléments de la phrase apportent des valeurs aspectuelles. En espagnol et en français l'aspect définitif est nécessairement analytique, c'est-à-dire il se résout dans le cadre de la phrase. L'aspect final relève de la synthèse de plusieurs éléments : le contenu sémantique du prédicat, des outils grammaticaux (les formes verbales simples ou composées, la détermination, les modifieurs) et des outils lexicaux (les prédicats, les verbes supports).

Certains verbes supports véhiculent des valeurs aspectuelles relatives au prédicat nominal qu'ils actualisent, c'est-à-dire des informations sur la nature interne du développement de l'action. Soit les exemples²⁹⁴ :

(60) a. *Estar (en guerra, en cuarentena)*

b. *Estallar (la guerra, *la cuarentena)*

L'emploi du verbe support *estar* dans (60a) est considéré standard, car il n'apporte aucune valeur sémantique nouvelle à la valeur sémantique des prédicats nominaux *guerra* et *cuarentena*. La valeur aspectuelle durative que l'on peut observer dans la phrase (60a) provient de l'aspect interne du nom prédicatif. Par contre, le verbe support *estallar* (60b) apporte une valeur inchoative au prédicat. L'exemple (60b) illustre que la distribution des verbes supports aspectuels ne dépend pas de la signification du verbe support en soi, mais elle dépend du nom prédicatif.

Nous tenons à souligner que les informations aspectuelles des verbes supports sont des valeurs collocationnelles ; autrement dit cette information peut différer en fonction du prédicat qui les sélectionne. Soit les phrases²⁹⁵ :

²⁹³ D'ailleurs, Harris (1976), dont les travaux ont inspiré ceux du lexique-grammaire, considère que l'aspect est un opérateur de niveau supérieur qui s'applique à toute la phrase.

²⁹⁴ Ces exemples ont été adaptés de Blanco, 2000, p. 5.

²⁹⁵ Ces exemples ont été empruntés à Blanco, 2000, p. 5.

- (61) a. *Caer la noche.*
b. *Caer la tarde.*

Dans (61a) le verbe support *caer* apporte une valeur aspectuelle inchoative, tandis que dans (61b), sa valeur est terminative.

Les verbes supports à valeur aspectuelle ne peuvent pas être en distribution complémentaire avec d'autres verbes supports (G. Gross, 1993b, 1996b, 2004a). Soit les phrases :

- (62) a. *Léa commence une étude sur l'aspect.*
b. *Léa entreprend une étude sur l'aspect.*
c. *Léa entame une étude sur l'aspect.*

Dans ces phrases les verbes apportent une valeur aspectuelle inchoative. Il semble difficile de percevoir une différence sémantique entre les trois verbes. Pourtant, du point de vue syntaxique, leur statut est différent. *Commencer* et *entreprendre* ne font pas partie du même paradigme que *faire* puisqu'ils peuvent apparaître ensemble dans la même phrase :

- (63) a. *Léa commence à faire une étude sur l'aspect.*
b. *Léa entreprend de faire une étude sur l'aspect.*
c. **Léa entame de faire une étude sur l'aspect.*

En effet, *commencer* et *entreprendre* n'opèrent pas sur le prédicat nominal, mais sur le verbe support *faire*. Ils ne sont donc pas des verbes supports, mais des verbes auxiliaires aspectuels. En revanche, *entamer* ne peut pas opérer sur le verbe support *faire* car il fait partie du même paradigme que *faire*. *Entamer* est donc un verbe support à valeur aspectuelle.

Il faudra prendre des précautions analogues lors de la description de l'aspect terminatif (G. Gross, 1996b, 2004a). Observons les exemples suivants :

- (64) a. *Léa (finit, achève, cesse) son travail.*
b. *Léa (interrompt, suspend) son travail.*
(65) a. *Léa (finit, achève, cesse) de faire son travail.*

b. **Léa (interrompt, suspend) de faire son travail.*

Des verbes comme *finir, achever, cesser* (64a) sont des verbes auxiliaires, tandis qu'*interrompre* ou *suspendre* (64b) sont des verbes supports à valeur aspectuelle terminative.

La catégorie de l'aspect semble donc très difficile à cerner. La littérature à ce sujet est très nombreuse, mais le manque de consensus est évident. Nous présentons ci-dessous, sous forme de liste, certaines significations aspectuelles pertinentes pour les constructions à verbe support :

- signification 'inchoatif' (*entamer* dans *entamer des négociations*) ;
- signification 'progressif' (*poursuivre* dans *poursuivre une démarche*) ;
- signification 'terminatif' (*interrompre* dans *interrompre ses études*) ;
- signification 'télique' (*atteindre* dans *atteindre la sérénité*²⁹⁶) ;
- signification 'intensif' (*mourir* dans *Léa meurt de faim*) ;
- signification 'atténuatif' (*esquisser* dans *Léa esquissa un geste de désespoir*) ;
- signification 'distributif' (*Le découragement s'est emparé des participants*²⁹⁷) ;
- signification itératif' (*renouveler* dans *renouveler une demande*) ;
- signification 'distributif-itératif' (*distribuer des coups*²⁹⁸).

5.2.1.2.4. Variantes diastratiques²⁹⁹

Certaines variantes des verbes supports relèvent d'une différence de niveau de langue. L'utilisation de variantes diastratiques marque le discours comme familier, soutenu, etc. Nous présentons ici quelques exemples :

(66) (*attraper/choper*) un rhume

(67) (*faire/effectuer*) un voyage

²⁹⁶ Exemple emprunté à G. Gross (2004a, p. 353).

²⁹⁷ Exemple emprunté à Blanco et Buvet (2004, p. 331).

²⁹⁸ Exemple emprunté à Blanco et Buvet (2004, p. 331).

²⁹⁹ G. Gross les appelle *variantes stylistiques* (cf. G. Gross, 1993b, 2004a, 2009, 2012, 2017).

(68) *(donner/flanquer/administrer) une gifle*

Les variantes diastratiques, à part l'utilité pragmatique évidente, elles peuvent permettre de distinguer des classes syntacticosémantiques différentes. Ainsi, en espagnol, dans l'hyperclasse des événements, le verbe support *oficiar*, appartenant à un niveau de langue soutenu, permet de distinguer une sous-classe d'évènement à caractère liturgique (Blanco, 2000) :

(69) *celebrar (una misa, una fiesta)*

(70) *oficiar (una misa, *una fiesta)*

5.2.1.2.5. Verbes supports et métaphore

Certains verbes supports sont de nature métaphorique. Il s'agit de métaphores qui sont lexicalisées dans la langue, c'est-à-dire elles ne relèvent pas de la créativité linguistique individuelle³⁰⁰. On peut évidemment les classer parmi les verbes supports appropriés. Nous signalons ici quelques exemples (G. Gross, 2004a, 2017) : *coller une gifle, lancer un regard, caresser un rêve, nourrir un sentiment, ourdir un complot*.

Dans le cas de notre étude, il existe une relation entre le champ lexical de l'argent et le liquide. Ainsi, comme nous l'avons déjà signalé, le verbe support *verser* nous a permis de distinguer la sous-classe des <aides économiques> : *verser (une allocation, une indemnité)*.

5.2.1.2.6. Verbes supports passifs

Le verbe support passif entre dans une structure qui permet la thématization de l'objet et l'on obtient ce qu'on appelle le *passif nominal*. En fait, ce n'est pas le nom qui est passif mais la phrase où il figure. G. Gross (1993a) explique ce phénomène comme une « permutation d'arguments sans qu'il y ait, dans l'opération, changement de prédicat » (p. 110). Soit les phrases suivantes :

(71) a. *Léa a donné un pourboire au serveur.*

b. *Le serveur a reçu un pourboire de la part de Léa.*

³⁰⁰ Sur les figures de style libres et les figures de style lexicalisées, voir Polguère (2008, p. 200-202).

Le prédicat nominal reste constant (*pourboire*), mais on observe que les arguments ont été inversés, comme c'est le cas du passif verbal. On ne peut pas considérer que le verbe support *recevoir* soit une variante du verbe support *donner* car l'ordre linéaire des arguments est différent. Ce type de relation entre deux constructions à verbe support est appelé une *relation de conversion*.

G. Gross (1993a) établit une série de critères qui permettent de savoir si une phrase à verbe support est la converse d'une autre :

- elle a le même substantif prédicatif ;
- les déterminants du substantif prédicatif ont la même distribution dans les deux phrases associées ;
- les arguments du prédicat nominal sont identiques et ont la même distribution dans les deux phrases associées ;
- ces arguments sont inversés ;
- l'agent est introduit par les éléments *de* ou *de la part de* ;
- les supports actualisateurs forment un couple (ici, *donner/recevoir*) ;
- les deux phrases sont synonymes ;
- en cas de prédicat déverbal, la construction converse est synonyme du passif verbal :

Paul a reçu des conseils efficaces de la part de Luc

= *Paul a été conseillé efficacement par Luc* (p.110)

Ce type de constructions à verbe support rappelle donc les constructions du passif verbal. Observons les phrases suivantes :

(72) a. *Léa a aidé Luc.*

b. *Luc a été aidé par Léa.*

(73) a. *Léa a apporté de l'aide à Luc.*

b. *Luc a reçu l'aide de Léa.*

Le premier argument des phrases actives (72a et 73a) occupe la position de sujet, mais dans les phrases passives (72b et 73b), il occupe la position de premier complément. Inversement, le premier complément des phrases actives (72a et 73a) occupe la place de sujet dans les phrases passives (72b et 73b). Pour qu'il existe deux structures synonymes (un passif verbal dans l'exemple 72b et un passif nominal dans

l'exemple 73b), il est nécessaire qu'au nom prädicatif lui corresponde un verbe transitif susceptible de passivation.

Cependant, il existe des différences entre les deux types de constructions passives. Nous présentons celles qui ont été pertinentes dans le cas de notre étude des prédicats nominaux des <aides financières> (G. Gross, 1993a) :

- a) Les prédicats sont de différente nature grammaticale. La première différence et la plus évidente est que dans les passifs verbaux, le prédicat est un verbe, tandis que dans les passifs nominaux, le prédicat est un nom.
- b) Le sujet provient de compléments différents. Pour qu'un prédicat verbal puisse entrer dans une construction passive, il doit être un verbe transitif direct³⁰¹ :

(74) a. *Léa conseille Luc.*

b. *Luc est conseillé par Léa.*

(75) a. *Léa parle de Luc.*

b. **Luc est parlé par Léa.*

Pourtant, dans le passif nominal, le sujet provient d'un objet indirect.

(76) a. *Léa donne un conseil à Luc.*

b. *Luc reçoit un conseil (de + de la part de) Léa.*

- c) Le passif nominal est moins polysémique du point de vue aspectuel. Ainsi, la phrase passive :

(77) *Luc a été giflé par Léa.*

peut avoir une valeur aspectuelle ponctuelle ou itérative. Pourtant, les paraphrases équivalentes avec une construction de passif nominal sont bien plus définies du point de vue aspectuel :

(78) a. *Luc a reçu une gifle de Léa. (aspect ponctuel)*

b. *Luc a reçu des gifles de Léa. (aspect itératif)*

³⁰¹ Sauf pour les verbes *pardonner* (*Léa pardonne à Luc → Luc est pardonné par Léa*) et *obéir* (*Léa obéit à Luc → Luc est obéi par Léa*).

Ces relations converses se présentent souvent par paires (*donner/recevoir* ; *accorder/bénéficier de*). Les études menées sur ce type de relations dans le cadre du modèle des classes d'objets ont permis d'établir une dizaine de paires « de base »³⁰². Nous signalons les plus fréquents : *donner/recevoir* ; *faire/recevoir* ; *faire/subir* ; *infliger/subir* ; *exercer/subir* ; *procéder à/faire l'objet de* (G. Gross, 2004, p. 348). Ensuite, il faudrait dresser la liste de toutes les variantes des deux verbes de chaque couple.

L'application de ces paires de converses n'est pas automatique. La description de ce phénomène nécessite une étude au cas par cas ; par exemple³⁰³ :

(79) a. *Luc donne une conférence aux étudiants.*

b. ?**Les étudiants reçoivent une conférence de Luc.*

Le nombre de prédicats nominaux concernés par le passif nominal est d'à peu près 6.000 (G. Gross, 1993a). Les passifs nominaux sont aussi nombreux que les passifs verbaux. Il s'agit donc d'un phénomène de grande ampleur³⁰⁴. Dans le cas de notre étude, nous avons retenu les verbes supports passifs suivants : *bénéficier de*, *disposer de*, *obtenir*, *percevoir*, *recevoir*, *toucher*³⁰⁵, etc.

(80) a. *L'UE apporte de l'aide à plus de 18 000 personnes affectées par les inondations d'aout à Kanhan.* (27/09/2022 ; https://www.eeas.europa.eu/delegations/guinea/1%E2%80%99union-europ%C3%A9enne-apporte-de-1%E2%80%99aide-%C3%A0-plus-de-18-000-personnes-affect%C3%A9es-par_fr?s=356)

b. *Plus de 18 000 personnes affectées par les inondations d'aout à Kanhan bénéficient de l'aide de l'UE.*

³⁰² Voir G. Gross (1982, 1989, 1993a).

³⁰³ Exemples empruntés à G. Gross (1993a, p. 13).

³⁰⁴ Pour une étude détaillée des constructions converses nominales du français, voir G. Gross (1989).

³⁰⁵ On notera que la combinatoire de *toucher* est beaucoup plus restreinte que la combinatoire des autres verbes supports passifs de cette liste.

5.2.2. L'actualisation moyennant la détermination

La détermination est l'actualisation propre aux substantifs. L'étude de la détermination peut être réduite à l'étude des contraintes sur les déterminants. Ceux-ci étant de nature très diverse d'un point de vue grammatical et sémantique, la description des déterminants s'avère une tâche ardue. Pourtant, nous montrerons que l'application du modèle des classes d'objets peut faciliter cette description.

5.2.2.1. L'hétérogénéité de la détermination

La description de la détermination constitue l'une des tâches les plus pénibles de la grammaire. Cette complexité est due, en partie, au fait que dans la tradition grammaticale il n'y a pas de consensus sur le nombre et la nature des déterminants. Pour certains, la détermination ne désigne que les unités linguistiques antéposées au nom. Pour d'autres, la détermination comprend tous les constituants du groupe nominal qui ne sont pas la tête³⁰⁶ de celui-ci (cf. Blanco et al., 2001). Ainsi, il existerait deux catégories majeures de déterminants : les *prédéterminants* (ils vont en position frontale du groupe nominal) et les *modifieurs* (des séquences qui se trouvent généralement après le nom). Cette dernière optique est le point de vue adopté dans le cadre théorique que nous avons adopté dans notre recherche³⁰⁷.

Alors, du point de vue grammatical, à part les articles, la détermination regroupe aussi d'autres parties du discours, comme des adjectifs indéfinis (*quelques*), des substantifs (*un tas de*³⁰⁸), des adverbes (*trop de*). En outre, la nature grammaticale des modifieurs³⁰⁹ est très riche aussi : des adjectifs (*des bonbons mentholés*), des substantifs (*les bonbons de ce pâtissier*), des verbes (*des bonbons à sucer*), ou même des

³⁰⁶ La tête d'un groupe nominal est « le constituant qui détermine la nature des relations entre le groupe nominal et les autres éléments phrastiques et donc qui prévaut structurellement » (Blanco, Buvet et Gaviriilidou, 2001, p. IX).

³⁰⁷ Sur la diversité de traitements de la détermination, voir Blanco et al. (2001).

³⁰⁸ On observera que dans ce cas il s'agit d'un déterminant composé : le déterminant fait appel à des unités linguistiques qui ne sont pas des actualisateurs, notamment des substantifs et des adverbes.

³⁰⁹ Les modifieurs sont aussi des déterminants composés.

modificateurs de nature phrastique (*des bonbons qui ressemblent à des fraises*³¹⁰). L'extension des formes que recouvre la notion de détermination est donc très vaste.

Du point de vue sémantique, les déterminants peuvent être définis (*la, ce*), indéfinis (*un, quelque*), comparatifs, (*autant de...que*), anaphoriques (*le, ce*), aspectuels (*une série de*), génériques (*le, un*), intensifs (*beaucoup de*) et quantifieurs (*trois, un tas de*)³¹¹ (Buvet, 1998).

Une autre complication liée à l'hétérogénéité de cette catégorie est que, même dans les langues de typologie très proche, le comportement des déterminants est très différent. En ce qui concerne les deux langues de notre étude, on notera, par exemple, que l'espagnol n'a pas d'article partitif.

5.2.2.2. La détermination et les classes d'objets

Traditionnellement, le cadre d'analyse de la détermination est restreint au groupe nominal. On ne tient donc pas compte des nombreuses contraintes des déterminants imposées par les autres éléments de la phrase. Par exemple, certains verbes supports exercent des contraintes très fortes sur la détermination du nom prédicatif. Observons les deux phrases suivantes (Buvet, 1998, p. 93)³¹² :

(81) a. *Luc fait (du + *le) judo.*

b. *Luc pratique (*du + le) judo.*

En effet, certains verbes interdisent la commutation de déterminants similaires, comme c'est le cas des exemples ci-dessus. Alors, les contraintes des déterminants relèvent autant du lexique que de la syntaxe. Les classes d'objets, étant des outils de description syntacticosémantique, s'avèrent donc de puissants descripteurs de la détermination, notamment pour la description de la détermination des noms prédicatifs. Les travaux menés à ce sujet dans le cadre du modèle des classes d'objets ont montré qu'il est possible de rendre compte des contraintes de la détermination du substantif à partir de la

³¹⁰ Tous ces exemples sont empruntés à Buvet (2008a, p. 334).

³¹¹ Dans le cas du français, il faudrait y ajouter le type partitif (*du, de la*).

³¹² Buvet (1998) a mené une étude sur la détermination des noms prédicatifs des maladies, dans laquelle il a pu démontrer que ces altérations de la détermination des noms prédicatifs imposées par les verbes supports sont très nombreuses.

classe d'objets dont celui-ci relève³¹³. Ainsi, dans l'un de ces travaux, Blanco et Buvet (1999) soutiennent :

Nous proposons de décrire les contraintes sur les déterminants en ayant recours aux classes d'objets dans la mesure où ces dernières, en tant que descripteurs exhaustifs du lexique, permettent de systématiser les observations sur les conditions d'apparition d'un déterminant en fonction des particularités syntaxiques du nom qui l'accepte. Le recours méthodique aux classes d'objets permet de rapprocher des faits de détermination alors qu'une approche au cas par cas le permettrait plus difficilement et qu'une analyse ne tenant pas compte de la dimension lexicale de la détermination l'interdit de facto. (p. 534)

En effet, l'autre intérêt que présentent les classes d'objets est qu'elles permettent de relever des comportements massifs en ce qui concerne la compatibilité entre les déterminants et les noms.

La description en classes d'objets doit être fondée sur la bipartition syntaxique des substantifs car la détermination des substantifs prédicatifs présente des contraintes différentes à celles de la détermination des substantifs argumentaux. Par exemple, la détermination d'un nom prédicatif peut indiquer des informations aspectuelles, mais l'argument ne peut pas se combiner avec ce type de déterminants. Ainsi, dans la phrase suivante (Lim, 1998) :

(82) *Luc avait un début de grippe.*

le déterminant composé *un début de* marque l'aspect inchoatif de l'état exprimé par le nom prédicatif.

D'ailleurs, les contraintes sur la détermination des prédicats sont plus fortes que celles de la détermination des arguments. Blanco et Buvet (1999) affirment que « le nombre de situations où le mode d'actualisation du prédicat influe sur la détermination des arguments serait très limité » (p. 543). La détermination des arguments, étant plus

³¹³ Pour le français, Buvet (1993, 1998, 2008a 2009c.) a mené de nombreuses recherches. En espagnol, Alonso Ramos (2001c) et Blanco (1997) ont mené de remarquables travaux. Dans une perspective contrastive ces deux auteurs ont publié aussi de nombreux travaux : Blanco (1998) ; Blanco et Buvet (1998, 1999). Il est nécessaire de signaler aussi le numéro 145 de la revue *Langages* (Buvet, 2002), consacré à la détermination. Finalement, nous voudrions mentionner aussi le projet DéTAL (Détermination du Traitement Automatique du Langage), dont l'objectif est la modélisation de la détermination en français et en espagnol et des modules de correspondance entre ces deux langues, dans le but de les implémenter dans des systèmes de traduction automatique (cf. Blanco et Buvet, 1999).

ou moins libre, elle est extrêmement difficile à décrire. Cette détermination est le plus souvent indépendante du mode d'actualisation du prédicat.

5.2.2.2. La détermination des prédicats

Comme nous l'avons signalé précédemment, la possibilité pour un nom prédicatif donné d'admettre certains déterminants et pas d'autres dépend de son environnement syntaxique. Par conséquent, la définition des emplois prédicatifs doit être une tâche préalable à la description de la détermination.

Dans le lexique-grammaire, des études portant sur les noms prédicatifs et les constructions à verbe support ont montré qu'il existe un grand nombre de contraintes de sélection sur les déterminants de ces noms (cf. Giry-Schneider (1978b, 1987) ; G. Gross (1989) ; Meunier (1981) ; Vivès (1993), entre autres). Par exemple, dans le domaine des maladies, on a constaté que les noms prédicatifs *grippe*, *hépatite* et *variole* précédés du verbe support *avoir* ne sélectionnent pas les mêmes déterminants (Blanco et Buvet, 1999)³¹⁴ :

- (83) a. *Luc avait (la + une) grippe.*
- b. *Luc avait (*la + une) hépatite.*
- c. *Luc avait (la + *une) variole. a*

Dans le cas où plusieurs verbes supports seraient possibles, il s'avère nécessaire de spécifier les éventuelles contraintes sur les déterminants imposées par chacun de ces supports. Ainsi, la détermination du nom prédicatif *peur* diffère en fonction du verbe support qui l'accompagne :

- (84) a. *Luc (a + *éprouve + prend) peur. (article zéro)*
- b. *Luc (*a + éprouve + *prend) de la peur. (article partitif)*
- c. *Luc (a + *éprouve + *prend) très peur. (intensif)*
- d. *Luc (a + éprouve + *prend) une peur bleue. (article indéfini + modifieur)*

³¹⁴ Sur la détermination des noms de maladie, voir Buvet (1998) pour le français, et Balnco et Bonell (1998) pour l'espagnol.

- e. *Luc (a + éprouve + *prend) la peur de sa vie.* (article défini + modifieur)

La détermination avec le verbe support inchoatif *prendre* est beaucoup plus restreinte que celle des verbes supports généraux *avoir* et *éprouver*. De toute façon, ces deux derniers manifestent des différences en ce qui concerne la détermination. La détermination des noms prédicatifs dépend donc de la nature sémantique et de l'aspect du nom prédicatif, et du type d'arguments qu'il sélectionne³¹⁵.

Encore, les exemples (84d) et (84e) mettent en évidence les relations de sélection qui existent entre certains déterminants et des modifieurs, comme nous l'avons signalé plus haut. Ainsi, les phrases :

- (85) a. **Luc a la peur bleue.*
b. **Luc éprouve une peur de sa vie.*

ne sont pas grammaticales. G. Gross (2012) illustre aussi des corrélations existantes entre des groupes de déterminants. À titre d'exemple, observons les phrases suivantes (G. Gross, 2012, p. 187) :

- (86) a. *Luc a de la fierté.*
b. *Luc a une grande fierté.*
c. *Luc a du courage.*
d. *Luc a un grand courage.*

Ainsi, dans les constructions des prédicats nominaux d'état avec le verbe support *avoir*, il existe une corrélation forte entre l'article partitif et le modifieur du type *un-modif*³¹⁶.

En somme, une étude systématique et exhaustive de la détermination à la lumière des classes d'objets permet de systématiser la description de la détermination et

³¹⁵ Certains auteurs affirment que les circonstants peuvent aussi exercer une influence sur la sélection des déterminants du nom prédicatif. Comme ce sujet dépasse la cadre de notre analyse, nous n'avons pas voulu y approfondir. Pour plus d'informations, voir G. Gross (2012).

³¹⁶ Cet encodage désigne les déterminants complexes composés d'un article indéfini + un modifieur, comme c'était le cas de la phrase *Luc a une peur bleue* (85a).

de compacter les informations récoltées. L'emploi de la notion de classe d'objets dans la description de la détermination constitue donc une économie importante dans la représentation des contraintes de sélection.

En outre, nous avons montré que la détermination dépend autant du prédicat nominal et que du verbe support. Les caractéristiques aspectuelles de ce dernier imposent des contraintes sur la sélection des déterminants. Alors, une telle étude de la détermination permet aussi de faire la lumière sur l'aspect des prédicats nominaux

6. La phrase complexe

Les textes sont construits autant par de phrases simples que par de phrases complexes. Ce qu'on appelle *phrase complexe* correspond à deux types de structures différentes : des subordonnées complétives et des subordonnées circonstancielles (G. Gross, 2012).

6.1. Les complétives constituent des arguments propositionnels

Il y a des phrases dont le prédicat a un argument de nature phrastique. Ce phénomène s'explique par le *principe de récursivité*, selon lequel « on peut insérer une phrase dans une autre en position argumentale » (G. Gross, 2012, p. 63). Autrement dit, une phrase élémentaire devient l'argument du prédicat de la principale³¹⁷. Ces phrases sont appelées *complétives* ou *enchâssées*. Toutes les langues ne se comportent pas de la même manière à cet égard. Dans des langues comme l'espagnol et le français pour qu'une phrase simple devienne un argument, elle doit subir des modifications³¹⁸ :

(87) a. **Je pense Léa réussira.*

b. **Je pense la réussite de Léa.*

³¹⁷ On notera que cette interprétation diffère de celle de Tesnière (1959), qui considère qu'un argument est fondamentalement un nom. Alors, quand une phrase est en position d'argument, elle acquiert corollairement le statut de nom. Par contre, selon le modèle des classes d'objets, toutes les catégories grammaticales sont syntaxiquement ambiguës. Donc le lexique ne peut pas constituer à lui seul le point de départ d'une description de la langue utile pour le traitement automatique (G. Gross, 1999, 2006).

³¹⁸ Sauf rares exceptions, comme c'est le cas de : *Je sais, Léa réussira.*

c. *Je pense que Léa réussira.*

Dans l'exemple, la conjonction *que* joue le rôle de « facteur » d'arguments. Il existe d'autres moyens linguistiques qui ont cette propriété (G. Gross, 1999, 2012) :

a) La réduction infinitive, surtout après des verbes de perception :

(88) a. **Je regarde Luc fait du jardinage chez moi.*

b. *Je regarde Luc faire du jardinage chez moi.*

Elle est possible également en cas de coréférence des sujets :

(89) a. **Je veux que Luc fait du jardinage chez moi.*

b. *Je veux que Luc fasse du jardinage chez moi.*

c. *Luc veut faire du jardinage chez moi.*

b) Le participe présent :

(90) *Je regarde Luc faisant du jardinage chez moi.*

c) La proposition relative :

(91) *Je regarde Luc qui fait du jardinage chez moi.*

d) Dans certaines conditions, la conjonction *quand* peut perdre sa valeur temporelle et elle devient une conjonction qui introduit une complétive :

(92) *J'aime quand tu montres tes sentiments.*

e) Les mots d'interrogation :

(93) *Est-ce que Luc fait du jardinage chez moi ?*

a. **Je me demande est-ce que Luc fait du jardinage chez moi.*

b. *Je me demande si Luc fait du jardinage chez moi.*

(94) *Que fais-tu ?*

a. **Je me demande que fais-tu ?*

b. *Je me demande ce que tu fais.*

f) La perte de l'inversion interrogative :

(95) *Quand Luc fera-t-il du jardinage chez moi ?*

- a. **Je me demande quand Luc fera-t-il du jardinage chez moi.*
- b. *Je me demande quand Luc fera du jardinage chez moi*

Nous considérons ces modifications comme des *indicateurs d'arguments*. Le fait qu'en français et en espagnol il soit nécessaire de « passer » par ces modifications pour qu'une phrase simple devienne un argument est très utile pour le traitement automatique de la langue. Ainsi, le recensement de ce type d'informations dans un dictionnaire électronique permettra la reconnaissance et la génération de ces complétives de manière automatique.

6.2. La subordonnée circonstancielle

La dénomination *phrase complexe* correspond notamment à des structures qui contiennent deux prédicats reliés par un autre prédicat, qui est considéré un *prédicat du second ordre*. G. Gross (2012) définit ce prédicat du second ordre comme « un élément prédicatif qui relie deux arguments de nature phrastique, c'est-à-dire contenant eux-mêmes un élément prédicatif » (p. 29). Observons les phrases suivantes :

- (96) a. *Max est parti car il devait récupérer son enfant à l'école.*
- b. *Cette année, Léa n'a pas pris ses congés payés du fait de sa maladie.*³¹⁹
- c. *La situation actuelle résulte du conflit entre les deux gouvernements.*

On notera que les prédicats du second ordre sont d'une grande variété morphologique, syntaxique et sémantique. Dans le cas des exemples précédents, la

³¹⁹ Comme cette construction contient un prédicat nominal, et qu'une phrase est composée d'un prédicat et de ses arguments, nous considérons que *du fait de sa maladie* constitue une proposition subordonnée causale avec le même statut que *du fait qu'elle est malade* (G. Gross, 2012 : 302). La tradition grammaticale a toujours opposé la notion de complément (la préposition est suivie d'un nom : *du fait de sa maladie*) à celle de subordonnée circonstancielle (la préposition est suivie d'une phrase : *du fait qu'elle est malade*) parce qu'elle a oublié qu'il existe des substantifs prédicatifs et qu'ils constituent le noyau d'une phrase autonome (*du fait de, sans peur, etc.*). « On doit donc considérer toute préposition ou locution prépositive suivie d'un substantif prédicatif comme une proposition subordonnée » (G. Gross, 2006 : 35).

fonction de prédicat du second ordre correspond à une conjonction (*car* dans 96a), à une locution prépositive (*du fait de* dans 96b) et à un verbe (*résulter* dans 96c).

Étant donné que la phrase élémentaire est l'unité minimale d'analyse, pour analyser ces phrases complexes il est nécessaire de les réduire à leurs parties constitutives, c'est-à-dire à des phrases élémentaires.

Avant de poursuivre la description des structures circonstancielles, nous tenons à ajouter deux considérations essentielles. Primo, la subordonnée circonstancielle n'est pas facultative, comme la tradition grammaticale l'a toujours affirmé (G. Gross, 2002, 2006, 2012). Le sens de la subordonnée est essentiel pour interpréter le contenu de la principale. Soit la phrase (G. Gross, 2012, p. 301) :

(97) *Paul ne part pas parce qu'il a peur, mais parce qu'il s'ennuie.*

La suppression de la subordonnée changerait complètement la nature de la principale.

Secundo, le connecteur n'appartient pas à la circonstancielle (G. Gross, 2006, 2012). Les subordonnées circonstancielles sont introduites essentiellement par une conjonction, une préposition ou une locution conjonctive ou prépositive. Traditionnellement on a considéré que cet élément connecteur appartient à la subordonnée. Pourtant, il n'y joue aucun rôle. Il s'agit d'un élément autonome. Soit la phrase complexe :

(98) a. *Paul est parti très tôt sous prétexte que demain il se lève à 6 heures.*

Si on la décompose en phrases élémentaires, on peut bien identifier :

b. *Paul est parti très tôt.*

c. *Demain il (Paul) se lève à 6 heures.*

On observera que la locution *sous prétexte que* n'appartient à aucune des phrases élémentaires mentionnées dans (98a). Par conséquent, cette phrase complexe ne comprend pas deux sous-ensembles (la principale et la subordonnée), mais trois : la principale, la subordonnée et la locution *sous prétexte que*. La fonction de cette locution est de relier la première phrase (en l'occurrence, la principale ; 98b) avec la seconde (la subordonnée ; 98c). Si un prédicat est à la base d'une relation entre plusieurs arguments, les connecteurs sont des prédicats. Les connecteurs sont donc des

prédicats du second ordre dont les arguments sont les propositions principale et subordonnée. Le schéma argumental de cette phrase serait donc le suivant :

(99) *prétexte (partir, se lever)*

6.2.1. La description des connecteurs

Une grande partie des connecteurs comportent un substantif prédicatif. Ces prédicats nominaux qui apparaissent dans des locutions constituent deux classes différentes du fait de la nature phrastique ou non phrastique de leurs arguments (G. Gross, 2006).

Soit la locution *avec l'intention de* dans la phrase suivante :

(100) *Léa est partie en vacances avec l'intention de se déconnecter du travail quotidien.*

Le substantif *intention* est un prédicat dont l'objet est *se déconnecter du travail quotidien* et le sujet n'est pas exprimé parce qu'il est le même que celui de la principale (*Léa*). Nous savons qu'une phrase est composée d'un prédicat, ses arguments et une actualisation. Le prédicat *intention* n'est pas actualisé dans la locution, mais il pourrait être actualisé au moyen du verbe support *avoir* :

(101) *Léa est partie en vacances ; elle a l'intention de se déconnecter du travail quotidien.*

Le prédicat nominal *intention* opère donc un argument non phrastique (le sujet humain coréférent) et un argument phrastique.

Observons maintenant le cas du verbe *causer* dans :

(102) a. *Les inondations ont causé des coupures d'électricité.*

Le verbe *causer* est un prédicat du second ordre qui opère deux arguments prédicatifs : *inondation* et *coupure d'électricité*. Le schéma prédicatif de la phrase est donc le suivant : *causer (inondation, coupure d'électricité)*. Ce prédicat du second ordre peut avoir la forme nominale :

b. *Les inondations ont été la cause des coupures d'électricité.*

Ces phrases peuvent subir des modifications. Par exemple, on peut thématiser l'objet grâce à la construction passive :

c. *Des coupures d'électricité ont été causées par les inondations.*

d. *Les coupures d'électricité ont été à cause des inondations.*

On pourrait même transformer la principale et obtenir :

e. *Il y a eu des coupures d'électricité, à cause des inondations.*

Donc, la locution *à cause de* n'a pas de statut propre. Il s'agit d'un prédicat nominal qui est entré dans une structure passive nominale³²⁰. Selon G. Gross (1999) : « Une locution du type *à cause de* n'est rien d'autre qu'un prédicat qui a perdu son actualisation propre et qui se range sous l'actualisation d'une phrase argument, qui devient alors la proposition principale » (p.79). Toutes les réalisations de cette racine *caus-* ont donc deux arguments phrastiques, à différence du connecteur *à l'intention de*, comme nous venons de le montrer (G. Gross 1999, 2005a, 2006, 2012).

Ce dernier exemple met en évidence l'intérêt de ne pas cloisonner l'analyse dans des catégories grammaticales ou dans la distinction entre mots lexicaux et mots grammaticaux. L'analyse en racines prédictives nous a permis de relier toutes les paraphrases que nous venons de présenter (G. Gross, 1999). En plus, il nous a permis de rendre compte de la variété morphologique de ce que l'on connaît comme des connecteurs.

D'autres conjonctions ou prépositions ne comportent pas de substantif. Elles doivent être considérées comme des prédicats également, car elles réalisent la fonction de connecteur entre deux arguments. Soit les phrases :

(103) a. *Max est arrivé avant que le train ne parte.*

b. *Max est arrivé avant le départ du train.*

Qu'il s'agisse d'une préposition ou d'une locution conjonctive, dans les deux cas c'est le même prédicat du second ordre (sous la forme *avant* ou *avant que*) et le même schéma argumental³²¹.

³²⁰ C'est le cas aussi, par exemple des formes *à l'abri de*, *à l'essai*, *aux arrêts*, etc. (cf. G. Gross, 2005a, p 122).

³²¹ D'ailleurs, dans le cadre du modèle des classes d'objets, on ne considère pas les conjonctions et les prépositions (et leurs locutions) comme des catégories différentes. Comme nous venons de le montrer, ces catégories ont la même fonction syntaxique. La différence ne tient pas à leur statut de connecteur, mais à

En somme, les connecteurs étant des prédicats, ils doivent donc être décrits avec les mêmes outils que les autres prédicats (schéma argumental³²², degré de figement, variations morphologiques, actualisations, modifications, paradigme du substantif, etc.), en tenant compte de leurs particularités en tant que prédicats du second ordre. Étant donné que les connecteurs sont d'une grande variété aux niveaux morphologique, syntaxique et sémantique, le traitement automatique de ces constructions oblige à établir une typologie sémantique des circonstancielles (cause, conséquence, finalité, manière, lieu, temps) et de dresser une classification de ce type de prédicats, car il s'agit de suites d'une grande variété morphologique, syntaxique et sémantique. Nous avons montré qu'une description exhaustive des racines prédictives permettra de relever les régularités et d'identifier et produire toutes les paraphrases possibles de manière automatique.

7. Conclusions

Le traitement automatique des langues nécessite une description de la langue cohérente et exhaustive. Nous considérons que le modèle des classes d'objets permet d'élaborer des ressources lexicales (notamment des dictionnaires électroniques et des grammaires locales) performantes en traitement automatique. Nous tenons à rappeler

la nature du complément : la forme *avant que* introduit un prédicat conjugué, tandis que la forme *avant* introduit un prédicat nominal non actualisé. Dans ce même sens, la différence entre la locution conjonctive *afin que* et la locution prépositive *afin de* s'explique par la coréférence des sujets de la principale et de la subordonnée. En effet, dans les phrases avec *afin de* le prédicat de la subordonnée perd son actualisation, en héritant celle du prédicat de la principale. Toutes ces différences sont donc des phénomènes extérieurs aux connecteurs et elles n'ont aucun effet sur leur statut de prédicat du second ordre. C'est pourquoi on considère que les prépositions et les conjonctions (de même que leurs locutions respectives) n'appartiennent pas à des catégories différentes (G. Gross, 2006).

³²² L'établissement d'une typologie sémantique de chaque type de structure circonstancielle (cause, finalité, manière, conséquence, etc.) où l'on préciserait la nature sémantique des arguments sélectionnés par les opérateurs du second ordre serait d'un énorme intérêt pour le traitement automatique (G. Gross, 1996a). Nous comprenons qu'il s'agit d'une tâche de longue haleine car le spectre argumental des prédicats du second ordre est bien moins contraint que celui des prédicats du premier ordre.

pourquoi ce modèle théorique permet une description de la langue qui s'adapte aux spécificités des automates.

Le point de départ est le postulat qui veut que la phrase simple soit l'unité minimale d'analyse, celle-ci se rapportant à une structure prédicat-argument. Nous avons démontré que le sens des mots est caractérisé en fonction des éléments contextuels. Le cadre de la phrase permet de désambigüiser les cas de polysémie et d'homonymie et de bien reconnaître le figement. Ainsi, la description du lexique ne doit pas se faire seulement sur la base de la sémantique, mais elle demande aussi une description des propriétés syntaxiques.

De cette conception de la description de la langue se dégage la notion de l'emploi prédicatif. Il s'agit d'une notion qui intègre tous les niveaux d'analyse, ce qui la rend performante en traitement automatique de la langue. La description d'un emploi réunit des propriétés morphologiques, syntaxiques et sémantiques. Ainsi, la description des langues dans le cadre de ce modèle n'est pas constituée de niveaux autonomes. En effet, nous trouvons que le lexique, la syntaxe et la sémantique n'ont pas d'autonomie fonctionnelle, c'est-à-dire il s'avère impossible de les étudier de manière indépendante les uns des autres.

Une notion capitale de ce modèle théorique est la notion de classe. Les classes sont au départ des notions sémantiques (au moins pour la définition et la dénomination des classes), mais elles sont validées par l'examen systématique des propriétés syntaxiques (le schéma d'arguments, dans le cas des classes de prédicats, et la combinatoire lexicale). Elles sont définies par le principe d'homogénéité du fonctionnement linguistique, ce qui permet d'établir des taxonomies fiables. Les classes sont donc définies de manière inductive à partir de l'analyse des propriétés linguistiques d'une unité lexicale donnée dans son contexte de réalisation, contrairement aux systèmes logicoconceptuels. Les classes syntacticosémantiques sont un outil économique pour la description du lexique, d'autant plus que la notion d'héritage linguistique entre classes et hyperclasses permet de rentabiliser les ressources linguistiques mises en œuvre.

Enfin, le modèle des classes d'objets permet de décrire le lexique d'une façon exhaustive en ayant systématiquement recours à des descripteurs normalisés. Ces descriptions peuvent donc être informatisées en constituant des dictionnaires

électroniques conçus dans la perspective du traitement automatique des langues. La spécificité des dictionnaires électroniques résultants résidera dans leur capacité de décrire tous les emplois lexicaux à l'aide d'informations linguistiques et métalinguistiques standardisées, c'est-à-dire régulières et explicites.

CHAPITRE V : LE TRAITEMENT AUTOMATIQUE DES LANGUES

Fuchs et Habert (2004) définissent l'objectif du traitement automatique des langues (désormais TAL) comme « la conception de logiciels capables de traiter de manière automatique des données exprimées dans une langue (dite "naturelle", par opposition aux langages formels de la logique mathématique) » (p. 1).

Les langues naturelles sont d'une extrême complexité. Le manque d'univocité entre la forme et le sens génère des phénomènes de polysémie et de figement. De ce fait, l'analyse automatique des textes se heurte souvent à ce qu'on appelle en TAL le *bruit* (plusieurs analyses possibles pour une seule séquence) et le *silence* (le fait de ne pas reconnaître une séquence donnée). Le principe du traitement automatique repose sur la possibilité de « traduire » tout texte dans un formalisme donné que l'automate puisse gérer, pour ensuite exécuter les tâches qu'on lui demande de faire. Les différents formalismes qui existent pour effectuer cette représentation de la langue ont une caractéristique commune et absolument nécessaire : un symbole donné a une signification unique. Alors, cette « traduction » se fait d'un système qui obéit au jeu de la polysémie et du figement (les langues naturelles) à un autre système où l'ambiguïté doit être complètement éradiquée.

Le TAL a dû surmonter d'autres barrières d'ordre plutôt psychologique : le scepticisme, voire pessimisme, de l'opinion commune, et ce même dans la communauté scientifique. Dans les années 90, G. Gross (1995) dénonce que de nombreux linguistes et informaticiens ont été découragés par l'idée qu'une description exhaustive réunissant des faits lexicaux, syntaxiques et sémantiques requerrait tant de temps qu'il ne serait pas réaliste de croire qu'on pourrait réussir à confectionner des dictionnaires électroniques qui réunissent ces caractéristiques. D'autre part, peut-être les informaticiens qui ont travaillé dans le domaine du TAL ont-ils sous-estimé la complexité des langues naturelles et n'ont-ils pas compris la place fondamentale des rapports entre le lexique, la syntaxe et la sémantique. Ainsi, on a obtenu des outils peu performants, ce qui aurait provoqué un intérêt quasi exclusif pour la traduction automatique assistée (TAO), comme si on se croyait incapables d'atteindre l'automatisation complète. L'échec du versant pratique des projets dans le domaine du TAL (des projets de traduction

automatique) a mené les spécialistes vers une vision du TAL moins utilitariste et plus scientifique³²³ (Cori, 2008). Alors, au début du XXI^e siècle, les ressources pour le TAL en français et en espagnol étaient très peu développées. Dès lors, de grands efforts ont été déployés pour développer des ressources langagières dans ces deux langues³²⁴.

Le TAL a entraîné une vraie révolution en linguistique. En effet, la linguistique a perçu le TAL comme un champ d'application de grande valeur. D'une part, l'informatique offre la possibilité de traiter un très grand nombre de données qui permettent de tester et de valider des résultats d'une théorie linguistique donnée. D'autre part, l'outil informatique exige une formalisation explicite, cohérente et opératoire des données linguistiques. Le traitement automatique pose alors un grand défi de formalisation des mécanismes de la langue pour que la machine puisse les reconnaître. En même temps, cela implique un grand avantage, voire un élan, pour le développement de modèles de description de la langue plus rigoureux et moins ontologiques, qui répondent aux exigences scientifiques de cohésion, de reproductibilité et de vérifiabilité. Comme le signale Buvet (2013), la linguistique informatique est soumise aux impératifs suivants :

- 1) observer les faits de langue étudiés pour en dégager des régularités généralisables ;
- 2) exprimer ces régularités en éliminant le flou, l'implicite, le non-dit, les évidences allant de soi ;
- 3) vérifier la cohérence de la formulation qui garantit l'objectivité et donc la reproductibilité de la démarche. (p. 196)

Ainsi, cette visée informatique va façonner les approches théoriques de description de la langue et leur méthodologie de recherche, car l'élaboration de ressources et d'applications informatiques permet de tester la validité des concepts théoriques.

Dans ce chapitre, nous allons approfondir sur deux des grands défis du traitement automatique : la polysémie et le figement. Nous allons décrire ces deux phénomènes linguistiques qui caractérisent les langues naturelles et nous allons

³²³ C'est à ce moment que naît le terme *Computational Linguistics*.

³²⁴ Pour un survol des ressources lexicales du français des dernières décennies et sur leur impact dans la recherche en TAL, voir Choi, 2022 ; Choi et al., 2023 ; Mariani et al. (2019a, 2019b).

présenter des solutions pour le traitement automatique. Ensuite, nous confronterons deux méthodes très utilisées actuellement en TAL, à savoir la méthode statistique (fondamentalement mathématique) et la méthode mixte (profitant des outils statistiques pour développer une étude de la langue de base qualitative). À continuation, nous présentons quelques ressources lexicales de TAL élaborées dans le cadre de méthodes de travail mixtes. L'intérêt de ces ressources tient au fait qu'elles sont exploitables par de nombreux logiciels de TAL (elles sont transversales) et qu'elles sont consultables « à l'œil humain ». Enfin, nous expliquerons comment on aborde le TAL dans le cadre du modèle des classes d'objets (qui est le modèle de description linguistique que nous avons choisi pour notre recherche). Nous présenterons les ressources linguistiques et informatiques nécessaires pour le système de TAL et les différentes étapes de traitement de ce système.

1. Les difficultés du TAL

Nous avons déjà signalé que les deux grands problèmes d'analyse auxquels se heurte le TAL sont la polysémie et le figement. Dans ce qui suit nous exposerons l'envergure de ces difficultés et comment le modèle des classes d'objets les affronte et les résout.

1.1. Le traitement de la polysémie

La polysémie constitue l'une des difficultés les plus ardues de la description des langues. Du point de vue théorique, de nombreuses approches se sont intéressées à la délimitation et la relation des sens d'un mot polysémique³²⁵. Son traitement dépend du degré d'abstraction de l'analyse et de la précision des discriminations en sens différents³²⁶. Par conséquent, la polysémie pose d'importants problèmes au traitement

³²⁵ En ingénierie linguistique, le domaine d'activité qui vise le développement d'outils servant à désambiguïser le sens d'un mot polysémique dans un texte donné est connu sous le nom de *word sense disambiguation* (WSD).

³²⁶ Nous n'aborderons pas l'étude des différentes approches théoriques visant la polysémie, car nous préférons focaliser notre intérêt sur les notions et les positionnements théoriques qui sont à la base de la configuration de notre dictionnaire. Néanmoins, nous proposons ici quelques références intéressantes qui pourraient résumer l'évolution des études à ce sujet : Apresjan, 1973 ; Martin, 1983 ; Lakoff, 1987 ;

automatique de la langue. D'autant plus si l'on essaie d'analyser la polysémie du prédicat, car, rappelons-le, celui-ci détient un rôle capital dans la phrase en imposant ses contraintes de sélection au reste des éléments de son environnement linguistique.

1.1.1. La notion de polysémie en TAL

La plupart des approches classiques de TAL ont considéré la polysémie des langues comme un phénomène marginal³²⁷. Souvent on a abordé le problème de la polysémie moyennant de procédures spécifiques de levée d'ambiguïté : les systèmes construits sont basés sur l'idée que les unités linguistiques sont à priori monosémiques, et puis il a été prévu de produire des analyses spécifiques dans les cas où une unité aurait plusieurs sens. Cela revient à construire autant d'analyses qu'il y a d'acceptions, ce qui implique qu'on déploie la majorité des efforts sur la tâche d'essayer de corriger après coup les analyses erronées. Il est évident que ces approches ont sous-évalué au départ le problème de la polysémie, ce qui leur a valu leur inefficacité pour le TAL (Victorri et Fuchs, 1996).

Mel'čuk, Clas et Polguère (1995) affirment que « la polysémie est un fait linguistique de base – un trait fort répandu et important de toutes les langues naturelles » (p. 156-157). En effet, presque toutes les unités lexicales sont polysémiques. Par exemple, *Le Petit Robert 2024 (version numérique)* contient plus de 60 000 mots et quelques 300 000 sens, ce qui revient à dire que chaque mot aurait en moyenne cinq sens différents³²⁸. Actuellement, il est admis que la polysémie n'est pas un accident ou un phénomène marginal, mais elle constitue une constante régulière de la sémantique. La polysémie, loin d'être un « défaut » qui nuirait l'efficacité de la

Cruse, 1986, 1995 ; Kleiber 1990, 1999 ; Pustejovsky, 1995 ; Fuchs, 1996 ; Victorri et Fuchs 1996 ; Cardiot et Habert, 1997 ; Nunberg et Zaenen, 1997 ; entre autres.

³²⁷ On a parfois atteint des positions aussi extrêmes que celle de Montague (1970), qui soutenait qu'il existe une équivalence totale entre langues naturelles et langues formelles.

³²⁸ Ce ratio pourrait être bien plus élevé si l'on considère que le *Petit Robert* est un dictionnaire d'usage général qui ne rend compte que des sens les plus fréquents.

communication, elle est considérée une qualité de la langue, qui permet aux systèmes linguistiques d'être très puissants et plus économiques³²⁹.

On peut facilement comprendre le phénomène de la polysémie de manière intuitive. Pourtant il s'avère fort difficile de trouver une définition rigoureuse qui recueille le consensus de la communauté de lexicographes et lexicologues³³⁰ (Fuchs, 1996). On n'a trouvé qu'un certain accord sur la définition de la polysémie comme un phénomène comprenant : « (i) une pluralité de sens liée à une seule forme ; (ii) des sens qui ne paraissent pas totalement disjoints, mais se trouvent unis par tel ou tel rapport » (Kleiber, 1999, p. 55).

Le terme polysémie comprend une grande diversité de phénomènes : métaphore, métonymie, méronymie, spécialisations du sens, etc. En ce qui concerne le TAL, la description et formalisation de certaines de ces formes de polysémie peuvent être systématiques, c'est-à-dire décrites par des règles générales assez simples. D'autres règles sont spécifiques d'une unité lexicale donnée, du fait de la singularité de son histoire individuelle³³¹. Entre ces deux cas, il existe un vaste éventail de phénomènes, qui dépassent le cadre des règles individuelles, mais qui ne peuvent pas non plus être décrits par des règles trop générales (Jacquet et al., 2005).

1.1.2. L'analyse automatique de la polysémie

D'habitude, la polysémie ne constitue pas un problème de communication parce que les êtres humains sont capables d'interpréter l'emploi d'un mot grâce au contexte

³²⁹ Victorri et Fuchs (1996) approfondissent sur les avantages de la polysémie pour la langue et pour la communication : elle apporte une grande souplesse à l'expression ; elle est utile pour le développement des sens figurés ; elle joue un rôle capital dans la paraphrase ; et elle est un mécanisme puissant d'évolution des langues.

³³⁰ La grande querelle étant la distinction entre polysémie et homonymie, c'est-à-dire la spécification des rapports de parenté qui justifient la rupture de sens.

³³¹ G. Gross (2010a) signale que souvent un mot dérivé créé par métonymie a des opérateurs appropriés qui lui sont spécifiques. Selon cet auteur, du moment où les différentes interprétations d'un mot polysémique n'ont pas les mêmes propriétés syntaxiques, on est face à des emplois différents, qui doivent figurer dans des entrées différentes d'un dictionnaire électronique.

linguistique et grâce aussi à la situation de communication. Pourtant, l'automate ne sera capable de rien interpréter si on ne lui fournit pas des mécanismes généraux de désambiguïsation fondés sur une description de la langue exhaustive, entièrement explicite et systématique.

Actuellement dans la communauté de l'ingénierie linguistique il existe un certain accord sur le fait que pour désambiguïser il est nécessaire de faire intervenir le contexte. Le contexte est un élément décisif pour l'interprétation du sens. En effet, souvent, les mots ne sont polysémiques que si on essaie de les appréhender isolément. Dans les phrases réelles la plupart de ces mots deviennent monosémiques. Ainsi, si les mots sont décrits dans une structure phrastique, la polysémie devient un phénomène secondaire³³². La notion d'emploi de la méthode des classes d'objets s'avère, donc, très pertinente pour résoudre le problème de la polysémie.

En conséquence, la polysémie pose une autre difficulté majeure pour les systèmes automatiques de représentation et de calcul du sens : la circularité. Comme le signalent Jacquet et al. (2005) :

Puisque la plupart des unités lexicales peuvent prendre des sens différents, leur sens dans une phrase donnée dépend du sens de cette phrase, et, bien entendu, le sens de la phrase est lui-même fonction du sens des unités qui la composent. (p. 99)

Quel que soit le cadre théorique ou le type de représentation que l'on adopte, le risque de tomber sur des cercles vicieux y est présent.

Dans des utilisations très concrètes, telles que l'extraction d'information sur des corpus spécialisés, la tâche de désambiguïsation est très légère. Même si l'on ne peut pas exclure d'autres interprétations du sens en dehors du domaine, la tendance lexicale dans les corpus spécialisés est plutôt monosémique ; tandis que si l'on travaille sur des corpus « généralistes » le calcul du sens du mot devient plus difficile.

1.1.3. Le traitement lexicographique de la polysémie

³³² Cette observation vaut pour les prédicats et pour les arguments.

La question se pose d'envisager le traitement automatique de la polysémie depuis une perspective de regroupement ou de dégroupement des sens. La lexicographie classique est fondée sur une conception fondamentalement unitaire du mot : tous les sens d'un mot sont regroupés dans un même article³³³. Cette conception pose deux types de problèmes. Premièrement, la disposition des différents sens dans l'article : disposition linéaire ou arborescente, ordre logique ou chronologique, distinction entre langue générale et langues de spécialité, etc. Deuxièmement, quel que soit le choix effectué pour la disposition des sens, on a toujours comme résultat des structures d'une énorme complexité et qui sont très différentes d'un dictionnaire à l'autre (Mathieu-Colas, 1996).

Ce modèle unitaire a connu quelques aménagements depuis les années 70 dans certains dictionnaires papier : nombreux mots polysémiques ont été dégroupés, c'est-à-dire ils sont traités comme des homographes. C'est le cas, par exemple, du *Dictionnaire du français contemporain* (1975), où *bureau*, *commander* ou *cher* se trouvent dégroupés en deux ou trois entrées. Pourtant, Mathieu-Colas (1996) dénonce que la méthode n'est pas menée jusqu'au bout parce qu'au sein de chaque entrée on retrouve encore une multiplicité d'emplois :

1. **bureau** n.m. 1° Table, munie ou non de tiroirs, dont on se sert pour écrire [...] – 2° Pièce où est installée cette table [...] – 3° Mobilier de cette pièce [...]
2. **bureau** n.m. 1° Établissement public où sont installés des services administratifs [...] – 2° Caisse d'un théâtre [...] – 3° Ensemble des employés ou des fonctionnaires qui travaillent dans une administration [...] – 4° Membres d'une assemblée, d'une association, élus pour diriger les travaux [...]. (*Dictionnaire du français contemporain*, 1975, p. 171)

On distingue deux entrées de *bureau*, mais elles sont encore subdivisées en trois ou quatre sens différents. Alors, le problème que la polysémie pose au traitement automatique n'a pas été résolu.

Le dégroupement n'est pas une tendance généralisée dans l'analyse lexicographique. On pourrait dire que les dictionnaires papier et les dictionnaires informatisés suivent chacun une tendance plus ou moins forte vers le dégroupement.

³³³ Sont exclus de ce système les homonymes, appartenant à des étymons différents (p. ex., les trois unités lexicales *baie*) ou à des familles anciennement éclatées (p. ex., les deux unités lexicales *voler*).

Prenons, par exemple, le traitement que le *Dictionnaire Larousse en ligne* accorde à *bureau* :

Bureau n.m. **1°** Meuble comportant une table pour écrire [...] - **2°** Pièce destinée au travail intellectuel ou à la réception des clients [...] - **3°** Ensemble des meubles de cette pièce. - **4°** Lieu de travail des employés d'une administration ou d'une entreprise [...] - **5°** Guichet. - **6°** Service assuré dans un bureau. - **7°** Ensemble des employés travaillant dans une administration [...] - **8°** Organe directeur d'une association [...]. (*Dictionnaire Larousse en ligne*)

On observe les mêmes emplois (sauf pour le sixième sens, qui n'apparaît pas dans le *Dictionnaire du français contemporain*), mais dans le cas du *Dictionnaire Larousse en ligne* il n'y a qu'une seule entrée de *bureau*. Cependant, dans ce même dictionnaire on a bien tenu à accorder un traitement d'homonymie à *cher* :

- 1. cher, chère** adj. **1°** Qui est l'objet d'un grand attachement [...] - **2°** Pour qui on éprouve une vive tendresse [...] - **3°** S'emploie dans les formules de politesse [...]
- 2. cher, chère** adj. **1°** Qui est d'un prix élevé [...] - **2°** Qui vend des objets ou qui fournit des services à un prix élevé - **3°** Qui exige des dépenses élevées.
- 3. cher** adv. **1°** À un prix élevé. - **2°** Au prix de grands sacrifices [...]. (*Dictionnaire Larousse en ligne*)

D'autres dictionnaires, comme le *Trésor de la Langue Française Informatisé* ou le *Petit Robert 2024*, privilégient les traitements unitaires. Voici l'article du *Petit Robert 2024* (version numérique) pour *cher* :

cher, chère adjectif et adverbe

- I. 1.** (Attribut ou épithète) Qui est aimé [...] - **2.** (Épithète, av. le nom) Dans des tournures amicales, des formules de politesse [...]. - **3.** (Attribut) Que l'on considère comme précieux [...].
- II. 1.** Qui est d'un prix élevé [...]. - **2.** Qui exige de grandes dépenses [...] - **3.** Qui fait payer un prix élevé [...] - **4. Adv.** À haut prix [...]

Certaines approches en lexicologie privilégient une présentation « continuiste » du sens (Kayser, 1987 ; Fuchs, 1988 ; Levrat, 1993). Ces lexicologues penchent pour une sorte de sémantisme de base qui s'enrichit sous l'influence du contexte qui donne

lieu à un ensemble de valeurs apparentés³³⁴. Ces études sont très intéressantes du point de vue théorique, mais elles sont difficilement exploitables du point de vue du TAL (ces modèles sont très difficiles à implémenter et à appliquer sur de gros corpus). Par ailleurs, le manque de systématisme en ce qui concerne le traitement de la polysémie, comme nous venons de le montrer, empêche que certains dictionnaires informatisés puissent être exploités par des systèmes de traitement automatique, par exemple, des grammaires locales. C'est pourquoi nous considérons que la représentation discontinue est celle qui s'accommode le mieux au traitement lexicographique automatique.

En effet, le traitement automatique s'avère bien plus effectif si chaque emploi est considéré comme un mot et qu'il constitue une entrée différente dans le dictionnaire électronique. Rappelons que la description de chaque emploi dans un dictionnaire électronique est très riche : des informations morphologiques (structure, forme, flexions, variantes graphies), des informations sémantico-syntaxiques (le schéma d'arguments, les classes d'objets), des données sémantiques (les domaines), des informations sur la combinatoire (l'actualisation), des informations paraphrastiques (les modifications du schéma prédicatif), possibles traductions en d'autres langues, etc.³³⁵ Ces caractéristiques relèvent de tous les niveaux d'analyse et elles sont intégrées : chaque niveau est dépendant des autres et il ne peut pas faire l'objet d'une étude isolée (G. Gross, 2010a). Or, toutes ces propriétés peuvent avoir des expressions linguistiques différentes selon les différents emplois d'un mot³³⁶, ce qui pourrait donner lieu à des analyses et des générations erronées ou à des descriptions tellement complexes qu'elles seraient ingérables par l'automate.

³³⁴ Rappelons les approches de Pustejovsky (1995), de Cruse (1986) et ultérieurement de Croft et Cruse (2004), qui prônent l'existence d'un sens ou concept global subsumant les différents sens (ou *facettes*, selon la terminologie de Cruse et de Croft).

³³⁵ On notera que notre conception des dictionnaires électroniques est très proche des bases de données.

³³⁶ Les différents emplois peuvent avoir, par exemple, des genres différents (*un espace / une espace*), une flexion différente (*ciels / cieux*), des types de conjugaison différente (*Il pleut*, comme verbe prédicatif défectif / *Des coups pleuvaient sur la tête de Luc*, comme verbe support conjugué), une syntaxe différente : *briser / *casser / *rompre (une révolte, une mutinerie, un soulèvement, une opposition)* (ce dernier exemple a été emprunté à Dubois, 1964, p. 14-15).

Mathieu-Colas (1996) a mis en relief les nombreux avantages que la description de chaque emploi d'un mot comporte pour le traitement automatique :

- a) les données lexicales peuvent être décomposées en paramètres indépendants, ce qui permet de faire des extractions d'informations et rend possible un jeu de requêtes beaucoup plus riche ;
- b) on gagne en précision et en simplicité sur les plans linguistique (clarté et lisibilité) et informatique (traitement automatique plus simple) ;
- c) la notation pour chaque emploi de toutes ces données que nous avons mentionnées plus haut est la garantie d'une représentation plus fine de la polysémie, ce qui facilite la levée de l'ambiguïté ;
- d) le traitement des mots composés et des locutions est plus facile, car chaque unité complexe possède sa propre entrée ;
- e) les critères de description sont purement linguistiques et reflètent directement la pratique de la langue ; tandis que la représentation unitaire possède un caractère ontologique et elle est sujette à l'interprétation des lexicographes.

Souvent le principe du dégroupement est critiqué du fait de son émiettement ou éclatement des descriptions. Pourtant, on pourrait regagner cette articulation des différents emplois d'un mot, par exemple, en ajoutant un champ où figurerait l'information relative aux liens logiques ou historiques entre les différents emplois, ou en intégrant dans la description une arborescence qui représenterait ces rapports (traditionnellement exprimés avec de chiffres ou de lettres).

Enfin, du point de vue méthodologique, le dégroupement de sens s'avère aussi plus efficace, comme le signale Mathieu-Colas (1996) :

Il s'avère plus opératoire, du point de vue linguistique comme du point de vue informatique, de commencer par rendre compte de la **diversité**³³⁷ des éléments (dégroupement maximal) avant de pouvoir décrire, avec plus de précision, les **liens** qui les unissent. Disons-le autrement : étant donné une structure à la fois une et multiple – comme l'est la polysémie –, il est techniquement plus simple de partir du multiple pour y introduire l'unité que d'effectuer l'opération inverse. (p. 325)

³³⁷ En gras dans le texte original.

En somme, en ce qui concerne notre cadre méthodologique, nous pouvons bien comprendre que la notion d'emploi du modèle des classes d'objets est un outil qui répond très bien aux problèmes posés par le traitement automatique de la polysémie. Ainsi, le modèle des classes d'objets permet de manière très naturelle de mettre en évidence les différentes interprétations et de factoriser les analyses pour faciliter le calcul automatique.

1.2. LE TRAITEMENT DU FIGEMENT

Le traitement automatique de textes écrits procède à l'analyse à partir des constituants les plus atomiques : les caractères. L'automate regroupe ces caractères pour former des unités lexicales à partir d'un lexique et puis il analyse les relations entre les unités lexicales pour arriver aux sens. Dans ce mode de fonctionnement, chaque étape dépend de l'étape qui précède. La notion de mot est celle qui précède toutes les autres, elle tient une importance capitale pour le traitement automatique des langues (cf. Issac, 2011).

En TAL on a parfois défini le mot comme une suite de caractères figurant entre deux blancs. Pourtant, cette conception ne rend pas compte de la réalité lexicale des langues naturelles. Toute langue utilisée dans une communauté produit inévitablement des séquences figées (G. Gross, 1996a). Comme l'affirme Mejri (2007), « frozenness is a process that is inherent in all modern languages » (p. 685). En effet, il s'agit d'un phénomène d'une ampleur extraordinaire. En langue française, M. Gross (1993) a recensé près de 35 000 verbes figés face à seulement 12 000 verbes simples ; Mathieu-Colas (1996) a fait l'inventaire de 200 000 noms composés (distribués en 700 moules de formation), un chiffre bien supérieur aux environ 70 000 noms simples. D'ailleurs, on estime que 70 à 90 % de la surface d'un texte relève du figement³³⁸.

³³⁸ Ce chiffre peut varier en fonction de la conception large ou étroite de ce qui est phraséologique (cf. section 2.2. *Objet d'étude et limites de la phraséologie*, Chap. III). Nous tenons à rappeler que n'avons déjà commenté l'importance de la présence du phénomène phraséologique dans la langue (cf. section 2.3. *Phénomène marginal ou central de la langue*, Chap. III).

Les analyses systématiques des expressions phraséologiques menées aux laboratoires LADL et LDI ont mis en évidence que les structures morphosyntaxiques dans lesquelles elles s'insèrent ne diffèrent guère des structures libres. En effet, ces séquences posent des problèmes de reconnaissance automatique du fait de leur double lecture. Seule une description fine aux niveaux syntaxique et sémantique des deux lectures possibles permet d'éviter d'erreurs potentielles d'interprétation.

Alors, si le TAL a besoin d'une description exhaustive de la langue et que les séquences figées apparaissent dans le discours dans une proportion aussi importante, il s'avère nécessaire d'aborder leur traitement automatique. Notre travail de recherche ne vise pas le recensement et classement des séquences figées, mais, suite aux données fournies par notre corpus, nous avons été confrontée à la description de nombreux noms composés (p. ex., *coup de main*, *micro crédit*, *assurance médicale*).

1.2.1. La notion de figement dans le modèle des classes d'objets³³⁹

Traditionnellement on a établi une limite très nette entre les constructions libres³⁴⁰ et les constructions figées. Les premières relèvent du régulier, c'est-à-dire ce qui est formulable à partir de règles de syntaxe. Les secondes, elles relèvent du préconstruit. Or ni la combinatoire libre est totalement libre ni les séquences figées sont strictement figées. En effet, la phrase libre repose sur un ensemble de contraintes que les unités lexicales s'imposent mutuellement au niveau de la combinatoire, et en ce qui concerne les séquences figées, on trouve des constructions qui sont plus figées que d'autres. Autrement dit, il existe des degrés de figement et de degrés de liberté (Mejri, 2011b). Ainsi, dans *rater le coche*, on peut remplacer *rater* par *louper* ou *manquer* (G. Gross, 1996a, p. 16). Étant donné que cette suite présente la possibilité de paradigme, elle est moins figée que, par exemple, *chercher midi à quatorze heures* dont le figement

³³⁹ Nous tenons à rappeler que le Chap. III a été entièrement consacré à la phraséologie. Le lecteur y aura trouvé un parcours des différentes approches à l'étude du phénomène phraséologique, ainsi qu'une description des objets d'étude, des propriétés, des typologies, etc.

³⁴⁰ Rappelons que nous avons abordé cette notion dans la perspective du modèle des classes d'objets dans la section 3.2. *Caractéristiques des unités phraséologiques* du Chap. III.

est presque total³⁴¹. Il y a donc un continuum entre les séquences libres et les séquences figées : « la délimitation n'est pas simple à tracer, au point que syntaxiquement il semble exister un continuum entre formes figées et libres » (M. Gross, 1982, p. 160). Souvent, on réduit cette gradation aux variations combinatoires acceptées ou rejetés par les séquences polylexicales en fonction de leur degré de figement, mais le figement est présent à tous les niveaux de l'analyse linguistique (Mejri, 1997).

De même qu'au niveau syntaxique il existe des degrés de figement, au niveau sémantique on a pu constater différents degrés d'opacité. On comprendra que *fait historique* est bien moins opaque que *fait divers*. Dans la littérature on a fait de l'opacité sémantique un élément définitoire des unités polylexicales, comme nous avons pu l'observer dans la définition ci-dessus. Pourtant, si l'on admet qu'il y a des séquences figées opaques, transparentes ou plus ou moins opaques ou transparentes³⁴², on ne peut pas accepter la non-compositionnalité comme un critère définitoire de figement (cf. Mejri, 2003a). Si une séquence est sémantiquement opaque, c'est qu'il s'agit d'une séquence figée ; mais si la séquence est transparente, ça ne signifie pas nécessairement qu'elle soit une séquence libre. Il conviendrait mieux de considérer la non-compositionnalité comme une caractéristique facultative³⁴³.

Parfois on a corrélié l'opacité sémantique aux restrictions syntaxiques, c'est-à-dire plus une séquence est opaque, plus elle est figée du point de vue syntaxique (G. Gross, 1996a, 2012)³⁴⁴. Or certaines unités polylexicales opaques admettent des

³⁴¹ Même si cette locution pourrait sembler complètement figée (opacité sémantique, blocage des paradigmes synonymiques et des propriétés transformationnelles...), nous tenons à signaler que du point de vue syntaxique, elle n'est que partiellement figée parce que l'insertion est possible (*Luc cherche, comme c'est normal pour lui, midi à quatorze heures*).

³⁴² Dans le cas de la nomenclature de notre dictionnaire, la plupart des unités polylexicales sont transparentes (*aide médicale, asile politique*), mais d'autres sont complètement opaques (*parachute doré, coup de pouce*).

³⁴³ Pour plus de détails sur la dimension sémantique, voir Gréciano (1983, 2000c) ; Mejri (2003a).

³⁴⁴ De nombreux auteurs se sont penchés sur cette question. Il y en a qui considèrent qu'il existe une relation de dépendance réciproque l'opacité sémantique et le figement syntaxique ; par exemple, selon Gaatone (1984) « plus une expression serait transparente, moins elle serait figée et vice-versa » (p. 77). D'autres linguistes font dériver la flexibilité syntaxique de l'opacité sémantique (Bennet, 1997 ; Gibbs,

transformations. Ainsi, *briser la glace* accepte la passivation : *Une fois la glace brisée, nous avons abordé de tels sujets* (Mejri, 2003a, p. 30).

À part ces quelques caractéristiques, les paramètres d'étude du figement considérés dans le modèle des classes d'objet sont plus nombreux. Ces critères permettraient donc de déterminer si une séquence est figée ou non. Nous proposons ci-dessous une présentation succincte³⁴⁵ (G. Gross, 1996a, 2012).

- a) La polylexicalité.
- b) L'opacité sémantique.
- c) Le blocage des propriétés de restructuration et de transformation : il existe une corrélation inverse entre le nombre de propriétés de recompositions d'une séquence donnée et son degré de figement, c'est-à-dire moins elle accepte de transformations, plus elle est figée.
- d) La non-actualisation des éléments constitutifs de la séquence figée : il s'agit de l'un des critères les plus sûrs du figement.
- e) La portée du figement : le plus souvent le figement n'affecte pas à la totalité de la suite, mais seulement à un sous-ensemble.
- f) Le degré de figement.
- g) Le blocage de paradigmes synonymiques : dans les suites figées au degré le plus élevé, la possibilité de substitution synonymique est exclue.
- h) La non-insertion : dans les séquences figées, l'insertion d'éléments nouveaux (comme un adjectif, un adverbe, une relative) est très réduite.
- i) Le défigement : le figement peut être mis en évidence grâce à l'effet provoqué par le jeu du défigement, c'est-à-dire à l'ouverture des paradigmes (*silence, on assassine – silence on tourne*).

1995 ; Gibbs et Nayak, 1989 ; Ruwet, 1983). Pour d'autres, comme Abeillé (1995), il n'existerait pas de corrélation entre ces deux phénomènes.

³⁴⁵ Les caractéristiques qui ont déjà été analysées ne seront que mentionnées dans cette brève présentation.

- j) L'étymologie : se poser la question de l'origine d'une séquence figée donnée (*être riche comme Crésus*) implique que la structure n'est pas libre car on ne se pose pas la question de l'origine des séquences librement générées (*être très riche*). D'autre part, certaines séquences figées contiennent des éléments ou des constructions syntaxiques qui reflètent un état antérieur de la langue (par exemple, l'absence d'article dans *chercher noise*).
- k) La réduction des séquences figées à des parties du discours : une unité polylexicale n'est pas une catégorie primaire (elle se sert des catégories primaires), mais elle a le même statut syntaxique d'une catégorie simple.

Les unités polylexicales ont donc une double combinatoire, qui est d'ailleurs le socle sur lequel repose une description fiable des séquences figées. D'une part, la combinatoire externe d'une séquence figée est constituée par l'ensemble des relations que cette séquence entretient avec son contexte linguistique. Par exemple, pour la séquence figée *tirer les vers du nez*, la combinatoire externe serait : *hum <tirer les vers du nez> hum*. D'autre part, la combinatoire interne concerne les relations syntaxiques entre les constituants de la séquence qui assurent la pérennité du sens. Cette combinatoire interne sert à mesurer le degré de figement des séquences figées.

1.2.2. Reconnaissance et génération automatique des unités polylexicales

Comme nous l'avons déjà signalé, détecter automatiquement les séquences figées constitue l'un des grands problèmes du traitement automatique du fait de leur double lecture. Le cas le plus simple pour le TAL serait un mot à trait d'union (*eau-de-vie*), dont le traitement lexicographique est comme celui d'un mot simple. Il n'est pas non plus très difficile pour l'automate de reconnaître une séquence entièrement figée, dont la signification est totalement opaque et qui ne se prête pas à aucune lecture compositionnelle (p. ex., *à huis clos*). À partir de ces cas de figure assez simples, il existe un éventail de degrés de figement et d'opacité.

Pour le traitement automatique, le premier problème qui se pose est le découpage du texte en unités. La solution la plus naturelle est d'élaborer des dictionnaires électroniques dont les entrées sont les unités polylexicales, et qui portent les mêmes indications que les mots simples (G. Gross, 2004b ; Blanco et Buvet, 2000).

Ainsi, au moment du traitement automatique, les séquences figées doivent être repérées les premières, pour éviter la traduction compositionnelle. Dans cette première tâche, il faut que l'automate prenne compte de la suite la plus longue, car il existe des cas d'« enchâssement » de séquences figées (p. ex., *corps de garde / plaisanterie de corps de garde ; foie de morue / huile de foie de morue ; chemin de fer / chemin de fer de ceinture*) (G. Gross, 2004b). Ainsi, l'automate doit être en mesure d'éviter de disloquer ces noms composés longs qui contiennent d'autres noms composés. Cela n'est possible que sur la base de dictionnaires électroniques contenant des indications sur la portée du figement.

En conclusion, la prise en compte des séquences figées en TAL nécessite de recensements exhaustifs et systématiques de ces séquences, ainsi que de l'ensemble de leurs propriétés, dans des dictionnaires électroniques³⁴⁶. Dans le cadre du modèle des classes d'objets, le classement des prédicats polylexicaux se fait essentiellement à partir du modèle adopté pour la description des emplois prédicatifs simples, en ajoutant juste quelques particularités.

2. Méthodes d'analyse automatique : méthode statistique vs méthode mixte

Le développement de l'informatique a supposé une rouée vers des méthodes de TAL strictement statistiques. Pourtant, la tendance actuelle est de prendre un positionnement plutôt intermédiaire entre les méthodes qualitatives traditionnelles et les nouvelles méthodes quantitatives. Dans cette section, nous présenterons brièvement ces approches et nous expliquerons pourquoi nous considérons que l'approche mixte est la plus avantageuse.

2.1. La méthode statistique

Depuis le début des années 90 et grâce à l'importance accrue accordée aux méthodes numériques, les méthodes statistiques ont été massivement utilisées dans le processus d'analyse (cf. Abney, 1991 ; Leacock et al. 1996). Leur grand atout est

³⁴⁶ Le système INTEX (Silberztein, 1993a) intègre l'ensemble des dictionnaires produits au LDI.

qu'elles permettent d'obtenir des résultats sans faire appel à de complexes ressources linguistiques.

La démarche statistique consiste à identifier et encoder les unités linguistiques en fonction de la fréquence de cooccurrence dans différents contextes. Ainsi, on considère que les informations nécessaires à la reconnaissance du sens d'une unité linguistique ou des relations entre elles sont définies par leur environnement strict, dégagées par de logiciels d'extraction qui opèrent sur de grands corpus (Massoussi et Mejri, 2011).

Cependant, les éléments contextuels qui définissent une unité lexicale ne lui sont pas forcément contigus (on pourrait, par exemple, trouver des insertions de circonstants ou des effacements d'arguments). Par ailleurs, le premier problème auquel se heurte le traitement statistique est la reconnaissance des unités lexicales. Les séquences figées constituent de faux environnements que cette approche aurait du mal à reconnaître. En effet, les méthodes statistiques se sont avérées incapables de résoudre les problèmes de génération de bruit ou de silence.

Néanmoins, l'outil statistique est pratique en termes de temps de traitement et d'occupation de mémoire. Il a donné lieu à des outils pratiques intéressants dont certains chercheurs ont tiré des apprentissages pour les appliquer aux fondements et les méthodes de la linguistique (Cori, 2008). L'utilité des méthodes statistiques dépendrait donc des applications envisagées.

2.2. La méthode mixte

Depuis une vingtaine d'années, ces méthodes statistiques semblent avoir atteint leurs limites. Comme l'affirment Kauffer et Keromnes (2022), « une approche quantitative de la phraséologie n'est pas la panacée » (p. 13). Actuellement, les méthodes mixtes (cf. De Loupy et al., 2003) s'avèrent plus performantes. Tout en utilisant des outils statistiques, elles exploitent aussi des ressources linguistiques. Ainsi, Granger et Meunier (2008), dans leur introduction, soutiennent que l'approche statistique (ou approche quantitative) n'est pas appelée à prendre la place de l'approche

qualitative³⁴⁷ traditionnellement adoptée, mais ces deux approches devraient plutôt s'enrichir mutuellement. Nous sommes d'accord avec Buvet (2013a) lorsqu'il affirme : « la complémentarité entre les deux approches, dite approche mixte, permet d'allier la finesse de l'approche linguistique à la puissance des calculs de l'approche mathématique » (p. 197).

Il existe une grande variété de méthodes mixtes. Cela est dû à trois facteurs : le type de méthodologie linguistique adopté, la description des faits de langue sous-jacente à la méthode linguistique utilisée, et, finalement, l'articulation entre la méthode linguistique et la méthode statistique (Buvet, 2009b).

Le premier facteur a trait à la nature des ressources linguistiques intégrées dans les systèmes. Par exemple, certains étiqueteurs ne se servent que de données métalinguistiques. Ces méthodes, destinées à des heuristiques d'apprentissage fondées sur des calculs statistiques, n'utilisent que des données métalinguistiques (cf. Russel et Norvig, 1995). Brill (1994) les définit comme des méthodes linguistiques minimalistes. À l'autre extrême, il existe des méthodes linguistiques bien plus maximalistes, qui utilisent une grande variété de ressources exhaustives (cf. Silberstein, 1993a).

Le deuxième facteur concerne la variété descriptive des méthodes linguistiques. D'abord, les descriptions des faits de langue peuvent être de différente nature : description morphologique, syntaxique, sémantique, morphosyntaxique, morphosémantique et sémanticosyntaxique. Ensuite, ces descriptions peuvent être compilées sous des outils différents : des dictionnaires électroniques, des listes de règles et des banques de graphes. Enfin, les annotations métalinguistiques poursuivent des objectifs différents. L'étiquetage³⁴⁸ des mots garantit l'uniformisation et la catégorisation de ces mots. Les annotations relatives aux phrases servent à les analyser et à les structurer. Celles qui concernent les sens participent à l'interprétation des textes.

Le troisième facteur porte sur les rôles de la méthode linguistique et de la méthode statistique dans le processus de traitement automatique. Ainsi, il est nécessaire

³⁴⁷ Nous avons déjà eu l'occasion d'expliquer la différence entre les approches qualitatives et les approches quantitatives (cf. section 1. *Bref aperçu historique de la phraséologie* du Chap. III).

³⁴⁸ Nous utilisons de manière indistincte les termes *étiquetage*, *annotation* et *notation*.

de définir les tâches attribuées à chaque méthode, à quel moment du processus de traitement ces méthodes interviennent et dans quelle mesure.

3. Ressources lexicales pour l'étiquetage

Buvet et Blanco (2000) résument le fonctionnement du TAL en deux processus : l'analyse et la génération. D'une part, le processus d'analyse, également appelé de reconnaissance, consiste à produire la représentation d'un énoncé ou d'un texte sous la forme d'un ensemble de codes métalinguistiques qui sont reliés entre eux d'une façon structurée. D'autre part, le processus de génération consiste à produire des énoncés à partir d'informations encodées appartenant aux niveaux sémantique, syntaxique et morphologique. Les systèmes de traduction automatique ou de résumés automatiques utilisent autant le processus de génération que celui d'analyse.

L'étiquetage est une tâche intermédiaire du processus du traitement automatique des textes écrits. Cette tâche consiste à attribuer une valeur morphosyntaxique, syntaxique ou sémantique aux unités lexicales ou aux groupes d'unités lexicales. Ces trois types de notation se font de manière successive. D'abord, l'analyse lexicale identifie les unités lexicales et les morphèmes grâce à des outils d'étiquetage morphosyntaxique. Ensuite, l'analyse syntaxique utilise les ressources d'étiquetage syntaxique pour repérer la combinatoire et les relations entre les unités lexicales (les emplois prédictifs et les schémas d'arguments) qui configurent les phrases et les textes. Enfin, l'analyse sémantique, avec ses propres outils d'étiquetage, identifie les significations transmises comme résultat de la relation entre le lexique et la syntaxe. L'étiquetage automatique fonctionne donc par couches disjointes, mais elles sont interdépendantes. L'étiquetage automatique des niveaux morphologique et morphosyntaxique est relativement simple. Par contre, l'étiquetage sémantique est bien plus complexe.

Comme le signale Choi (2022), « les ressources lexicales informatisées ou *machine-readable* constituent des données indispensables pour développer des outils et des méthodes réalisant des tâches de Traitement Automatique des Langues (TAL) » (p. 55). Leur influence sur la performance des systèmes et des programmes de TAL est décisive. La notion de ressource lexicale est très large (cf. Gala, 2013). Elle désigne des

objets très hétérogènes tels que des corpus, des dictionnaires électroniques, des lexiques, des réseaux lexicosémantiques, etc. Une ressource lexicale décrit la langue, mais les buts poursuivis, le pan de linguistique visé, la perspective théorique adoptée, leur construction, leur niveau de description linguistique, leur forme, etc. varient énormément, donnant lieu à des ressources de nature très différente. De nos jours, l'élaboration de ces ressources se fait souvent de manière automatique ou semi-automatique (avec validation manuelle par des linguistes)³⁴⁹ et souvent on utilise des ressources pré-existantes pour créer de nouvelles ressources plus complètes ou dans d'autres formalismes³⁵⁰.

Dans ce qui suit, nous allons nous intéresser à différentes ressources d'étiquetage lexical dont se servent les programmes de traitement automatique³⁵¹, et qui sont également lisibles et mis à disposition pour l'être humain. Il s'agit d'outils de différente nature ayant différentes finalités. Notre propos est d'offrir un panorama de certaines ressources électroniques de libre utilisation qui ont été élaborées à des fins de recherche. Nous allons présenter quelques ressources pour l'étiquetage morphosyntaxique, pour l'étiquetage sémantique, pour l'étiquetage en rôles sémantiques et des ressources pour l'étiquetage de la temporalité et de l'aspect. Enfin, nous consacrons une section à l'informatisation des dictionnaires papier.

³⁴⁹ Il existe aussi des ressources construites de manière collaborative faisant participer des personnes non expertes. Cette pratique par laquelle on sollicite la collaboration d'un grand nombre de personnes reste plutôt restreinte à de petites tâches dont l'automatisation est difficile (Lafourcade et Joubert, 2013). Cette méthode permet d'obtenir un grand nombre de données de manière très rapide et peu coûteuse, mais en revanche elle soulève la question de la qualité des données. À titre d'exemple, nous pouvons citer Wiktionary (<https://www.wiktionary.org/>) ou JeuxDeMots (<https://www.jeuxdemots.org/jdm-accueil.php>) (Lafourcade et Joubert, 2008 ; Lafourcade et Le Brun, 2020), une ressource lexicosémantique sous forme de réseau lexical créée à partir de différents jeux en ligne.

³⁵⁰ Sur les méthodes de construction des ressources lexicales informatiques du français et leur alignement, voir Choi (2022) et Polguère (2014).

³⁵¹ Nous avons exclu les plongements lexicaux (p. ex. : *FastText*, cf. Bojanowski et al., 2017, Grave et al., 2018 ; *Glove*, cf. Pennington et al., 2014 : *Word2vec*, cf. Fauconnier, 2015, 2022, Mikolov et al. 2013). Bien qu'ils connaissent un important essor depuis une quinzaine d'années, ils sont relativement peu utilisés en tant que ressources. Ils sont plutôt utilisés en tant qu'outils, c'est-à-dire des « modèles qui sont utilisés pour générer des plongements lexicaux propres à une application » (Choi et al., 2023, p. 28).

3.1. Ressources pour l'étiquetage morphosyntaxique³⁵²

3.1.1. DicoValence

DicoValence (cf. Mertens, 2010 ; van den Eynde et Mertens, 2003 ; van den Eynde et al., 2017) est un dictionnaire des verbes du français développé dans le cadre de l'*approche pronominale*³⁵³. Il répertorie les cadres valenciels³⁵⁴ de plus de 3700 verbes simples du français en 8000 entrées. En ce qui concerne la nomenclature, un verbe a autant d'entrées qu'il a de cadres valenciels. Ceux-ci sont définis selon le nombre, la nature, le caractère facultatif et la composition des paradigmes pronominaux.

VAL\$	supprimer: P0 P1
VTYPE\$	predicator simple
VERB\$	SUPPRIMER/supprimer
NUM\$	80500
EG\$	nous avons supprimé tous les obstacles à la publication de ce dico
TR_DU\$	afschaffen, opheffen, intrekken, weghalen, weglaten, schrappen, doen verdwijnen
TR_EN\$	suppress
FRAME\$	subj:pron n:[hum], obj:pron n:[hum,nhum,?abs]
P0\$	que, qui, je, nous, elle, il, ils, on, ça, celui-ci, ceux-ci
P1\$	que, qui, te, vous, la, le, les, se réfl., se réc., en Q, ça, ceci, celui-ci, ceux-ci, l'un l'autre
RP\$	passif être, se passif, se faire passif
AUX\$	avoir

Tableau 4 : Entrée supprimer du DicoValence (extrait de Van den Eynde et Mertens, 2006, p. 4)

Comme nous pouvons l'observer sur la Tableau 4, le premier ensemble de champs permet d'identifier l'entrée avec sa désignation conventionnelle (*VAL*), le type de verbe (prédicateur simple, verbe adjoint, verbe auxiliaire, etc.), l'infinitif, le numéro identifiant de l'entrée, un exemple et des traductions en néerlandais et en anglais. Ensuite, se trouve l'ensemble où l'on précise le schéma valenciels du verbe : le nombre de valences, leur nature, leur caractère facultatif et la composition des paradigmes pronominaux (dans le cas de notre exemple, les paradigmes pronominaux de *P0* et *P1*), ainsi que les autres propriétés de construction associées à cette configuration (*RP* étant la reformulation passive). Enfin, cette entrée contient un champ additionnel (*AUX*) qui informe sur le verbe auxiliaire utilisé dans les temps composés.

³⁵² Nous comprenons par morphosyntaxe « les phénomènes de variation formelle des morphèmes déterminés par la structuration syntaxique de l'énoncé » (Neveu, 2011, p. 239).

³⁵³ Pour une présentation de l'approche pronominale, voir van den Eynde et Blanche-Benveniste (1978).

³⁵⁴ Selon van den Eynde et Mertens (2006), par *cadre valenciels* on entend : « le nombre et la nature des compléments valenciels du verbe, y compris le sujet, avec mention de leur fonction syntaxique » (p. 2).

3.1.2. Lefff (Lexique des Formes Fléchies du Français) – Leffe (Léxico de Formas Flexionadas del Español)

Le *Lefff* (cf. Clément et al., 2004 ; Sagot, 2010 ; Sagot et Danlos, 2008) et le *Leffe*³⁵⁵ (Molinero et al., 2009a, 2009b) sont des lexiques morphologiques et syntaxiques à large couverture. La particularité du modèle lexical de ces ressources tient au fait qu'il se veut indépendant d'un modèle théorique de description de la langue. Ainsi, les informations retenues sont le fruit d'un consensus de spécialistes de différentes équipes de recherche. Il s'agit du lexique du français le plus utilisé en TAL pendant ces 20 dernières années (Choi et al., 2023)

Le *Lefff* contient 100 000 lemmes et 520 000 entrées, acquises de manière automatique à partir des tables du Lexique-Grammaire (M. Gross, 1975), de la ressource lexicale Les Verbes Français (LVF) (Dubois et Dubois-Charlier, 1997)³⁵⁶, de Dicovalence et de l'acquisition automatique avec validation manuelle de données sur corpus.

Dans ce lexique flexionnel, à chaque entrée sont associées autant des informations morphologiques (le lemme et la partie du discours) que des informations syntaxiques très riches :

Ces informations syntaxiques donnent le cadre de sous-catégorisation (la valence, la rection) d'un emploi de verbe plein sous forme de liste de fonctions syntaxiques (Suj, Obj, Objà, etc.) assorties de leurs diverses réalisations de surface (réalisations sous forme de groupe nominal, de groupe adjectival, de complétive, d'interrogative indirecte, d'infinitive ou de clitique, chacune de ces réalisations pouvant être précédées d'une préposition). (Danlos et Sagot, 2008, p. 131)

³⁵⁵ Le *Leffe* a été créé à partir du *Lefff*. En effet, dans le *Lefff* on a développé le formalisme lexical de modélisation et d'acquisition lexicale appelé Alexina (*Architecture pour les LEXiques INformatiques et leur Acquisition*) (Sagot, 2010).

³⁵⁶ L'ouvrage de Dubois et Dubois-Charlier (1997) est la version papier (légèrement réduite) de la base de données électronique des verbes du français. Il s'agit d'un thésaurus de classes syntacticosémantiques, c'est-à-dire de classes sémantiques définies par la syntaxe, constitué en appliquant les méthodes de la grammaire distributionnelle et transformationnelle. Cette ressource lexicale est disponible sur le site : <http://rali.iro.umontreal.ca/rali/?q=fr/versions-informatisees-lvf-dem>. Pour une présentation de cette ressource et de ses principes de constitution, voir François et al. (2007).

<i>Luc apprend la conduite / à conduire à Marie</i>
apprend v [pred='apprendre_teach_1<Suj:sn cln, Obj:(sn cla à-sinf scompl qcompl), Objà:(à-sn cld)>', cat=v, @personnel, @CtrlObjàObj, @Complnd, @P3s]
<i>Marie apprend la conduite / à conduire</i>
apprend v [pred='apprendre_learn_2<Suj:sn cln, Obj:(sn cla à-sinf scompl qcompl)>', cat=v, @personnel, @CtrlObjàObj, @Complnd, @P3s]
<i>La conduite / conduire s'apprend facilement</i>
apprend v [pred='apprendre_learn_2<Suj:sn cln sinf>', cat=v, @personnel, @pronominal, @P3s]

Figure 7 : *Trois entrées du verbe apprendre dans le Lefff* (extrait de Sagot et Danlos, 2008, p. 131)

Entre crochets se trouve la liste des fonctions syntaxiques de chaque emploi, assorties de leurs réalisations et un ensemble de couples attributs-valeurs (@) pour décrire d'autres informations morphosyntaxiques (catégorie syntaxique, pronominalité, impersonnalité, contrôles, attributifs, traits morphologiques, etc.).

Le *Lefff* et le *Leffe* sont en un certain sens des lexiques sémantiques car l'inventaire d'entrées lexicales est déterminé par des propriétés morphologiques, syntaxiques et sémantiques. D'ailleurs, Sagot (2018) signale que l'une des perspectives d'évolution de ces ressources serait d'« associer chacune de ses entrées à des identifiants sémantiques tels que disponibles dans le lexique sémantique WOLF » (p. 133)³⁵⁷.

3.2. Ressources pour l'étiquetage sémantique

Les ressources présentées dans cette section fournissent des informations sémantiques au moyen de la description des emplois de mots dans un espace de relations sémantiques stables qui permettent la constitution de réseaux sémantiques. Bien que les sources d'informations sémantiques soient nombreuses, notamment les versions informatisées des dictionnaires imprimés, seulement certaines de ces ressources sont exploitables pour l'étiquetage de corpus et l'analyse linguistique de la langue.

3.2.1. WordNet

³⁵⁷ Pour une analyse critique et perspectives d'évolution du *Lefff*, voir Sagot (2018). Par contre, le *Leffe* n'a pas avancé depuis plusieurs années.

WordNet³⁵⁸ (cf. Fellbaum, 1998, 2006 ; Miller, 1995) est une base lexicale initialement développée pour l'anglais, par des psychologues cognitifs du Cognitive Science Laboratory à Princeton University. Il s'agit de la première grande tentative de construction d'un dictionnaire sémanticoconceptuel du langage courant (cf. Cartier, 2009). WordNet est fondé sur une approche essentiellement relationnelle qui organise le lexique en hiérarchies de *synsets* (ensemble de quasi-synonymes représentant un concept lexical), reliés entre eux par différentes relations sémantiques (hypéronymie, hyponymie, antonymie, méronymie...) et lexicales³⁵⁹ (dérivation). Cette hiérarchisation est basée sur une approche psycholinguistique du lexique et sur des théories sur la mémoire sémanticoconceptuelle. La modélisation du lexique est donc mentale : accéder aux mots par les concepts. Le postulat qui régit la structuration des contenus de WordNet est le suivant : plus les lexicographes s'inspirent de la façon dont on accède mentalement à l'information lexicale, plus il sera facile aux utilisateurs de naviguer dans le dictionnaire.

Ainsi, douze grandes classes nominales, trois grandes classes de verbes et deux grandes classes d'adjectifs ont été établies. Ensuite, ces grandes classes ont été subdivisées en sous-classes ou *synsets*. Les noms sont classifiés au moyen de relations d'hypéronymie, hyponymie, méronymie, holonymie, antonymie et dérivation. Les verbes sont organisés par des relations d'hypéronymie, hyponymie, troponymie, antonymie, implication, cause et des dérivés morphologiques. Parmi les adjectifs, les descriptifs sont hiérarchisés en fonction des relations de similarité, antonymie ou attribut ; les adjectifs relationnels sont liés au nom dont ils sont dérivés. Finalement, les adverbes, qui sont souvent dérivés morphologiquement d'un adjectif, ils sont mis en relation avec celui-ci. Également, les adverbes héritent généralement les relations sémantiques des adjectifs dont ils dérivent (par exemple, l'antonymie est très fréquente).

À titre d'exemple, nous montrons ici le résultat de la recherche du nom *help* :

³⁵⁸ Disponible sur ce lien : <http://wordnetweb.princeton.edu/perl/webwn> .

³⁵⁹ Dans ses débuts, cette base lexicale était conçue pour la description des relations sémantiques. Les relations lexicales n'ont été ajoutées qu'ultérieurement, mais elles restent secondaires face au vaste réseau sémantique mis en place.

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- **S: (n) aid, assist, assistance, help** (the activity of contributing to the fulfillment of a need or furtherance of an effort or purpose) *"he gave me an assist with the housework"; "could not walk without assistance"; "rescue party went to their aid"; "offered his help in unloading"*
 - [direct hyponym / full hyponym](#)
 - [direct hypernym / inherited hypernym / sister term](#)
 - [derivationally related form](#)
- **S: (n) assistant, helper, help, supporter** (a person who contributes to the fulfillment of a need or furtherance of an effort or purpose) *"my invaluable assistant"; "they hired additional help to finish the work"*
 - [direct hyponym / full hyponym](#)
 - [direct hypernym / inherited hypernym / sister term](#)
 - [derivationally related form](#)
- **S: (n) aid, assistance, help** (a person or thing that is a resource that helps make something easier or possible to do) *"visual aids in teaching"; "his job was to give technical assistance over the phone"*
 - [direct hypernym / inherited hypernym / sister term](#)
 - [derivationally related form](#)
- **S: (n) avail, help, service** (a means of serving) *"of no avail"; "there's no help for it"*
 - [direct hypernym / inherited hypernym / sister term](#)
 - [derivationally related form](#)

Figure 8 : Entrées du nom *help* dans WordNet (extrait de

<http://wordnetweb.princeton.edu/perl/webwn>)

Chaque entrée du nom *help* contient des synonymes, une définition (ou glose), des exemples et d'autres relations sémantiques et morphologiques. Les hyponymes et hypéronymes sont hiérarchisés en fonction de leur relation directe ou indirecte avec le mot donné. Ainsi, pour la première entrée du nom *help*, voici les hyponymes qui lui sont associés :

Word to search for: Search WordNet

Display Options: Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
 Display options for sense: <lexical filename >

Noun

- <noun.act>S: (n) [aid](#), [assist](#), [assistance](#), [help](#)
 - [direct hyponym](#) / [full hyponym](#)
 - <noun.act>S: (n) [self-help](#)
 - <noun.act>S: (n) [facilitation](#)
 - <noun.act>S: (n) [hand](#), [helping hand](#)
 - <noun.act>S: (n) [recourse](#), [resort](#), [refuge](#)
 - <noun.act>S: (n) [thanks](#)
 - <noun.act>S: (n) [relief](#), [succor](#), [succour](#), [ministration](#)
 - <noun.act>S: (n) [consolation](#), [comfort](#), [solace](#)
 - <noun.act>S: (n) [mercy](#)
 - <noun.act>S: (n) [lift](#)
 - <noun.act>S: (n) [service](#)
 - <noun.act>S: (n) [childcare](#), [child care](#)
 - <noun.act>S: (n) [daycare](#), [day care](#)
 - <noun.act>S: (n) [community service](#)
 - <noun.act>S: (n) [community service](#), [public service](#)
 - <noun.act>S: (n) [help desk](#), [helpdesk](#)
 - <noun.act>S: (n) [seating](#)
 - <noun.act>S: (n) [accommodation](#)
 - <noun.act>S: (n) [service](#)
 - <noun.act>S: (n) [curb service](#)
 - <noun.act>S: (n) [self-service](#)
 - <noun.act>S: (n) [valet parking](#)
 - <noun.act>S: (n) [boost](#), [encouragement](#)
 - <noun.act>S: (n) [morale building](#), [morale booster](#)
 - <noun.act>S: (n) [comfort](#)
 - <noun.act>S: (n) [support](#)
 - <noun.act>S: (n) [attachment](#), [adherence](#), [adhesion](#)
 - <noun.act>S: (n) [ecclesiasticism](#)
 - <noun.act>S: (n) [kabbalism](#), [cabalism](#)
 - <noun.act>S: (n) [royalism](#)
 - <noun.act>S: (n) [traditionalism](#)
 - <noun.act>S: (n) [advocacy](#), [protagonism](#)
 - <noun.act>S: (n) [drumbeat](#)
 - <noun.act>S: (n) [insistence](#), [insistency](#)
 - <noun.act>S: (n) [urging](#)
 - <noun.act>S: (n) [sponsorship](#)
 - <noun.act>S: (n) [endorsement](#), [indorsement](#)
 - <noun.act>S: (n) [auspices](#), [protection](#), [aegis](#)
 - <noun.act>S: (n) [blessing](#), [approval](#), [approving](#)
 - <noun.act>S: (n) [adoption](#), [acceptance](#), [acceptation](#), [espousal](#)
 - <noun.act>S: (n) [embrace](#), [bosom](#)
 - <noun.act>S: (n) [backing](#), [backup](#), [championship](#), [patronage](#)
 - <noun.act>S: (n) [admiration](#), [appreciation](#)
 - <noun.act>S: (n) [adoration](#), [idolization](#), [idolisation](#)
 - <noun.act>S: (n) [glorification](#)
 - <noun.act>S: (n) [idealization](#), [idealisation](#), [glorification](#)
 - <noun.act>S: (n) [sentimentalization](#), [sentimentalisation](#), [romanticization](#), [romanticisation](#)
 - <noun.act>S: (n) [reward](#), [reinforcement](#)
 - <noun.act>S: (n) [carrot](#)
 - <noun.act>S: (n) [reassurance](#)

Figure 9 : Hyponymes de la première entrée du nom *help* dans WordNet360 (extrait de <http://wordnetweb.princeton.edu/perl/webwn>)

En somme, il s'agit d'une organisation sémantique « de sens commun » très réussie du fait de sa profonde réflexion, de son exhaustivité et de l'ampleur au niveau de la couverture des unités lexicales. Cependant, WordNet manque d'informations sur les contextes dans lesquels les mots apparaissent pour acquérir un sens donné. Dans la hiérarchisation sémantique de WordNet, un mot appartient à autant de synsets qu'il a de sens, mais aucune information sur les structures argumentales et les restrictions

³⁶⁰ Par souci de clarté, nous avons caché les informations relatives à la définition (ou glose) et aux exemples.

sélectionnelles de chaque emploi n'est donnée. En conséquence, l'exploitation informatique de cette ressource va générer de l'ambiguïté, parce que chaque mot sera associé à tous les synsets décrits dans la ressource, mais qui appartiennent à des emplois différents.

Actuellement il existe des ressources de type WordNet en plusieurs langues, la version anglaise étant la plus complète. Ces différentes ressources monolingues ont été rassemblées dans des réseaux multilingues : EuroWordNet/Mutiwordnet³⁶¹ (anglais, français, espagnol, allemand, néerlandais, italien, estonien et tchèque) ; BalkaNet³⁶² (serbe, roumain, bulgare, tchèque, grec, turc) ; Arabic WordNet³⁶³ ; Russian WordNet³⁶⁴, entre autres³⁶⁵.

Le WOLF³⁶⁶ (WOrdNet Libre du Français)³⁶⁷ est la version française, issue du projet EuroWordNet. Ce lexique sémantique, acquis automatiquement, a été développée à partir du Princeton WordNet et de diverses ressources multilingues et de lexiques bilingues extraits de Wikipedia³⁶⁸ et de Wiktionary³⁶⁹. Les entrées monosémiques ont été traduites du Princeton WordNet, tandis que le traitement des vocables polysémiques repose sur l'alignement d'un corpus parallèle en cinq langues (l'anglais, le bulgare, le français, le roumain et le tchèque). Ainsi, cette ressource contient plus de 32 000 synsets

³⁶¹ Voir Vossen (1998).

³⁶² Voir Tufis et al. (2004).

³⁶³ Voir Black et al. (2006) ; Rodríguez et al. (2008).

³⁶⁴ Voir Azarova et al. (2002). Disponible sur ce site web : <https://wordnet.ru/>.

³⁶⁵ L'organisation Global WordNet Association (GWA) a été créée dans le but de connecter les bases de données WordNet des différentes langues et ainsi promouvoir leur standardisation. Sur le site web de GWA (<http://globalwordnet.org/>) se trouve la liste actualisée de WordNet développés dans le monde.

³⁶⁶ Disponible sur ce lien : <http://alpage.inria.fr/~sagot/index-fr.html>.

³⁶⁷ Disponible sur ce lien : https://almanach.inria.fr/software_and_resources/WOLF-en.html. Pour la version initiale du WOLF (WOLF 0.1.4), voir Fišer et Sagot (2008) ; Sagot et Fišer (2008). Pour son extension à large couverture (WOLF 0.2), voir Sagot et Fišer (2011, 2014).

³⁶⁸ https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal

³⁶⁹ <https://www.wiktionary.org/>

et 38 000 lexèmes³⁷⁰. En espagnol, l'Universitat Autònoma de Barcelona a développé le WordNet Español³⁷¹ (cf. Fernández-Montraveta et Vázquez, 2010 ; Fernández-Montraveta, Vázquez et Fellbaum, 2008), contenant 30 000 entrées lexicales nominales y verbales.

3.2.2. DiCo (Dictionnaire de Combinatoire)

Cette base de données lexicale du français a été développée à l'Observatoire de Linguistique Sens-Texte (OLST) de l'Université de Montréal. Sous la direction d'Igor Mel'čuk et Alain Polguère, le DiCo (cf. Polguère, 2000a, 2000b, 2003 ; Steinlin et al., 2005) a été élaborée dans le cadre de la théorie Sens-Texte (cf. Mel'čuk, 1997 ; Mel'čuk, Class et Polguère, 1995 ; Mel'čuk et Polguère, 1987 ; Polguère, 1998a ; Wanner, 1996), dont la composante lexicale est la lexicologie explicative et combinatoire (cf. Mel'čuk, Clas et Polguère, 1995). Le Dico est une version plus formelle et moins riche du *Dictionnaire explicatif et combinatoire* (DEC) (Mel'čuk et al., 1984, 1988, 1992, 1999), conçu aussi selon les principes de la lexicologie explicative et combinatoire. Le développement du DiCo est essentiellement manuel. Selon Polguère (2003),

la finalité première de cette base est de décrire chaque lexie apparaissant dans la nomenclature du DiCo selon deux axes : les *dérivations sémantiques* (relations sémantiques fortes)³⁷² qui la lient à d'autres lexies de la langue et les *collocations* (expressions semi-idiomatiques) qu'elle contrôle. Cette description s'accompagne d'une modélisation des structures syntaxiques régies par la lexie et d'une modélisation de son sens, sous forme d'étiquetage sémantique. (p. 40)

S'agissant d'un travail de longue haleine, le DiCo ne contient que les lexies les plus courantes du français : 1125 lexies (*mots-clés*) et 26 093 liens lexicaux. Il possède une structure plutôt traditionnelle, c'est-à-dire une structuration en articles lexicographiques inspirée des dictionnaires papier d'usage courant tels que le Petit

³⁷⁰ Le French WordNet est un autre WordNet du français issu du projet EuroWordNet, mais il n'est pas libre d'utilisation et il contient moins de *synsets* que le WOLF (22 121 *synsets*). Il a donc été peu exploité.

Nous ne le présentons pas car il a été très peu exploité en raison de sa licence restrictive et

³⁷¹ Disponible sur ce lien : <http://grial.uab.es/synset/synset2.php>.

³⁷² Pour une définition de la notion de dérivation sémantique, voir Polguère (2000a, p. 517-518).

Robert ou Larousse. Chaque entrée du dictionnaire décrit un *vocable*, qui peut contenir une ou plusieurs *lexies* (unités lexicales) correspondant aux différentes acceptions. Chaque lexie (*mot-clé*) est décrite dans une fiche. Les vocables sont stockés comme des regroupements de fiches. Alors, comme toute base de données le DiCo est un ensemble de fiches. Chaque fiche contient onze champs principaux. Nous montrons ci-dessous une description partielle de la lexie *meurtre*³⁷³ (Polguère, 2000a) :

Champ	Valeur
vocable	MEURTRE
caractéristiques grammaticales	nom, masc
étiquette sémantique	action de tuer
forme propositionnelle	~ PAR L'individu X DE L'individu Y
tableau de régime	X = I = de N, A-poss Y = II = de N, A-poss
(quasi-)synonymes	{QSyn} assassinat, homicide#1; crime
fonctions lexicales	{V0} tuer {A0} meurtrier-adj /*Nom pour X*/ {S1} auteur [de ART ~] //meurtrier-n /*Nom pour Y*/ {S2} victime [de ART ~] /*Très choquant*/ {Magn} atroce, affreux, brutal, horrible, inqualifiable, odieux /*Qui a été préparé*/ {A2Prepar1} avec préméditation, prémédité postpos //assassinat {Tel qu'il y a deux/trois/quatre Y} double/triple/quadruple antépos ["Les victimes de ce double meurtre sont un père de famille et son fils de 15 ans."] /*Faire un M.*/ {Oper1} accomplir, commettre, perpétrer [ART □]; tremper [dans ART ~] ["Il a refusé de tremper dans ce meurtre odieux."] /*Causer que X fasse un M.*/ {CausOper1} pousser [N=X au ~] /*Raison d'un M.*/ {S1CausOper1} mobile [de ART ~] /*S'occuper d'un M.*/ {Real-I} enquêter [sur ART ~] {Real-II} élucider [ART ~], trouver l'auteur de ART ~ {Real-III} punir, châtier [ART ~]; venger [ART ~] /*Service de police qui s'occupe des M.*/ {S1Real-I/II} brigade criminelle /*Préparer un M.*/ {Prepar1} préméditer, préparer [ART ~] //comploter

³⁷³ Selon les conventions d'écriture des dictionnaires explicatifs et combinatoires, nous présentons le contenu des champs de la lexie *meurtre* en police d'écriture Courier New. Dans un souci de praticité, nous ne présentons pas les champs spécifiques à la gestion automatique de la base de données ni celui qui correspond aux exemples. Pour une présentation plus complète de la structure et les formalismes du DiCo, voir Polguère (2000a, 2003).

	{Essayer de faire croire qu'un M. est un N} maquiller [ART ~ en N] ["Il a maquillé ce meurtre en accident/suicide."] {Fait de tenter un M.} tentative [de ~] {Cri lancé par qqn. qui assiste à un M.} « Au ~ ! »
phraséologie	_appel au meurtre_ _crier au meurtre_

Tableau 5 : Description partielle de l'unité lexicale *meurtre* dans le DiCo (adaptée de Polguère, 2000a)

La caractérisation sémantique de l'unité lexicale est modélisée sous forme d'étiquette sémantique³⁷⁴. Celle-ci inscrit l'unité lexicale dans une hiérarchie. L'étiquette sémantique et la structure actantielle correspondraient à la définition lexicographique. Ainsi, cette définition précise que le sens principal de *meurtre* est 'action de tuer' et que ce sens implique deux actants sémantiques³⁷⁵ : celui qui tue (X) et celui qui est tué (Y). Nous tenons à signaler que dans la forme propositionnelle que nous avons copiée, les éléments en majuscules n'appartiennent pas à la formalisation du DiCo, mais ils ont été ajoutés pour être compris par l'humain. En TAL, cette formule aurait la forme suivante :

(1) `action_de_tuer (individu :X, individu :Y)`

Le tableau de régime indique « par quelles structures syntaxiques régies s'expriment les actants sémantiques » (Steinlin et al. 2004, p. 178). Dans notre cas de figure, les deux actants peuvent apparaître dans le discours en tant que compléments du nom (*le meurtre de X ; le meurtre de Y*) ou en tant qu'adjectifs possessifs (*son meurtre*).

Les relations lexicales (dérivations sémantiques et collocations) sont modélisées sous forme de fonctions lexicales, dont le nom est noté entre accolades avant la liste de valeurs (p. ex. {QSyn} ou {Magn}). Les synonymes, quasi-synonymes et termes génériques sont décrits dans un champ séparé des autres relations sémantiques afin de permettre d'obtenir l'ensemble de synonymes en activant une procédure de recherche simple. En ce qui concerne les autres relations lexicales, les relations paradigmatiques

³⁷⁴ Polguère (2011a) définit l'étiquette sémantique d'une lexie comme « la forme normalisée du genre prochain de la définition de L qui établit l'appartenance de L à une classe sémantique de lexies d'au moins deux éléments » (p. 202). Pour approfondir sur l'étiquetage sémantique dans le cadre de la théorie Sens-Texte, voir, Milićević (1997) et Polguère (2003).

³⁷⁵ Pour la notion d'actant sémantique dans la perspective de la théorie Sens-Texte, voir section 1.2.3.1. *Le prédicat nominal dans la théorie Sens-Texte*, Chap. II.

apparaissent en premier lieu (dans notre cas de figure, les dérivations sémantiques V0, A0, S1, S2) ; ensuite, les relations paradigmatiques (les collocations)³⁷⁶.

Enfin, le champ phraséologie recense l'ensemble d'unités phraséologiques construites sur l'unité lexicale en question.

DiCo a été développé dans la perspective d'en faire la source d'autres produits lexicographiques. Ainsi, cette base lexicale est à la base de trois dictionnaires différents, qui rendent accessibles les données DiCo aux apprenants de français et au grand public, ne possédant pas forcément des connaissances sur le formalisme de la théorie Sens-Texte. À continuation, nous ferons une brève présentation de ces dictionnaires.

3.2.2.1. DiCouèbe

Le DiCouèbe³⁷⁷ (cf. Jousse et Polguère, 2005 ; Steinlin, Kahane, Polguère et El Ghali, 2004) est une interface de type formulaire qui permet de faire des requêtes sur la base lexicale DiCo et ainsi d'améliorer les possibilités de consultation par des utilisateurs humains. Le formulaire de requête en mode standard³⁷⁸ est donné ci-dessous :

³⁷⁶ Nous n'expliquons pas en détail les fonctions lexicales décrites dans l'entrée de la lexie meurtre, mais nous trouvons que la plupart d'entre elles sont compréhensibles grâce aux valeurs fournies.

³⁷⁷ Disponible sur ce lien :

http://idefix.ling.umontreal.ca/dicouebe/main.php?MODE_D_ACCES=acces_standard.

³⁷⁸ Il existe deux modes de consultation : le mode standard, qui donne accès aux champs fondamentaux dans la structuration du réseau lexical ; et le mode expert, qui donne accès à tous les champs et qui est utile uniquement aux lexicographes du DiCo (cf. Jousse et Polguère, 2005).

Aucune requête effectuée pour l'instant.

Chercher Chercher (nouvelle fenêtre) Effacer tout Effacer tous les champs

Mode expert

LEXIE

nom vocable [lexie:vocable]

no. acception [lexie:num]

carac. grammaticales [lexie:cgs]

étiquette sém. [lexie:formuleEtiquette]

LIENS DE FONCTIONS LEXICALES + FL(2)...

fonction lexicale [FL:formuleFL]

glose [FL:glose]

valeur fusionnée? [FL:estFusionnee]

marque d'usage [FL:marqueDUsage]

valeur [FL:lexie]

régime valeur [FL:regime]

contrainte [FL:contrainte]

exemple [FL:exemple]

LOCUTIONS CONSTRUITES AVEC LE MOT-CLÉ + phraseme(2)...

locution [phraseme:phraseme]

EXEMPLES D'EMPLOI DU MOT-CLÉ + exemple(2)...

phrase d'exemple [exemple:exemple]

afficher requête SQL

Figure 10 : Formulaire de requête du DiCouèbe (extrait de <http://olst.ling.umontreal.ca/dicouebe/main.php>)

Ainsi, si l'on effectue une requête de l'unité lexicale *meurtre*, on obtiendra les mêmes informations que nous avons vues dans la Figure 10 correspondant à la base lexicale DiCo :

lexie vocable	lexie cgs	lexie formule FL	FL glose	FL lexie	FL régime	FL contrainte	phraseme
MEURTRE	nom, masc	crime de sang QSyn		assassinat			_appel au meurtre_
MEURTRE	nom, masc	crime de sang QSyn		homicide-n			_appel au meurtre_
MEURTRE	nom, masc	crime de sang QSyn		crime#l.b			_appel au meurtre_
MEURTRE	nom, masc	crime de sang V0		tuer	[N=Y]		_appel au meurtre_
MEURTRE	nom, masc	crime de sang A0		meurtrier-adj			_appel au meurtre_
MEURTRE	nom, masc	crime de sang S1	Nom pour X	auteur	[de ART ~]		_appel au meurtre_
MEURTRE	nom, masc	crime de sang S1	Nom pour X	meurtrier-n			_appel au meurtre_
MEURTRE	nom, masc	crime de sang S1	Nom pour X	criminel-n			_appel au meurtre_
MEURTRE	nom, masc	crime de sang S2	Nom pour Y	victime	[de ART ~]		_appel au meurtre_
MEURTRE	nom, masc	crime de sang Magn	Très choquant	atroce			_appel au meurtre_
MEURTRE	nom, masc	crime de sang Magn	Très choquant	affreux			_appel au meurtre_
MEURTRE	nom, masc	crime de sang Magn	Très choquant	brutal			_appel au meurtre_
MEURTRE	nom, masc	crime de sang Magn	Très choquant	horrible			_appel au meurtre_
MEURTRE	nom, masc	crime de sang Magn	Très choquant	inqualifiable			_appel au meurtre_
MEURTRE	nom, masc	crime de sang Magn	Très choquant	odieux			_appel au meurtre_
MEURTRE	nom, masc	crime de sang A2Prepar1	Qui a été préparé	avec préméditation		postpos	_appel au meurtre_
MEURTRE	nom, masc	crime de sang A2Prepar1	Qui a été préparé	prémédité		postpos	_appel au meurtre_
MEURTRE	nom, masc	crime de sang A2Prepar1	Qui a été préparé	assassinat			_appel au meurtre_
MEURTRE	nom, masc	crime de sang Tel qu'il y a deux/trois/quatre Y		double		antépos	_appel au meurtre_
MEURTRE	nom, masc	crime de sang Tel qu'il y a deux/trois/quatre Y		triple		antépos	_appel au meurtre_
MEURTRE	nom, masc	crime de sang Tel qu'il y a deux/trois/quatre Y		quadruple		antépos	_appel au meurtre_
MEURTRE	nom, masc	crime de sang Oper1	[X] commettre ~	accomplir	[ART ~]		_appel au meurtre_
MEURTRE	nom, masc	crime de sang Oper1	[X] commettre ~	commettre	[ART ~]		_appel au meurtre_
MEURTRE	nom, masc	crime de sang Oper1	[X] commettre ~	perpétrer	[ART ~]		_appel au meurtre_
MEURTRE	nom, masc	crime de sang Oper1	[X] commettre ~	tremper	[dans ART ~]		_appel au meurtre_

Tableau 6 : Entrée meurtre du DiCouèbe (adaptée au format Excel, par souci de lisibilité)³⁷⁹

3.2.2.2. DiCoPop

Le DiCoPop³⁸⁰ (cf. Polguère, 2008b) est une interface pour naviguer de manière simplifiée dans les données du DiCo. Le DiCoPop adopte la forme d'un dictionnaire électronique (au sens de Atkins, 1996³⁸¹) dont les descriptions lexicographiques sont construites de manière entièrement automatique en réponse aux requêtes adressées à la base de données DiCo. Le mode de consultation et la disposition des informations de cette interface sont bien plus simples que sur le DiCouèbe. Le DiCoPop offre de multiples modes de navigation dans les données lexicales : une navigation séquentielle à partir de la liste de lexies ordonnées par ordre alphabétique ; une toute novatrice

³⁷⁹ Nous ne montrons qu'une partie du tableau de résultats de l'unité lexicale *meurtre* dans le DiCouèbe.

³⁸⁰ Disponible sur ce lien : <http://olst.ling.umontreal.ca/dicopop/liste.php>.

³⁸¹ Atkins (1996) signale les avantages suivants des dictionnaires électroniques : « hypertext functionality eliminating linear text restrictions and opening the way to new types of information by offering new ways of presenting it; no space constraints other than the need to avoid swamping the user; no distortion of the source language description by the needs of the target language; flexible compiling liberated from alphabetical order; alternative ways of presenting information, as for example graphics; rapid access to large amounts of lexicographical evidence in corpora; large-scale user customization » (p. 11).

navigation basée sur la hiérarchie des étiquettes sémantiques (Figure 11) ; et une navigation fondée sur le réseau de liens paradigmatiques et syntagmatiques. Ainsi, selon Polguère (2008b), l'un des grands avantages de Dicopop est qu'il permet « d'obtenir des descriptions qui ne gravitent pas nécessairement autour d'une unité lexicale donnée » (p. 1283), mais aussi autour de l'ensemble de données constituant la description lexicographique. Nous montrons ci-dessous une partie de la structuration hiérarchique des étiquettes sémantiques (mots en minuscules) et des unités lexicales (mots en majuscules) telle qu'elle est présentée dans le DiCoPop.



Figure 11 : Exemple de navigation hiérarchique dans le DiCoPop (extrait de <http://olst.ling.umontreal.ca/dicopop/hes.php>)

À titre d'exemple, nous montrons également une partie du résultat de la requête de la lexie *meurtre* dans le DiCoPop :

« *fiche précédente* » **MEURTRE** *fiche suivante* »
 nom, masc

crime de sang

~ de l'**individu Y** par l'**individu X**

*C'est ici que ce triple meurtre odieux a été commis.
 Soupçonné du meurtre de son épouse, il a été arrêté par les gendarmes mercredi.
 Il devrait comparaître aux assises dans trois semaines comme auteur présumé du meurtre d'un quinquagénaire.
 Il a été reconnu coupable sur la personne de Jules Gémal.*

 assassinat, homicide_n; crime#I.b

Verbe
 tuer [N_v]

Adjectif
 meurtrier_{adj}

Nom pour X
 auteur [de ART ~] //meurtrier_n; criminel_n

Nom pour Y
 victime [de ART ~]

Très choquant
 atroce, affreux, brutal, horrible, inqualifiable, odieux

Qui a été préparé
 avec préméditation, prémédité | *postpos* //assassinat

Tel qu'il y a deux/trois/quatre Y
 double, triple, quadruple | *antépos* [Les victimes de ce double meurtre sont un père de famille et son fils de 15 ans.]

[X] commettre ~
 accomplir, commettre, perpétrer [ART ~]; tremper [dans ART ~] [Il a refusé de tremper dans ce meurtre odieux.]

Figure 12 : Entrée meurtre du DiCoPop (extrait de <http://olst.ling.umontreal.ca/dicopop/lexie.php?id=467>)

Dans un souci de vulgarisation, les entrées du DiCoPop ne contiennent pas des noms de fonctions lexicales. Les relations lexicales sont encodées au moyen de formules facilement compréhensibles³⁸² telles que « Nom pour X » ou « Très choquant ». On utilise aussi « Verbe » pour V0 et « Adjectif » pour A0, voire des symboles (☞) pour marquer la synonymie. Ces formules constituent donc une sorte de paraphrases des fonctions lexicales.

3.2.2.3. Lexique actif du français (LAF)

³⁸² Elles sont également présentes dans la base de données DiCo entre le balisage /*...*/ (p. ex. : /*Nom pour X*/ ou /*Très choquant*/).

Nous avons tenu à inclure dans cette présentation cet ouvrage papier, parce qu'il est entièrement dérivé de la base de données lexicales du DiCo et qu'il est très proche des deux interfaces que nous venons de présenter, notamment du DiCoPop. Le *LAF* (Mel'čuk et Polguère, 2007) a un fort caractère pédagogique. Il s'agit d'un dictionnaire de collocations et de dérivés sémantiques dont l'objectif est de rendre accessibles les fonctions lexicales au grand public. Le *LAF* constitue donc la première tentative pour combler le fossé entre la lexicographie « théorique » (p. ex. le *Dictionnaire explicatif et combinatoire*, Mel'čuk et al. 1984, 1988, 1992, 1999) et la lexicographie « commerciale »³⁸³.

Selon Polguère (2008b), « il tire son originalité du fait qu'il est à la fois un manuel de lexicologie destiné, en tout premier lieu, aux enseignants de langue et un échantillon de dictionnaire du français » (p. 1282). En effet, le *LAF* est divisé en deux parties. Dans la première partie sont présentées les notions de base que les auteurs considèrent au cœur du fonctionnement du lexique. La seconde partie est le dictionnaire proprement dit. Elle contient un ensemble de descriptions lexicographiques qui constituent une sorte de mini-dictionnaire et servent de support descriptif du manuel. Malgré sa courte nomenclature, cet ouvrage offre un traitement profond des collocations sur les plans syntaxique et sémantique³⁸⁴.

Nous montrons ci-dessous l'entrée de l'unité lexicale *meurtre* dans le *LAF* :

³⁸³ Polguère (2000) note que

³⁸⁴ Tutin (2005) le qualifie comme « le dictionnaire de collocations grand public le plus ambitieux et le plus audacieux » (p. 45). Cette autrice attirera l'attention aussi sur « le peu d'informations sur les contextes d'emploi, les alternances syntaxiques, les fréquences ou les marques d'usage » dans le *LAF* (Tutin, 2008b, p. 49).

MEURTRE, nom, masc

CRIME DE SANG

Meurtre de l'individu Y [= de N, A_{poss}, soutenu sur la personne de N] par l'individu X [= de N, A_{poss}, par N]

☞ assassinat, homicide_N; crime I.b

Verbe tuer [N_v] **Adjectif** meurtrier_{Adj} **Nom pour X** auteur [de ART ~] //meurtrier_N; criminel_N **Nom pour Y** victime [de ART ~] **Très choquant** atroce, affreux, brutal, horrible, inqualifiable, odieux **Qui a été préparé avec préméditation, prémédité** | postpos //assassinat **Tel qu'il y a deux/trois/quatre Y** double/triple/quadruple | antépos [Les victimes de ce double meurtre sont un père de famille et son fils de 15 ans.] [X] **commettre un M.** accomplir, commettre, perpétrer [ART ~]; tremper [dans ART ~] [Il a refusé de tremper dans ce meurtre odieux.] [Qqn./Qqch.] **causer que X commette un M.** pousser [N_x au ~] **Raison d'un M. mobile** [de ART ~] [Qqn.] **s'occuper d'un M.** enquêter [sur ART ~]; élucider [ART ~], trouver l'auteur de ART ~; punir, châtier [ART ~]; venger [ART ~] **Service de police qui s'occupe des M.** brigade criminelle_{Adj} I.1 [X] **préparer un M.** comploter, préméditer, préparer [ART ~] [X] **essayer de faire croire qu'un M. est un N** maquiller [ART ~ en N] [Il a maquillé ce meurtre en accident/suicide.] **Fait de tenter un M.** tentative [de ~] **Ce qu'on crie quand on risque d'être victime d'un M.** <quand on voit que qqn. risque d'être victime d'un M.> « Au ~ ! »

C'est ici que ce triple meurtre odieux a été commis. Soupçonné du meurtre de son épouse, il a été arrêté par les gendarmes mercredi. Il devrait comparaître aux assises dans trois semaines comme auteur présumé du meurtre d'un quinquagénaire. Il a été reconnu coupable de meurtre sur la personne de Jules Gémal.

◇ «appel au meurtre», «crier au meurtre»

Figure 13 : Entrée meurtre du LAF (extrait de Mel'čuk et Polguère, 2007, p. 312)

Nous pouvons donc observer que le contenu des articles du LAF (y compris son adaptation dans un effort de « vulgarisation ») est le même que celui du DiCoPop³⁸⁵. Les différences quant à la mise en page (séparation en différentes lignes) relèvent des contraintes d'espace qui s'appliquent aux dictionnaires papier³⁸⁶.

3.2.3. Réseau Lexical du Français (RL-fr)

Le Réseau Lexical du français (désormais RL-fr) (Lux-Pogodalla et Polguère, 2011 ; Polguère, 2014) est un modèle formel du lexique du français contemporain en cours de construction au laboratoire ATILF³⁸⁷ construit selon les principes théoriques et méthodologiques de la Lexicographie Explicative et Combinatoire (cf. Mel'čuk et al., 1995 ; Mel'čuk, 2006a).

³⁸⁵ Nous tenons à signaler que le projet d'élaboration du LAF est antérieur au dictionnaire électronique DiCoPop.

³⁸⁶ Sur la méthodologie d'élaboration du LAF, voir Polguère (2000a).

³⁸⁷ Analyse et traitement informatique de la langue française (UMR 7118, CNRS – Université de Lorraine).

Du point de vue formel, cette ressource lexicale de nouvelle génération constitue un système lexical, autrement dit une modélisation du lexique sous forme de graphe (cf. Polguère, 2009, 2014, 2015a, 2016). Son architecture est un réseau d'unités lexicales reliées entre elles par une large gamme de liens codant des relations lexicales sémantiques ou combinatoires. Selon Polguère (2014), « lexical systems are graph metaphors of natural language lexicons and their construction should be performed as a graph weaving rather than text writing activity » (p. 13). Face aux modèles textuels bidimensionnels (les dictionnaires), les systèmes lexicaux sont des entités multidimensionnelles qui représenteraient de manière plus plausible les lexiques mentaux (cf. Aitchison, 2003, Wierzbicka, 2009)³⁸⁸. La structure non ontologique des systèmes lexicaux permet de mettre en valeur la nature relationnelle des lexiques, réfutant ainsi les classifications hiérarchiques des unités lexicales en classes et sous-classes et l'héritage des propriétés. La structure du réseau, du type *petit monde* (cf. Gader et al., 2014 ; Watts et Strogatz, 1998), est composée de sommets (les sens lexicalisés de la langue, c'est-à-dire les unités lexicales et des unités quasi-lexicales telles que les clichés linguistiques, les proverbes, etc.) et d'arcs (les liens paradigmatiques et syntagmatiques)³⁸⁹.

Le RL-fr fait partie d'une nouvelle génération de ressources lexicales du type *dictionnaire virtuel* (Selva et al., 2003). Les données lexicographiques ne sont pas stockées sous forme de texte, mais comme des bases de données (SQL) d'entités connectées formant un graphe. L'éditeur lexicographique Dicet (Gader et al., 2012) structure ces données comme un graphe d'unités lexicales, mais il permet également la génération automatique de représentations textuelles du type articles du dictionnaire, avec des hyperliens vers les unités lexicales ciblées. De cette manière, les données lexicographiques peuvent être visualisées sous forme de réseau de sens lexicaux (Figure 14) ou sous forme de texte lexicographique (Figure 15).

³⁸⁸ Le terme *lexique mental* est utilisé pour désigner les connaissances lexicales stockées dans l'esprit du locuteur.

³⁸⁹ Polguère (2014, p. 7) fait une présentation détaillée des composantes des systèmes lexicaux.

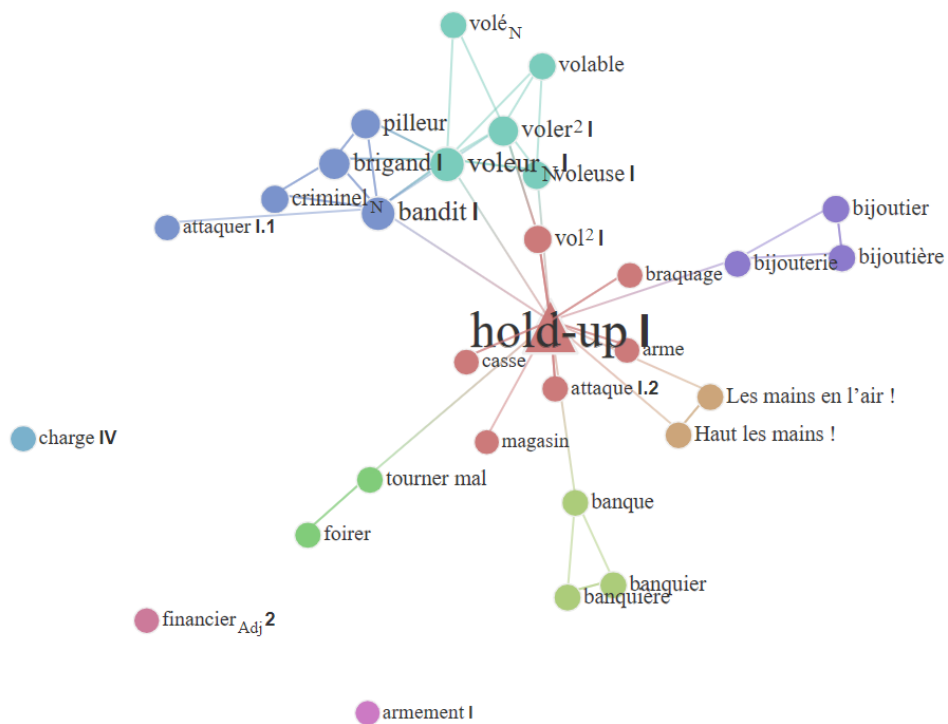


Figure 14 : Graphe de l'unité lexicale *hold-up I* du RL-fr (extrait de https://spiderlex.atilf.fr/fr/q/*hold-up***)

Ce graphe (Figure 8) illustre l'espace sémantique contrôlé par l'unité lexicale *hold-up I*. La longueur des arcs et la couleur des sommets représentent de manière visuelle la proximité sémantique entre les unités lexicales appartenant à cet espace sémantique.

[CG]	
angl	
nom	
masc	
invar	
graphie rectifiée : « holdup »	
[DF]	
méfait	
hold-up de X contre Y	
=	
x est un individu y est un lieu public	
attaque par x de y	
au moyen d'une arme	
afin d'y commettre un vol	
[FL]	
Syn _c	● attaque I.2
Syn ₀	● vol ² I ; ● fam braquage ; ● fam casse
S _{1c}	● bandit I, ● voleur _N I, ● voleuse I
S ₂ ^{prototyp}	● banque ; ● magasin ; ● bijouterie
S _{instr}	● arme
Oper ₁	● accomplir I [ART ~], ● commettre [ART ~], ● exécuter I.1 [ART ~], ● faire II.1 [ART ~], ● perpétrer [ART ~], ● réaliser II [ART ~]

Figure 15 : Échantillon de l'article lexicographique de l'unité lexicale *hold-up I* du RL-fr (extrait de https://spiderlex.atilf.fr/fr/q/*hold-up***)

Chaque sommet du graphe constitue un *article virtuel* contenant une description lexicographique de l'unité lexicale donnée. Nous expliquons ci-dessous les caractéristiques d'un article virtuel du RL-fr, en relation au contenu de la Figure 15 :

- Les caractéristiques grammaticales : la partie du discours, le genre, la flexion.
- La définition : une étiquette sémantique (cf. Polguère, 2011a) (*méfait*), la structure actancielle (*hold-up de x contre y ; x est un individu y est un lieu public*) et la définition sous forme de paraphrase (*attaque de x par y au moyen d'une arme afin d'y commettre un vol*).
- Les connexions paradigmatisées et syntagmatisées standardisées moyennant les fonctions lexicales de la théorie Sens-Texte (*Syn_c*, *S_{instr}*, *Oper₁*, etc.). Ces connexions sont implémentées comme des arcs du graphe.

- d) Des exemples : soit des citations d'un corpus, soit des adaptations artisanales de corpus ou de données Internet.
- e) La phraséologie : des locutions ou des clichés qui contiennent le mot vedette (dans notre cas de figure, il n'y en a pas).

La structure de l'article virtuel du RL-fr est assez similaire à celle du DiCo, puisqu'ils s'inscrivent dans le cadre de la Lexicographie Explicative et Combinatoire. Néanmoins, seulement le RL-fr constitue un véritable système lexical car les données sont structurées sous forme de réseau de connexions sémantiques et combinatoires entre les unités lexicales. Par contre, dans le DiCo, étant un modèle textuel bidimensionnel de lexique, les connexions décrites relient une unité lexicale donnée à une série d'unités lexicales comme le fait n'importe quel autre dictionnaire standard, mais aucune connexion n'est implémentée au niveau de la structure des données.

Le RL-fr entretient aussi une relation étroite avec WordNet car il a été conçu pour appartenir à la famille -Net (WordNet, FrameNet³⁹⁰, etc.). En effet, ces ressources sont structurées en réseau, ne sont pas un modèle de lexique « textuel » linéaire et sont construites manuellement par une équipe de lexicographes. La grande originalité du RL-fr réside sur sa vision globale de toutes les propriétés lexicales : définition lexicographique, définition grammaticale, fonctionnalités, combinatoire syntaxique, combinatoire lexicale, liens dérivationnels, etc. (Lux-Pogodalla et Polguère, 2011) ; tandis que les ressources -Net ne visent qu'un ensemble limité de propriétés, comme l'organisation hiérarchique des *synsets* (Wordnet) ou les *frames* contrôlant les interconnexions sémanticosyntaxiques FrameNet).

En ce qui concerne les perspectives d'exploitation, comme le RL-fr est conçu comme une base de données lexicales, il est possible d'en tirer de nombreuses ressources lexicales telles que des lexiques pour des systèmes de TAL, des entrées de dictionnaires similaires à celles des dictionnaires standards ou des ressources lexicales en ligne pour l'enseignement/apprentissage des langues (Lux-Pogodalla et Polguère, 2011).

3.2.4. Definiens

³⁹⁰ La base de données FrameNet sera décrite dans la section 3.3.1. *FrameNet*.

Le projet Definiens (cf. Barque et Polguère, 2009 ; Barque et al., 2010)³⁹¹ vise la construction d'une base de données de sémantique lexicale du français suffisamment structurée qui permette le traitement automatique du point de vue de la sémantique. Pour pouvoir y parvenir dans un délai relativement court, l'équipe de Barque et Polguère ont décidé de travailler sur les définitions du *Trésor de la Langue Française informatisé* (désormais TLFi).

Definiens est donc une base de définitions lexicographiques du TLFi explicitement structurées, c'est-à-dire chaque définition lexicographique est munie d'un balisage formel (cf. Barque et Polguère, 2009 ; Barque et al., 2010)³⁹². La conversion des définitions du TLFi sous une forme structurée a été faite essentiellement de manière manuelle. Definiens propose une structuration des définitions du TLFi en *composante centrale* (genre prochain) et *composantes périphériques* (différences spécifiques) et un balisage indiquant « le rôle joué par chaque composante dans la caractérisation du sens de l'unité définie (Barque et Polguère, 2009, p. 35). Nous montrons ci-dessous l'étiquetage sémantique de la définition du nom *brouette* :

```
<PARAPH>
<CC=véhicule>Véhicule</CC>
<CP=parties caractéristiques>à une roue et à deux brancards</CP>
<CP=fonction>servant au transport des matériaux</CP>
</PARAPH>
```

Figure 16 : Étiquetage sémantique de *brouette* dans *Definiens* (extrait de Barque et al., 2010, p. 251)

Cette base de données de définitions lexicographiques structurées est d'un grand intérêt autant pour le TAL que pour la recherche en linguistique lexicographique. Ainsi, Definiens a favorisé la valorisation du TLFi dans le contexte d'applications en linguistique informatique ou en TAL.

³⁹¹ Malheureusement, le projet Definiens n'est pas encore rendu disponible sur le site du CNRTL.

³⁹² Malheureusement, le projet Definiens n'est pas encore rendu disponible sur le site du CNRTL.

3.3. Ressources pour l'étiquetage en rôles sémantiques

La description des arguments peut se faire en types sémantiques (traits sémantiques) ou en rôles sémantiques. Les types sémantiques sont des classes sémantiques susceptibles d'instancier les arguments d'un prédicat (*individu, objet physique, locatif*, etc.). D'autre part, les rôles sémantiques sont les relations sémantiques qu'un prédicat entretient avec ses arguments (*agent, patient, instrument*, etc.).

Il n'existe pas un accord entre les différentes approches linguistiques sur le nombre de rôles sémantiques. Parmi ces approches, sont spécialement intéressantes pour le TAL la *sémantique des cadres* de Fillmore (1968, 1976, 1982), la *Lexical Conceptual Structure* de Jackendoff (1972, 1993), les *proto-rôles* de Dowty (1991) et les *classes verbales et alternances* de Levin (1993)³⁹³. Nous présentons ci-dessous trois ressources d'étiquetage en rôles sémantiques avec des structures prédicats-arguments. Ces ressources sont fondées sur les approches mentionnées ci-dessus. Si nous avons choisi ces trois ressources, c'est parce qu'elles ont donné lieu à des résultats tangibles et assez satisfaisants.

3.3.1. FrameNet

FrameNet³⁹⁴ est une base de données lexicale et un corpus annoté, développé à l'International Computer Science Institute à Berkeley University depuis 1997 (cf. Baker et al., 1998 ; Bouveret et Fillmore, 2008 ; Fillmore et al., 2003 ; Fillmore et al., 2004 ; Ruppenhofer et al. 2016). La théorie sous-jacente à FrameNet est la sémantique de cadres (*semantic frames*) de Fillmore (1968, 1976, 1982) dont le postulat de base veut que le sens de la plupart des mots peut être mieux compris sur la base d'un cadre sémantique. Ainsi, dans cette approche théorique, les sens lexicaux sont décrits au moyen de cadres sémantiques, qui comportent un noyau et des participants obligatoires (*frame elements* ou participants). Cette approche théorique est particulièrement centrée sur les rôles sémantiques des arguments. Le travail dans ce projet consiste donc à définir

³⁹³ Pour approfondir sur les différents aspects de l'étiquetage sémantique et avoir une vision générale des différentes approches et ressources, voir Palmer et al. (2010).

³⁹⁴ Disponible sur ce lien : <https://framenet.icsi.berkeley.edu/>.

les cadres sémantiques (*frames*) et à annoter les phrases pour illustrer comment les participants (*frame elements*) s'adaptent syntaxiquement à l'unité lexicale évoquée par le cadre.

En ce qui concerne le corpus annoté³⁹⁵, chaque unité lexicale des phrases annotées est associée à des patrons de réalisation illustrant les liens observés avec les participants (*frame elements*) et leurs réalisations syntaxiques (p. ex., fonction grammaticale, type de phrase, etc.). Les phrases annotées en XML constituent la base sur laquelle se construit la base de données lexicales FrameNet. Ce format prend en charge les requêtes par unité lexicale, par cadre, par participant ou par combinaison de ceux-ci.

La base lexicale contient la description des cadres, des participants, des unités lexicales et des relations entre les différentes entités, issues de l'annotation du corpus. Les unités lexicales évoquées par un cadre entretiennent des relations de synonymie, de conversivité et de conversion. De même, les cadres entretiennent entre eux des relations sémantiques, ce qui permet de placer chaque cadre dans un réseau sémantique global : hyponymie, hypéronymie, contiguïté, présupposition, méronymie (*sub-frames*), et des relations de perspective³⁹⁶. Quant au traitement de la polysémie, l'approche est discrète : aucune description des liens de polysémie n'y apparaît et on ne fait pas la distinction entre polysémie et homonymie.

Nous montrons ci-dessous un échantillon du frame *assistance* (Figure 17) et de ses *frame elements* (Figure 18).

³⁹⁵ Le corpus utilisé est le British National Corpus et des articles journalistiques du Linguistic Data Consistorium.

³⁹⁶ La relation de perspective est la relation existante entre des cadres à contenus identiques, mais qui sont pris d'un point de vue différent ; par exemple, le cadre « *employment_start* » ('commencer travailler') a deux formes en relation de perspective : « *hiring* » ('embaucher') et « *get-a-job* » ('décrocher un emploi') (cf. Cartier, 2009) ; les verbes *acheter* / *vendre* constituent deux *frames* en relation de perspective (« *commerce_buy* » et « *commerce_sell* ») ayant les mêmes rôles sémantiques (*buyer*, *seller*, *goods*, *money*, etc.).

Assistance

Definition:



A **Helper** benefits a **Benefited party** by enabling the culmination of a **Goal** that the **Benefited party** has. A **Focal entity** that is involved in reaching the **Goal** may stand in for it. This Frame has unique Frame-to-Frame Relations: it inherits from and uses the Intentionally_act frame. This is atypical of FrameNet's frame relations.

Will you **HELP** the Government find your brother?

Maybe Stephen should **ASSIST** him with the last manuscript.

They **HELPED** me psychologically to overcome the physical loss I had suffered.

You have **HELPED** him tremendously by showing him how to stand up for himself and by being his friend.

By bringing assistance to his troops wherever they were in trouble he **AIDED** them greatly.

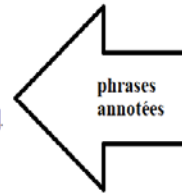


Figure 17 : Définition et phrases du frame assistance (extrait de <https://framenet.icsi.berkeley.edu/frameIndex>)

La caractérisation du frame mêle une description sémantique et des informations méta-linguistiques. Force est de constater que pour l'étiquetage sémantique des participants, on préfère utiliser des noms spécifiques à la situation dénotée (p. ex. : *Helper*, *Benefited_party*, *Focal_entity* etc.), au lieu des noms des rôles thématiques standards (p. ex. : *agent*, *patient*, *goal*, etc.). Dans les phrases annotées servant comme exemple, les couleurs montrent l'étiquetage en rôles sémantiques. Les mots en majuscules correspondent aux unités lexicales faisant partie du frame *assistance*.

FEs:

Core:

Benefited party [Ben]

The **Benefited party** receives a benefit from the action of the **Helper**.

Focal entity [Focal_e]

This FE identifies a **Focal entity** involved in achieving the **Goal**.
Whoever didn't cook has to **HELP** with the dishes.

Goal [Goal]

The desirable state of affairs that the **Benefited party** is involved in and which is enabled by the **Helper**.
Jack **HELPED** Jill climb Mt. Everest.

Helper [Help]

The **Helper** performs some action that benefits the **Benefited party**.

Non-Core:

Degree [deg]

Semantic Type: Degree

The measure to which the **Helper**'s assistance brings the **Benefited party** closer to accomplishing their **Goal**.
The Algebra Buster software **HELPED** my daughter very much.

Domain [dom]

The aspect of accomplishing the **Goal** to which the **Helper**'s assistance is relevant.
How can you be **HELPED** spiritually by his music, when he is defying God's Word with his very appearance?

Duration [dur]

Semantic Type: Duration

The time period during which the **Helper** gives assistance to the **Benefited party**.
Several nations **HELPED** ours during our most difficult years.

Figure 18 : Échantillon des frame elements du frame assistance (extrait de <https://framenet.icsi.berkeley.edu/frameIndex>)

Parmi les participants (*FEs*), on distingue les participants principaux (*Core* ; en l'occurrence, *Benefited_party*, *Focal_entity*, *Goal* et *Helper*) et les secondaires (*Non-Core* ; en l'occurrence, *Degree*, *Domain*, *Duration*, *Explanation*, *Frquency*, *Instrument*, *Manner*, *Means*, *Place*, *Purpose*, *Time*). Ils sont tous caractérisés (p. ex. : *Benefited_party* [*Ben*] = *The Benefited_party receives a benefit from the action of the Helper*) et souvent ils sont accompagnés d'une phrase annotée (p. ex. : *Jack HELPED Jill climb Mt. Everest*).

En ce qui concerne la description des unités lexicales (Figure 19 ci-dessous), elle contient l'indication du cadre évoqué par l'unité (*Assistance*), la définition extraite du Concise Oxford English Dictionary³⁹⁷ (*help or support*), l'indication des éléments du cadre relevant de la valence sémantique de l'unité lexicale en question, leur réalisation syntaxique et des phrases annotées.

aid [act].n

Frame: Assistance

Definition:

COD: help or support

Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

Frame Element	Number Annotated	Realization(s)
Benefited_party	(2)	DNI.-- (2) INI.-- (2)
Focal_entity	(3)	N.Dep (3)
Goal	(2)	DNI.-- (1) INI.-- (1)
Helper	(7)	PP[from].Dep (1) A.Dep (1) INI.-- (3) PP[of].Dep (2)

Valence Patterns:

These frame elements occur in the following syntactic patterns:

Number Annotated	Patterns		
3 TOTAL	Benefited_party	Focal_entity	Helper
(2)	INI --	N Dep	INI --
2 TOTAL	Benefited_party	Goal	Helper

Figure 19 : Échantillon de la description de l'unité lexicale *aid* (extrait de <https://framenet.icsi.berkeley.edu/frameIndex>)

³⁹⁷ D'autres fois la définition est élaborée par des lexicographes de FrameNet.

Des versions en français et en espagnol ont été développées. La structure FrameNet (cadres + relations entre cadres) étant fondamentalement sémantiques, elle peut être considérée comme conceptuelle. En effet, les généralisations obtenues dans FrameNet en anglais sont majoritairement faites indépendamment de la syntaxe de l'anglais, ce qui rend cette structuration compatible avec la description d'autres langues. Ainsi, le projet Asfalda³⁹⁸ (Candito et al., 2014) coordonné par l'Université Paris Diderot a développé un FrameNet du français (French FrameNet³⁹⁹). L'Universitat Autònoma de Barcelona et l'International Computer Science Institute de Berkeley University ont développé le FrameNet de l'espagnol (Spanish Framenet⁴⁰⁰) (cf. Subirats, 2013a, 2013b ; Subirats et Petruck. 2003). À part ces deux langues, des projets en allemand, japonais, suédois, brésilien, entre autres langues, ont déjà été lancés⁴⁰¹.

Pourtant, ce qui constitue un avantage pour la transposition de FrameNet à d'autres langues, suppose un problème majeur pour le TAL : la relation entre le niveau sémanticoconceptuel et le niveau syntagmatique n'est pas explicitée. Cette ressource vise surtout la description sémanticoconceptuelle, en délaissant la structure valencielle de surface. Comme conséquence de cette prise de position, il n'y a pas de description canonique des possibles séquences syntagmatiques pour chaque cadre (seul des exemples de phrases sont donnés, mais ils ne couvrent pas toutes les possibilités). Ainsi, il devient très difficile, voire impossible, de distinguer les insertions dans les formes canoniques, puisque les formes canoniques n'ont pas été explicitées.

D'autres inconvénients pour le TAL s'affichent : le nombre de cadres décrits (1224 cadres et 13 687 unités lexicales en anglais) reste insuffisant pour le TAL, le degré de granularité des cadres est irrégulier, et des incohérences ont été repérées dans les relations entre les cadres. Néanmoins, FrameNet est l'une des ressources lexicales

³⁹⁸ Ce projet a été développé entre les années 2012 et 2016.

³⁹⁹ Disponible sur ce lien : <http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml>.

⁴⁰⁰ Disponible sur ce lien : <http://spanishfn.org/data>.

⁴⁰¹ Par ailleurs, un projet a été lancé visant l'élaboration d'une base de données des alignements des cadres et des éléments des cadres des FrameNets en plusieurs langues. Pour un aperçu des recherches en cours sur des ressources FrameNet en d'autres langues et les progrès du projet du FrameNet multilingue, voir Baker et al. (2018) ; Torrent et al. (2018) ; et le site web de FrameNet (https://framenet.icsi.berkeley.edu/framenets_in_other_languages).

les plus fines dans la description des rôles sémantiques qui aient été développées jusqu'à présent. FrameNet fournit un ensemble de données en rôles sémantiques qui sont utilisés dans des applications telles que l'extraction d'information ou la traduction automatique.

3.3.2. VerbNet

Ce lexique verbal de l'anglais⁴⁰² (cf. Kipper-Schuler et al., 2000 ; Schuler, 2005) est une application et extension des classes de verbes de Levin (1993). L'objectif est de décrire pour chaque sens verbal, toutes ses contraintes syntagmatiques (toutes les réalisations de surface des structures argumentales) et les contraintes liées aux valeurs sémantiques possibles de chaque actant au moyen d'une courte liste de rôles sémantiques (agent, thème, location, etc.)⁴⁰³. Cette ressource est le plus grand lexique en ligne de l'anglais. Chaque classe est caractérisée en extension (liste de tous les verbes appartenant à la classe) et en intension (définition des arguments avec des informations sémantiques et syntaxiques). VerbNet s'intéresse d'avantage au comportement syntaxique que FrameNet.

⁴⁰² Disponible sur ce lien : <https://verbs.colorado.edu/verbnet/>.

⁴⁰³ Tandis que la sémantique des cadres de Fillmore préfère une liste indéfinie de rôles sémantiques, dans VerbNet il n'y en a que vingt-trois, car ces rôles sont plus génériques.

HIT class

<<MEMBERS>> [*hit*, 1], [*kick*, 1], [*slap*, 1], [*tap*, 1], ...
 <<THEMATIC ROLES>> Agent(A), Patient(P), Instrument(I)
 <<SELECT RESTRICTIONS>> Agent[+animate],
 Patient[+concrete],
 Instrument[+concrete,-animate]

<<FRAMES and PREDICATES>>

Basic Transitive	A V P	manner(during(E),directedmotion,A) ∧ manner(end(E),forceful,A) ∧ contact(end(E),A,P)
Transitive with Instrument	A V P with I	manner(during(E),directedmotion,I) ∧ manner(end(E),forceful,I) ∧ contact(end(E),I,P)
Together reciprocal	A V P[+plural] together	manner(during(E),directedmotion,P _i) ∧ manner(during(E),directedmotion,P _j) ∧ manner(end(E),forceful,P _i) ∧ manner(end(E),forceful,P _j) ∧ contact(end(E),P _i ,P _j)
Resultative	A V P Adj	manner(during(E),directedmotion,A) ∧ manner(end(E),forceful,A) ∧ contact(end(E),A,P) ∧ Pred(result(E),P)
Resultative	A V P Adj with I	manner(during(E),directedmotion,I) ∧ manner(end(E),forceful,I) ∧ contact(end(E),I,P) ∧ Pred(result(E),P)
Resultative	A V P PP	manner(during(E),directedmotion,A) ∧ manner(end(E),forceful,A) ∧ contact(end(E),A,P) ∧ Pred(result(E),P)
Resultative	A V P PP with I	manner(during(E),directedmotion,I) ∧ manner(end(E),forceful,I) ∧ contact(end(E),I,P) ∧ Pred(result(E),P)
Conative	A V at P	manner(during(E),directedmotion,A)
Conative	A V at P with I	manner(during(E),directedmotion,I)
With/against alternation	A V I against/on P	manner(during(E),directedmotion,I) ∧ manner(end(E),forceful,I) ∧ contact(end(E),I,P)
Body-part object or reflexive object	A V I[+body-part/+refl]	manner(during(E),directedmotion,I) ∧ manner(end(E),forceful,I) ∧ contact(end(E),I,?)
Body-part object or reflexive object	A V I[+body-part/+refl] against/on P	manner(during(E),directedmotion,I) ∧ manner(end(E),forceful,I) ∧ contact(end(E),I,P)
Transitive	I V P	manner(during(E),directedmotion,I) ∧ manner(end(E),forceful,I) ∧ contact(end(E),I,P)
Resultative	I V P Adj	manner(during(E),directedmotion,I) ∧ manner(end(E),forceful,I) ∧ contact(end(E),I,P) ∧ Pred(result(E),P)
Resultative	I V P PP	manner(during(E),directedmotion,I) ∧ manner(end(E),forceful,I) ∧ contact(end(E),I,P) ∧ Pred(result(E),P)

Tableau 7: *Entrée de la classe hit dans VerbNet* (extrait de Kipper-Schuler et al., 2000, p. 3)

L'utilisation de rôles génériques dans l'étiquetage sémantique fait que la description perde de la granularité, mais la rend plus facilement exploitable en TAL.

Quant aux versions françaises de VerbNet, plusieurs ressources ont été produites, mais elles ne sont pas toutes de libre disposition ni sont suffisamment développées. Néanmoins, un projet de traduction vers le français est très intéressant :

VerbNet⁴⁰⁴ (cf. Danlos et al., 2014 ; Danlos et al., 2016 ; Pradet et al., 2014). Dans VerbNet la hiérarchie des 270 classes de VerbNet a été respectée autant que possible. Les *frames* syntaxiques de VerbNet ont été adaptés vers la langue française et les classes n'ont été réorganisées que quand il a été nécessaire de le faire. Pour faire correspondre les classes de VerbNet avec celles de VerbNet les tables du Lexique-Grammaire (M. Gross, 1975) et le LVF (Dubois et Dubois-Charlier, 1997) ont été utilisées.

En espagnol, le lexique AnCorà-Verb⁴⁰⁵ de l'Universitat de Barcelona reprend en partie VerbNet. Il contient 7284 connexions avec des sens verbaux de VerbNet : définition de la classe sémantique, la structure lexicosémantique, les schémas syntaxiques, les rôles sémantiques et les restrictions. En outre de ces informations, chaque sens verbal est associé à d'autres informations provenant d'autres ressources : FrameNet (le schéma conceptuel) ; PropBank⁴⁰⁶ (les arguments et rôles qui correspondent à chaque sens sémantique) ; WordNet 3.0 (les sens verbaux) ; et OntoNotes⁴⁰⁷ (les sens verbaux).

3.3.3. PropBank (Propositional Bank)

PropBank⁴⁰⁸ (cf. Palmer et al., 2005 ; Pradhan et al., 2022) utilise les annotations syntaxiques du Penn TreeBank⁴⁰⁹ (cf. Kingsbury et Palmer, 2002 ; Marcus et al., 1993).

⁴⁰⁴ Disponible sur ce lien : <https://github.com/aymara/verbenet> .

⁴⁰⁵ Disponible sur ce lien : http://clic.ub.edu/corpus/es/ancoraverb_es .

⁴⁰⁶ Nous décrivons cette ressource ci-dessous.

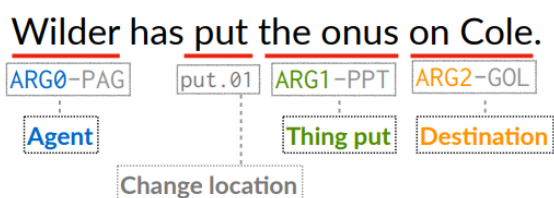
⁴⁰⁷ OntoNotes (cf. Weischedel, et al., 2013) est un corpus annoté, développé par BBN Technologies, l'University of Colorado, l'University of Pennsylvania et l'University of Southern Californias Information Sciences Institute. Il s'agit d'un corpus très large et riche du fait de sa diversité de genres (articles journalistiques, conversations téléphoniques, weblogs, émissions de radio et de télévision, etc.) et de sa diversité de langues (anglais, arabe et chinois). L'annotation offre des données sur la syntaxe, la structure argumentale des mots et sur la sémantique (le sens du mot en relation avec l'ontologie et la corréférence).

⁴⁰⁸ Disponible sur ce lien : <https://propbank.github.io/>.

Cette ressource est indépendante de quelconque théorie linguistique, bien qu'elle reçoive des influences de certaines théories linguistiques.

Comme c'était le cas chez VerbNet, le principe sous-jacent est que la syntaxe joue un rôle capital pour la désambiguïsation des sens et pour l'attribution de rôles sémantiques. Ce principe est héritier des alternances de Levin (1993), sauf que dans PropBank on n'établit pas de classes de verbes. L'autre grande différence avec VerbNet est l'utilisation du corpus. En effet, dans PropBank les différents sens des prédicats sont établis sur la base des structures argumentales et d'autres différences observables sur le corpus. Alors, l'étiquetage associé à un verbe spécifie le sens de celui-ci dans un contexte donné.

Pour faciliter son application en TAL, les étiquettes sont très générales. L'étiquetage sémantique est basé sur la théorie des proto-rôles de Dowty (1991). Les arguments sont marqués de Arg0 à Arg5. Seulement les marques 0 et 1 ont généralement la même valeur : « For a particular verb, Arg0 is generally the argument exhibiting features of a prototypical Agent (Dowty, 1991) while Arg1 is a prototypical Patient or Theme » (Palmer et al., 2005, p. 4). Le sens des autres arguments dépend du verbe : du sens du verbe, du type de constituants qui lui sont associés et du nombre de rôles qu'il présente dans une phrase donnée (cf. van der Plas et Apidianaki, 2014). En conséquence du manque de granularité, les rôles sont à interpréter en fonction du sens de chaque verbe. Pourtant, les douze étiquettes des adjoints sont plus spécifiques ; par exemple, ArgM-TMP pour le temps ; ArgM-MNR pour la manière ; ArgM-LOC pour le locatif ; ArgM-GOL pour le but, etc. Malgré le sens plus spécifique des étiquettes des adjoints, ceux-ci jouent un rôle secondaire dans la désambiguïsation et définition du sens du verbe. Nous montrons ci-dessous un exemple d'annotation :



⁴⁰⁹ Le Penn TreeBank est un corpus de référence issu des articles du *Wall Street Journal*. Il comprend un million de mots.

Figure 20 : Exemple de notation d'une phrase dans PropBank (extrait de Pradhan et al., 2022, p. 279)

Dans l'exemple de la Figure 20, le verbe *put* invoque le *roleset*⁴¹⁰ *put*.01 'changement de lieu' (*change location*), dans lequel le proto-agent (PAG) correspond à ARG0 et prend une valeur ARG0-PAG ; 'la chose mise' (*thing put*) est le proto-agent (PPT), qui prend la valeur ARG-1 ; et 'la destination' (*destination*) correspond au but (GOL), qui prend une valeur ARG-2GOL.

Ultérieurement, on a intégré dans PropBank des prédicats représentés par d'autres parties du discours. Nous voudrions attirer l'attention sur les prédicats nominaux, puisqu'il s'agit de l'objet de notre recherche. NomBank⁴¹¹ (cf. Meyers et al., 2004) a été conçu à l'image de PropBank. Cette nouvelle ressource couvre des nominalisations de verbes et des nominalisations d'adjectifs, des noms partitifs, des noms relationnels et des noms environnementaux. Souvent les verbes et les noms dérivés de ces verbes partagent les mêmes arguments sémantiques. Cela permet alors de réutiliser les fiches descriptives des verbes qui avaient été créées dans PropBank. Ainsi, ces deux projets sont coordonnés. Pour une phrase telle que *They gave the chefs a standing ovation*, PropBank et NomBank proposent la même annotation, mais l'une étant basée sur la structure prédicat-argument autour du verbe *gave*, et l'autre, sur le groupe nominal *a standing ovation* (Pradet, 2015).

Samardžić et al. (2010) ont soulevé deux grands avantages de PropBank. D'une part, le lexique est déterminé par le corpus. En effet, le lexique de PropBank est construit à partir de l'extraction et description de tous les prédicats qui apparaissent dans un corpus ; contrairement au lexique de FrameNet et de VerbNet, qui ne proviennent pas d'un corpus. Comme l'étiquetage est fait sur un corpus d'occurrences réelles, ces auteurs estiment que PropBank peut offrir une meilleure couverture⁴¹². D'autre part, le poids des hypothèses théoriques dans l'étiquetage est peu important dans PropBank. Tandis que les étiquettes de FrameNet contiennent une forte empreinte d'intuition linguistique, l'étiquetage de PropBank dépend surtout du comportement

⁴¹⁰ Les *rolesets* sont les structures argumentales des prédicats, associées à des sens généraux.

⁴¹¹ Disponible sur ce lien : <https://nlp.cs.nyu.edu/meyers/NomBank.html> .

⁴¹² Il faudrait préciser que ce corpus est limité au *Wall Street Journal*.

observable des mots. On s'attend donc à des critères plus tangibles pour décider de l'étiquetage de chaque élément de la structure argumentale trouvé dans le corpus.

Cependant, comme PropBank est très orienté vers la syntaxe, la rentabilisation de cette ressource pour d'autres langues devient impossible, bien que la méthodologie soit réutilisable. Différentes équipes de travail ont utilisé cette méthodologie pour faire un PropBank du français (cf. Gardent et Cerisaraï, 2010 ; van der Plas et al., 2010 ; van der Plas et Apidianaki, 2014). Quant à l'espagnol, les lexiques AnCora-Verb⁴¹³ et AnCora-Nom⁴¹⁴ (cf. Peris Morant et Taulé Delor, 2011), qui, rappelons-le, intègrent plusieurs ressources lexicales (PropNet, FrameNet, VerbNet et OntoNotes) sont les projets les plus développés de l'espagnol.

Dans cette section nous avons présenté des projets d'étiquetage en rôles sémantiques avec des structures prédicats-arguments. Nous avons pu constater que l'accent est porté sur la description du sens verbal, même si d'autres parties du discours peuvent bien assumer cette fonction. L'étude du nom prédicatif occupe une place secondaire, vis-à-vis des résultats obtenus. En outre, l'anglais est le grand privilégié dans ces études. D'autres travaux similaires ont été entamés en d'autres langues, mais la couverture est bien moindre.

3.4. Ressources pour l'étiquetage de l'aspect et de la temporalité

3.4.1. TimeBank

TimeBank 1.2⁴¹⁵ (Pustejovsky, et al., 2006) est un corpus de textes en anglais provenant de sources variées du domaine du journalisme (61 000 tokens⁴¹⁶). Ce corpus contient l'annotation des événements⁴¹⁷ et des expressions temporelles (appelées

⁴¹³ Déjà mentionné lors de la description de VerbNet (section 3.3.2. *VerbNet*, dans ce chapitre).

⁴¹⁴ Disponible sur ce lien : http://clic.ub.edu/corpus/es/ancoranom_es.

⁴¹⁵ Disponible sur ce lien : <https://catalog ldc.upenn.edu/LDC2006T08>.

⁴¹⁶ En TAL, un *token* est une suite de caractères qui constitue une unité linguistique minimale.

⁴¹⁷ La notion d'évènement dans TimeBank correspond à la notion élargie d'*éventualité* de Bach (1986) et elle comprend tous les types de situations (états, activités, achèvements, etc.). Selon Pustejovsky, Hanks et al. (2003), « events are taken to be situations that *occur* or *happen*. [...] Events are generally expressed

marquables), et, le plus important, l'annotation des relations temporelles existantes entre les événements ou entre un événement et une expression temporelle. Un langage de balisage a été développé à cet effet : ISO-TimeML⁴¹⁸ (Pustejovsky, Castaño et al., 2003). Ce langage compte six balises : deux pour les *marquables* (<EVENT> et <TIMEX3>), trois pour les relations (<TLINK>, <SINK> et <ALINK>), et une pour les marqueurs de relations (<SIGNAL>). Nous les présentons ci-dessous de manière succincte (cf. Bittar, Asmili et Denis, 2011) et nous présentons un exemple d'utilisation (Figure 21) :

- a. <EVENT> pour les événements (*marcher, être en retard*). Cette balise comporte des données sur les traits morphosyntaxiques et sémantiques de l'évènement annoté (classe sémantique, temps, aspect, mode, etc.).
- b. <TIMEX3> pour les expressions temporelles (*le 15 décembre 2013, la semaine dernière, mardi*). Cette balise comporte des attributs pour le type d'expression temporelle (date, heure, durée ou fréquence) et sa valeur normalisée⁴¹⁹.
- c. <SIGNAL> pour les marqueurs de relation temporelle entre différents événements (*après, pendant, lors de*).

by the means of tensed or untensed verbs, nominalisations, adjectives, predicative clauses, or prepositional phrases. Predicates describing states or circumstances in which something obtains or holds true are also considered as events » (p. 648). Par exemple, *mourir, lire, vol* ou *conversation* sont des événements (*events*). On notera que le terme *événement* ne désigne pas la même réalité linguistique dans TimeBank et dans la méthode des classes d'objets. Nous rappelons que dans le cadre de la méthode des classes d'objets, l'hyperclasse <événement> est définie par les verbes supports *avoir lieu, se produire* ou *arriver*, lorsqu'il s'agit d'un prédicat nominal (*rentrée scolaire, agression, championnat*). Quant aux constructions événementielles verbales, il s'agit notamment de phrases impersonnelles : verbes météorologiques (*Il pleut*), des constructions impersonnelles (*Il se vend beaucoup de journaux le dimanche*) ou des phrases clivées (*Il y a eu une réponse de sa part*) (G. Gross, 2012).

⁴¹⁸ TimeML étant la contraction de Time Markup Language.

⁴¹⁹ Cf. ISO (2008).

- d. <TLINK> pour les relations temporelles entre deux évènements ou entre un évènement et un marqueur temporel. Il existe différents sous-types de balise, introduits au moyen de l'attribut *relType*: BEFORE, AFTER, SIMULTANEOUS, BEGINS⁴²⁰, etc.
- e. <SLINK> pour les relations de subordination (modale) entre deux évènements. Généralement, cette balise marque la relation qui existe entre un verbe modal (*devoir, falloir*) ou de perception (*entendre, voir*) et son complément évènementiel. Cette balise comporte aussi des sous-types, toujours introduits au moyen de l'attribut *relType*, : MODAL, FACTIVE, NEGATIVE, etc.
- f. <ALINK> pour les relations aspectuelles entre deux évènements (un évènement aspectuel et son complément évènementiel (*commencer, continuer, cesser*). Des sous-types de balise spécifient le type de relation aspectuelle : INITIATION, TERMINATION, CONTINUATION, etc.

```

There was <SIGNAL sid="8"> no </SIGNAL> <EVENT EID="57" CLASS="STATE"
TENSE="PAST" ASPECT="NONE"> hint of trouble </EVENT> <SIGNAL id="11">
in </SIGNAL> the last <EVENT class="OCCURRENCE" aspect="NONE"
eid="10" tense="NONE"> conversation </EVENT> between controllers and
TWA pilot Steven Snyder. But <TIMEX3 TID="58" val="PT1M30S"
type="DURATION" temporalFunction="false"> a minute and a half
</TIMEX3> <SIGNAL SID="59"> later </SIGNAL>, a pilot from a nearby
flight <EVENT aspect="NONE" eid="18" tense="PRESENT"
CLASS="REPORTING"> calls </EVENT> in.

<SLINK eventInstanceID="eiid57" signalID="8" relType="NEGATIVE"/>
<TLINK eventInstanceID="eiid57" relatedToEvent="eiid10"
signalID="s11" relType="IS_INCLUDED"/>
<TLINK eventInstanceID="eiid18" relatedToEvent="eiid10"
signalID="s59" relType="AFTER" magnitude="t58"/>

```

Figure 21 : Exemple d'annotation ISO-TimeML (extrait de Pustejovsky et al., 2006, p. 651)

TimeBank est le corpus d'étiquetage temporel le plus fin et le plus vaste. Son système d'étiquetage poursuit quatre objectifs : la datation des évènements, l'articulation chronologique des évènements entre eux, le raisonnement sur les expressions temporelles sous-spécifiées et le raisonnement sur la durée des évènements.

⁴²⁰ Observons que dans ce cas, les exemples donnés ne sont plus de occurrences susceptibles d'apparaître de manière explicite dans le corpus, mais le codage qui figure dans l'analyse fait par ISO-TimeML.

L'inconvénient que nous y trouvons, c'est que les annotations sont de surface⁴²¹, c'est-à-dire que l'étiquetage se fait seulement à partir des marqueurs linguistiques explicites. Ce système n'est donc pas capable d'inférer ce qui est implicite. Toutefois, l'intérêt de cette ressource pour le TAL est capital.

Cette ressource a été développée pour le français et pour l'espagnol, ainsi que pour d'autres langues comme le chinois, le coréen et l'italien. Les équipes de recherche de l'Université Paris Diderot ont élaboré French TimeBank (Bittar, 2010 ; Bittar, Amsili, Denis et Danlos, 2011), un corpus annoté de plus de 16 200 tokens (signes de ponctuation exclus) provenant du journal *Est Républicain*. Quant à l'espagnol, une équipe de chercheurs de Barcelona Media a développé Spanish TimeBank 1.0⁴²² (Sauri et Badia, 2012). Ce corpus a été constitué à partir d'articles journalistiques, de nouvelles et de romans issus du corpus AnCora-ES⁴²³. Spanish TimeBank contient 68 000 tokens (signes de ponctuation exclus).

Autant pour le français que pour l'espagnol, il a été nécessaire d'adapter la norme ISO-TimeML aux spécificités de chaque langue. Les nécessités d'adaptation de la norme pour ces deux langues sont en partie communes. D'une part, le système d'expression du temps et de l'aspect en français et en l'espagnol diffère de celui de l'anglais. Ainsi, il a fallu proposer une série de valeurs appropriées pour l'annotation des temps verbaux (Bittar, Amsili et Denis, 2011 ; Sauri et Badia, 2012). D'autre part, le mode subjonctif tel quel n'existant pas en anglais, mais étant exprimé par d'autres moyens linguistiques, il a fallu faire en une adaptation pour l'espagnol et pour le français. Alors, la valeur sunjonctive, qui n'était pas prévue dans ISO-TimeML a été

⁴²¹ Les créateurs de cette ressource ont toujours souligné le caractère résolument surfacique de la norme ISO-TimeML car l'objectif n'est pas d'annoter les sens, mais de proposer une normalisation des formes linguistiques qui expriment la temporalité (Bittar, Amsili et Denis, 2011).

⁴²² Disponible sur ce lien : <https://catalog.ldc.upenn.edu/docs/LDC2012T12/>.

⁴²³ AnCora-ES est un corpus de l'espagnol contenant 500 000 mots. L'annotation morphologique de ce corpus a donné lieu aussi aux ressources lexicales AnCora-Verb (cf. section 3.3.2. *VerbNet*) et AnCora-Nom. En effet, le corpus AnCora a été étiqueté à plusieurs niveaux linguistiques (syntaxique, sémantique, pragmatique, etc.), mais ces données ne sont pas présentes dans la version utilisée pour Spanish TimeBank.

créée. Les verbes modaux représentent une autre différence de ces deux langues par rapport à l'anglais. Les verbes modaux espagnols (*deber, tener que*) et en français (*se pouvoir, falloir*) n'ont pas le même comportement que les auxiliaires modaux de l'anglais. Dans le cas de l'espagnol et du français, il s'agit plutôt de verbes lexicaux. C'est pourquoi il a été proposé de les baliser au moyen de <EVENT>, contrairement aux modaux anglais (Bittar, Amsili et Denis, 2011).

```
Jean est <EVENT id="e1" class="OCCURRENCE" pos="VERB" tense="PAST" vForm="PASTPART">né
</EVENT> <SIGNAL id="s1">avant</SIGNAL> l'<EVENT id="e2" class="OCCURRENCE"
pos="NOUN">introduction</EVENT> de l'euro.
<TLINK id="l1" eventID="e1" relatedToEvent="e2" signalID="s1" relType="BEFORE"/>
```

Figure 22 : Exemple n'annotation ISO-TimeML de French TimeBank (extrait de Bittar, Amsili et Denis, 2011, p. 260)

3.4.2. Nomage

Le projet Nomage (cf. Balvet et al., 2010 ; Balvet et al., 2011a, 2011b), mené au laboratoire STL⁴²⁴, a comme objectif la description des propriétés aspectuelles des noms déverbaux du français dans le but d'analyser dans quelle mesure les noms déverbaux héritent des propriétés aspectuelles de leurs verbes source. Ce projet a donné lieu à deux ressources : un corpus annoté et un lexique électronique de noms déverbaux. D'une part, le corpus est composé de l'ensemble de phrases, issues du French Treebank⁴²⁵, comportant au moins un nom déverbal. Ces noms y sont annotés sémantiquement et syntaxiquement. D'autre part, les occurrences de noms déverbaux retenues dans le corpus French Treebank ont servi à créer un lexique électronique des noms déverbaux du français. Ce lexique fournit des informations morphologiques (le verbe source, le suffixe dérivationnel, etc.), syntaxiques (la structure argumentale du nom prédicatif) et sémantiques (notamment la classe aspectuelle) sur les noms déverbaux du corpus et sur les verbes dont ils sont dérivés.

⁴²⁴ Savoir, Textes, Langages (UMR 8163, CNRS – Université Lille 3).

⁴²⁵ Le French Treebank est un corpus (un million de tokens) d'articles du journal *Le Monde*. Il fournit des annotations linguistiques de différents niveaux : tokenisation simple et composée, lemmes, balisage des parties du discours accompagnés d'informations morphologiques, entre autres (Abeillé et al., 2003 ; Abeillé et Barrier, 2004 ; Abeillé et al., 2019).

Selon Balvet et al. (2011a), la principale originalité de cette ressource est la double méthode d'annotation aspectuelle : l'une fondée sur la réflexion théorique, et l'autre, plus empirique. Dans l'approche théorique, des annotateurs formés en sémantique ont classifié les noms déverbaux et les verbes source (repérés dans French Treebank) en appliquant un certain nombre de tests aspectuels issus de la littérature sur le domaine⁴²⁶ à des exemples construits. Quant à l'approche empirique, des annotateurs non linguistes ont appliqué des tests transformationnels (partiellement différents à ceux appliqués sur les exemples construits) à des phrases du corpus pour vérifier la validité des transformations obtenues. La description est validée lorsque l'annotation faite sur le corpus (méthode empirique) concorde avec l'annotation faite côté lexique (méthode théorique). L'atout de cette double méthode d'attribution d'une classe aspectuelle repose sur le fait que ces deux approches s'alimentent mutuellement fournissant ainsi des données plus fiables :

la réflexion théorique a contribué à l'annotation en corpus en fournissant des tests linguistiques et une ontologie, et en retour, l'approche empirique a permis de remettre en question et d'améliorer l'analyse théorique, s'agissant notamment de la pertinence des classes aspectuelles retenues (Balvet et al., 2011a, p. 148).

Les tableaux ci-dessous illustrent les données fournies dans le lexique électronique de Nomage. Nous montrons la description de l'unité lexicale AMÉNAGEMENT#1 (Tableau 8), de son verbe source AMÉNAGER#1 (Tableau 9) et une occurrence dans le corpus issu de French Treebank (Tableau 10).

id	45
Lexeme	AMÉNAGEMENT#1
Argument structure	~ de Y par X
Aspectual class	ACC
Occurrences in the FT	{id:1794 ; id:1929}
Verbal base	id:44

Tableau 8 : Description du nom aménagement#1 dans Nomage (extrait de Balvet et al., 2011b, p. 10)

⁴²⁶ Dans Balvet et al. (2011a, 2011b) on cite des références telles que Flaux et Van de Velde, 2000 ; Haas et al. ; 2008 ; Haas et Huyghe, 2010 ; Heyd et Knittel, 2009 ; Vendler 1967.

Le nom déverbal a deux arguments (X et Y⁴²⁷) et il dénote un accomplissement, c'est-à-dire une situation dynamique, durative et culminante. Chaque description des noms déverbaux pointe vers la description verbe dont il est le dérivé (Tableau 9) et vers ses occurrences dans le corpus (Tableau 10) :

id	44
Lexeme	AMÉNAGER#1
Argument structure	X ~ Y
Aspectual class	ACC

Tableau 9 : Description du verbe *aménager#1* dans *Nomage* (extrait de Balvet et al., 2011b, p. 10)

id	1794
Deverbal	id:45
Occ.	Tout ce travail préparatoire sera fondamental pour l' aménagement universitaire au cours des cinq prochaines années.
Réal. Arg.	X:∅, Y:adj. rel.

Tableau 10 : Une occurrence de *aménagement#1* dans *Nomage* (extrait de Balvet et al., 2011b, p. 10)

Le travail de description a permis d'attester qu'il existe, dans la grande majorité des cas, un héritage aspectuel entre verbes et noms déverbaux⁴²⁸. Ainsi, la plupart de noms partagent la même classe aspectuelle que celle du verbe dont ils dérivent : états, activités, accomplissements, achèvements⁴²⁹, achèvements-statifs, accomplissements-statifs et activité accomplissement. Cependant d'autres classes sont spécifiques aux noms. L'ajout de plusieurs classes aspectuelles spécifiques aux noms donne lieu à une description plus fine de ceux-ci (Balvet et al., 2010).

Nous allons mettre l'accent sur ces spécificités de l'aspect nominal, étant donnée la relation directe avec l'objet de notre étude. D'abord, l'opposition massif/comptable

⁴²⁷ Dans une version ultérieure de *Nomage*, les arguments recevront une caractérisation sémantique.

⁴²⁸ Les données du projet révèlent que dans 67 % des cas l'aspect de la forme nominale coïncide exactement avec l'aspect des verbes source, dans 9 % les verbes et les noms déverbaux appartiennent à des classes aspectuelles légèrement différentes, et dans 24 % ils appartiennent à des classes complètement différentes.

⁴²⁹ On notera que ces quatre premières classes aspectuelles correspondent aux classes vendleriennes (cf. Vendler, 1967).

permet de distinguer, au niveau aspectuel, deux types de noms dérivés de verbes d'activité : les noms d'activité, comptables (p. ex., MANIFESTATION⁴³⁰, PROMENADE) et les noms d'habitude, massifs (p. ex., JARDINAGE). Tous ces noms dénotent des situations dynamiques, durables et non culminantes, mais seuls les noms comptables dénotent des actions délimitées dans le temps, c'est-à-dire des événements. Nous montrons ci-dessous (Figure 23) la hiérarchie des classes aspectuelles simples des noms déverbaux qui a résulté des analyses du projet Nomage :

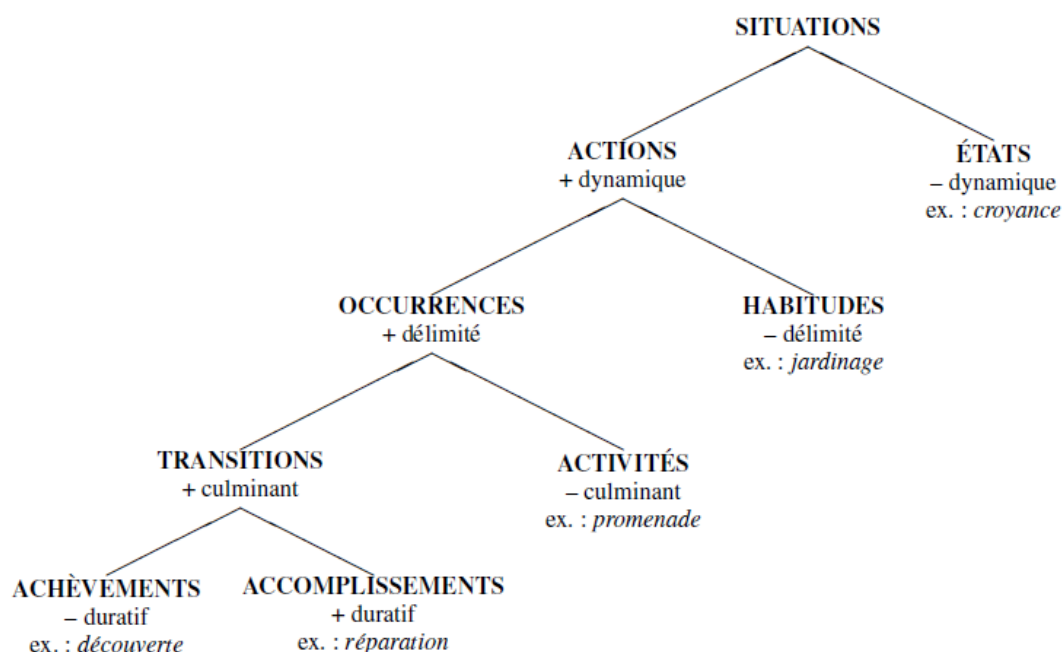


Figure 23 : Hiérarchie des classes aspectuelles simples des noms (extrait de Balvet et al., 2011a, p. 138)

Une deuxième spécificité des noms déverbaux tient au fait que, contrairement aux verbes, ceux-ci peuvent dénoter des objets. Les objets, rappelons-le, sont des entités dépourvues d'aspect. Ainsi, on distingue une classe des objets qui comprend : des noms dénotant des objets matériels (p. ex., CONSTRUCTION), des noms dénotant des objets ayant un contenu informationnel (p. ex., AFFIRMATION) et des noms dénotant quelque chose qui provoque un état psychologique (p. ex., OBSESSION peut dénoter quelque chose qui est l'objet d'une obsession).

Enfin, dans Nomage on a eu recours aussi à des classes aspectuelles complexes. Celles-ci comprennent des noms susceptibles de dénoter une situation

⁴³⁰ Nous utilisons les petites capitales afin de respecter les codes typographiques utilisés dans Nomage. Les lexèmes en majuscules indiquent qu'il s'agit d'un vocable.

(accomplissement, achèvement), un objet ou les deux en même temps (cf. Godard et Jayez, 1996 ; Milićević et Polguère, 2010 ; Pustejovsky, 1995). À titre d'exemple, le nom *exposé* dans la phrase *Son exposé fut long et ennuyeux*⁴³¹ reçoit la double étiquette ACC•OBJET (accomplissement et objet).

La ressource Nomage combine donc un corpus de noms déverbaux annoté sémantiquement et syntaxiquement et un lexique électronique fournissant des descriptions des propriétés morphologiques, syntaxiques et sémantiques des noms déverbaux et des verbes source issus de ce corpus. Il s'agit donc d'une ressource intéressante, mais très limitée car le lexique est composé de 746 noms déverbaux et 679 verbes.

3.5. Informatisation des dictionnaires papier

Les dictionnaires papier sont des ressources qui contiennent de nombreuses informations sémantiques exploitables automatiquement. Les plus importants de ces dictionnaires papier ont subi un traitement de rétroconversion en vue de le rendre disponible sur support informatique. En langue française, le grand exemple de rétroconversion pour rendre disponible un dictionnaire papier en support électronique est le *Trésor de la Langue Française informatisé* (TLFi⁴³²) (cf. Dendien, 1996 ; Dendien et Pierrel, 2003 ; Pierrel, 2006). En espagnol, le travail lexicographique de rétroconversion le plus important est la version numérique du *Diccionario de la Lengua Española* (<https://dle.rae.es/>)⁴³³. Les descriptions lexicographiques de ces dictionnaires

⁴³¹ Exemple emprunté à Balvet et al. 2011a (p. 11).

⁴³² *Trésor de la Langue Française informatisé* (TLFi), Nancy, CNRS, ATILF (Analyse et traitement informatique de la langue française), UMR CNRS-Université Nancy 2 [<http://atilf.atilf.fr/frantext.htm>]. Ce dictionnaire informatisé a été élaboré par une centaine de chercheurs (CNRS, INaLF, ATILF) entre 1960 et 1990. Il contient 100 000 mots et 270 000 définitions. La rétroconversion a été faite moyennant un balisage XML des différents types d'informations (domaine, définition, exemples, etc.).

⁴³³ Dans le cadre des ressources linguistiques académiques en espagnol, nous tenons à signaler l'apparition en 2017 de la plateforme Enclave RAE (<https://enclave.rae.es/>). Elle contient tous les dictionnaires généraux de la RAE, des dictionnaires spécialisés (*Diccionario jurídico*, *Diccionario panhispánico de dudas*, *Diccionario de americanismos* y *Diccionario del estudiante*), une grammaire (*Nueva gramática de la lengua española*), le corpus CORPES XXI (400 millions de formes), des ressources pédagogiques pour les professionnels de l'enseignement, une section de dépannage

contiennent les mêmes informations que le dictionnaire papier. Leur atout (à part le côté pratique et rapide de la recherche d'un mot), c'est qu'ils offrent de nouveaux modes de consultation lexicographique grâce aux liens hypertextuels automatiques et la possibilité de faire des requêtes complexes qui exploitent l'ensemble de la structure du dictionnaire.

Cependant, le traitement de l'information dans les dictionnaires papier n'est ni entièrement explicite ni absolument cohérent. Pour Robert Martin (1994), la rétroconversion des dictionnaires papier demande un aménagement des données sous trois aspects : un effort de systématisation (le même type d'information ne doit plus apparaître sous diverses formes), un effort d'explicitation (p. ex., que l'information syntaxique ne soit plus cachée) et un effort de complémentation (pallier les lacunes présentes dans les dictionnaires papier). En effet, les dictionnaires destinés à l'usage humain contiennent des lacunes et des incohérences dans la description de la morphologie ou la classification morphosyntaxique (cf. Dubois et Dubois-Charlier, 1990 ; Véronis, 2004). Dans ces dictionnaires de nombreuses informations restent implicites parce qu'elles sont évidentes pour l'utilisateur humain. Par contre, le programme de traitement automatique doit avoir accès à toutes les formes fléchies d'une unité lexicale pour être capable de les lemmatiser, ou il doit également disposer de listes des unités polylexicales. Alors, même si ces dictionnaires sont accessibles à des programmes informatiques, ils ne sont pas vraiment performants pour le traitement automatique des langues à cause de l'hétérogénéité de traitement des différentes données et des informations qui restent implicites⁴³⁴.

D'un point de vue sémantique, le traitement automatique de l'homonymie et de la polysémie demande une description beaucoup plus exhaustive et systématique que celle des dictionnaires papier. Également, il faut expliciter à l'automate de manière systématique toutes les relations lexicales, ce qui n'est pas toujours le cas dans les

linguistique, un correcteur et un étiqueteur morphologique. On comprend bien que l'ampleur de cette plateforme dépasse largement la rétroconversion d'un dictionnaire papier.

⁴³⁴ Néanmoins, un travail remarquable a été fait pour structurer et formaliser les définitions du TLFi dans le cadre du projet Définiens (L. Barque et A. Polguère, 2008). Nous avons présenté ce projet dans la section 3.2.4. *Définiens* de ce chapitre.

dictionnaires papier, et lui fournir des données nécessaires pour que les inférences lexicales ne produisent pas d'analyses erronées (les dictionnaires papier n'offrent jamais d'informations à propos des inférences). Ces problèmes ne nuisent en rien la consultation humaine, mais posent des obstacles insurmontables aux systèmes de traitement automatique des langues naturelles.

Ainsi, la véritable distinction n'est plus le support papier ou le support informatique car actuellement presque la totalité des dictionnaires papier ont aussi une version informatisée. La véritable distinction est la finalité des dictionnaires : une ressource pour l'humain ou pour la machine. Nous avons vu que parmi ces dictionnaires (ou ressources lexicales) destinés au traitement automatique, il y en a qui disposent d'interfaces qui rendent leurs données lisibles et compréhensibles à l'être humain.

4. Le traitement automatique et la méthode des classes d'objets

Le développement d'outils de traitement automatique des langues naturelles doit se fonder sur un modèle de description de la langue structuré selon les principes d'une théorie linguistique homogène et adaptée (ou adaptable) aux spécificités de l'automate. Nous adhérons au postulat de Polguère (1998b) selon lequel : « un modèle général de la lexicalisation en génération de texte doit être enchâssé dans une approche théorique linguistique donnée » (p. 61). En effet, il s'avère presque impossible de développer un système de TAL sans une approche linguistique particulière sous-jacente, parce que le fonctionnement du système hérite de la structuration des connaissances (dans un cadre théorique donné) qu'il manipule. Donc, à part le développement de bons algorithmes, le travail de formalisation détaillée du linguiste est fondamental.

Dans le cadre du modèle des classes d'objets, on a mené un projet de développement d'un système de TAL. Ce système, conçu comme une plateforme d'analyse syntacticosémantique de textes, nécessite de ressources linguistiques (une description exhaustive et systématique de la langue) et de ressources informatiques (des outils informatiques supportant cette description de la langue et des logiciels permettant l'exploitation informatique de ces ressources). Le système développé fonctionne en trois grandes phases : la segmentation de la chaîne graphique, l'analyse morphologique et l'analyse syntacticosémantique. Ces phases d'analyse linguistique sont nécessairement

précédées d'une phase de prétraitement : la normalisation des documents. À chaque étape, il y a un produit de sortie qui permet d'effectuer une évaluation des résultats et qui sera utilisé dans l'étape suivante. L'architecture du système est donc la suivante :

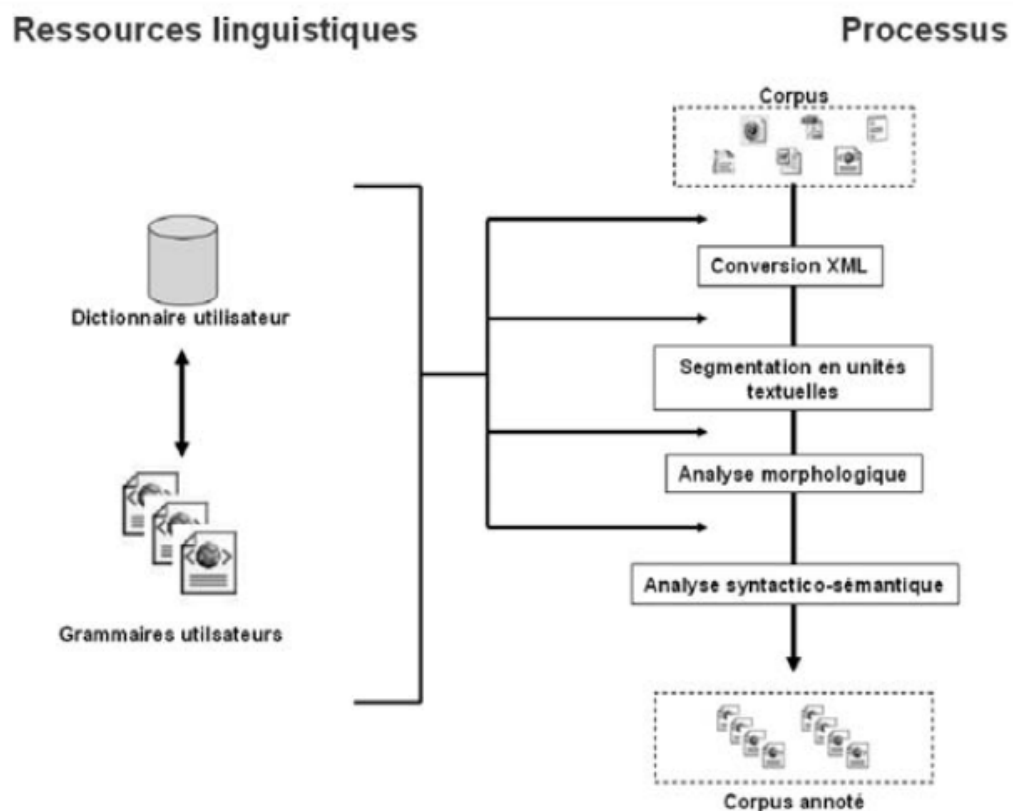


Figure 24 : *Architecture du système TextBox* (extrait de Buvet et al., 2007, p. 241)

La figure ci-dessus est divisée en deux zones : celle de gauche comprend les ressources linguistiques (le dictionnaire et les grammaires) ; celle de droite comprend les processus d'analyse linguistique, à partir des documents qui composent le corpus (en haut, sur l'image) jusqu'à la génération d'un corpus annoté en format XML. On notera que tout le système est basé sur un fonctionnement en cascade : chaque étape d'analyse travaille sur le document généré dans l'étape précédente. L'analyse des textes en traitement automatique implique que ceux-ci soient enrichis de données métalinguistiques.

À continuation, nous présenterons les ressources linguistiques dont ce système a besoin pour ensuite expliquer les différentes phases du fonctionnement du système de TAL en cours de développement dans le cadre du modèle des classes d'objets.

4.1. Ressources linguistiques pour le système de TAL du modèle des classes d'objets

Il y a trois sortes de ressources qui contribuent à cet enrichissement : les dictionnaires électroniques, les grammaires locales et les corpus. Seulement les deux premières y interviennent directement, car les corpus sont utilisés pour l'élaboration et la validation des dictionnaires électroniques et des grammaires locales. Nous les présenterons à continuation.

4.1.1. Dictionnaires électroniques

Nous avons déjà signalé que l'exploitation de dictionnaires traditionnels (destinés à l'usage humain) en TAL n'est pas effective. Buvet et Issac (2006) signalent trois grandes causes. D'une part, ces dictionnaires ne sont pas complets car seuls les sens les plus fréquents sont décrits. D'autre part, les descriptions linguistiques de ces dictionnaires sont insuffisantes pour le traitement automatique : il manque des informations relatives à la construction dans laquelle le mot est réalisé dans une phrase et, souvent, il n'y a aucune information sur l'aspect inhérent ou contextuel du mot. Alors, il s'avère nécessaire pour les systèmes de TAL d'élaborer des dictionnaires électroniques du type base de données où chaque entrée lexicale contient des descripteurs exhaustifs et normalisés (cf. Courtois et Silberztein, 1990). Nous tenons à préciser la notion de dictionnaire telle que nous l'entendons :

Notre conception des dictionnaires électroniques s'inspire largement de la pratique des bases de données : chaque entrée constitue un « enregistrement », cependant que la description se trouve répartie en une série de « champs » (rubriques) clairement définis, correspondant aux différents paramètres de l'information lexicographique (Mathieu-Colas, 1996, p. 318).

Les dictionnaires élaborés dans le cadre du modèle des classes d'objets sont donc des structures logiques qui associent des informations morphologiques et syntacticosémantiques à une forme lexicale. Ces dictionnaires servent à étiqueter les unités lexicales repérées dans les textes (les tokens). Ils ont les particularités suivantes : les unités lexicales sont rapportées à la distinction entre prédicats du premier ordre, arguments élémentaires et connecteurs ; les descripteurs de chaque entrée ne sont pas hiérarchisés ; ces descripteurs relèvent du modèle des classes d'objets. Cette notion de dictionnaire est donc très différente de ce que l'on comprend généralement comme un dictionnaire, qu'il soit un dictionnaire papier ou au format électronique. La finalité du

mode d'élaboration de ces dictionnaires est de décrire de manière exhaustive le lexique d'une langue, avec des propriétés morphologiques et syntacticosémantiques explicites et reproductibles.

4.1.1.1. Dictionnaires morphosyntaxiques

Les formes (fléchies) d'un dictionnaire sont regroupées en une forme canonique appelée *lemme*. Ainsi, l'opération de lemmatisation consiste à assigner à une forme fléchie son lemme correspondant ; par exemple, *accorderont*, *accorde*, *accordait* seront associés à l'infinitif *accorder*, considéré comme le prototype.

L'identification d'un mot à partir d'un token peut se faire par analyse ou par reconnaissance (Cartier et Issac, 2009). Dans l'analyse, à partir d'une forme fléchie et des règles de flexion, le système peut déduire le lemme associé. Par exemple, à partir de la règle « au pluriel on ajoute un *s* aux noms » et du lemme *allocation*, présent dans le dictionnaire, le système peut déduire que le token *allocations* est la forme fléchie au pluriel du lemme *allocation*. D'autre part, le procédé de reconnaissance utilise un dictionnaire contenant toutes les formes fléchies. Ce dictionnaire est donc composé d'une liste d'éléments : <forme><description> (<forme> étant le mot fléchi, et <description> étant les informations morphologiques associées à ce mot). En français, l'ensemble de formes fléchies est d'environ 700 000, un chiffre que l'ordinateur peut facilement gérer. Un exemple de ce type de ressource est le dictionnaire morphologique Morfetik (Mathieu-Collas, 1996-2006).

Morfetik est un dictionnaire électronique conçu comme un système modulaire. Le dictionnaire morphologique proprement dit est lié à des outils informatiques tels qu'un moteur de flexion (Morfetik Engine), d'interfaces de consultation et d'interrogation des lemmes et des formes fléchies et un ensemble d'outils assurant la maintenance et l'exploitation de ces ressources (cf. Buvet et al., 2009 ; Mathieu-Colas, 2009). Morfetik vise la génération automatique de l'ensemble des formes simples du français. Le schéma ci-dessous (Figure 19) illustre l'architecture générale de ce système :

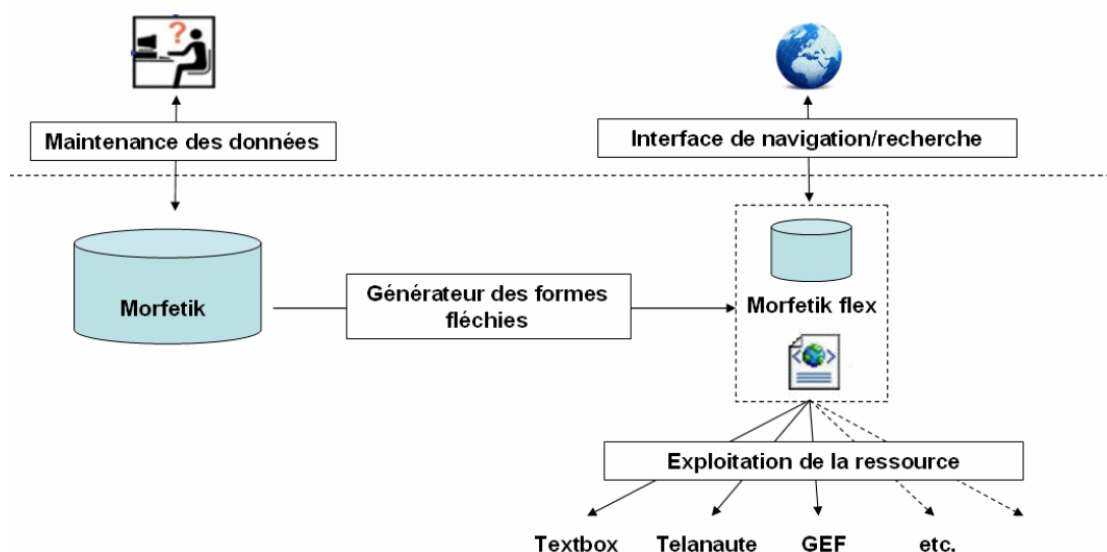


Figure 25 : Architecture du système Morfetik (extrait de Buvet et al., 2009, p. 220)

Le dictionnaire Morfetik se compose de deux volets : MORF-SIM (décrivant les unités monolexicales) et MORF-COM (décrivant les unités polylexicales). Ils contiennent un ensemble de formes fléchies du français et les informations morphologiques associées à chaque forme ; par exemple des informations de catégorie, de genre et de nombre pour les noms et les adjectifs, ou des informations de temps, de mode, de personne et de genre pour les verbes. Ces informations morphologiques sont structurées en tables différentes, en fonction de la catégorie morphosyntaxique dont il s'agit. Cinq groupes différents ont été mis en place. Pour décrire certains groupes, comme celui des mots invariables (prépositions, adverbes, conjonctions et interjections) et celui des déterminants et pronoms, il a suffi de faire un simple listage. Par contre, pour les autres catégories (adjectifs, noms et verbes), il a été nécessaire de créer deux tables complémentaires : d'une part, des tables de flexion pour identifier et coder les types flexionnels ; et d'autre part, des tables de lemmes pour attribuer à chaque lemme le code flexionnel qui lui correspond (Mathieu-Colas, 2009).

En ce qui concerne les noms, 63 codes ont été définis, couvrant toutes les formes attestées⁴³⁵. Dans le cas des flexions multiples, les codes ont été attribués sans faire

⁴³⁵ Le recensement lexical a été fait à partir de nombreuses ressources lexicographiques telles que le DELAS (Dictionnaire électronique du LADL, cf. Courtois, 1990), le *Petit* et le *Grand Robert*, le *Petit Larousse illustré*, le *Lexis*, le *Grand Larousse encyclopédique* et le *Grand Dictionnaire encyclopédique Larousse*, le *Trésor de la langue française*, le *Harrap's*, le *Robert & Collins*, des dictionnaires d'argot, des

distinction ni de la polysémie (*ciels / cieux*), ni des variantes (*barmans / barmen*). Seuls les mots homonymes ont été décrits sur des lignes différentes de la table (*sols* : ‘surface de la Terre’ / *sol* : ‘note de musique’). Quant aux adjectifs, les mêmes principes ont été appliqués, en ajoutant la flexion en genre. Ainsi, 59 codes ont été établis. Les déterminants et les pronoms reçoivent un code pour chaque forme, du fait de l’irrégularité de leurs flexions. Les mots invariables (prépositions, adverbes et interjections) sont les plus faciles à décrire. Ils n’ont pas reçu de codage, mais ils ont tous été recensés sur une liste, avec l’indication de leurs variables (*jusque / jusqu’*). Le codage le plus complexe est celui des verbes. Chaque code associé à un infinitif correspond à un type de conjugaison spécifique, défini de manière précise et exhaustive. Ainsi, toute variation dans le paradigme entraîne la création d’un nouveau code. Au total, 226 codes ont été définis. Les tables descriptives des verbes comprennent plus de champs que celles des adjectifs et des noms. Observons la fiche du verbe *acquérir* :

CHAMPS	EXEMPLE
numéro de code	036
exemple-modèle	ACQUÉRIR
radical (nombre de caractères à soustraire de la forme canonique)	-4
radical-modèle	ACQU
désinences de l’infinitif	<i>érir</i>
désinences des formes conjuguées (45 champs)	<i>iers, iers, iert, érons...</i>
désinences des participes (5 champs)	<i>erant, is, ise, is, ises</i>

Tableau 11 : *Structure de la table de flexion du verbe acquérir* (extrait de Mathieu-Colas, 2009, p. 143)

Force est de constater que, le découpage des radicaux et des désinences appliqué dans le domaine du traitement informatique ne correspond pas à l’analyse que l’on fait en morphologie. Pour des raisons d’efficacité en traitement automatique, le radical a été défini comme le plus petit dénominateur commun ; par exemple, *v-* pour le verbe *vouloir*, à cause de *veux* et *veut*.

Ces tables seront ensuite utilisées par le moteur de flexion Morfetik Engine pour produire le lexique des formes fléchies (Morfetik Flex). La description exhaustive et

tables de conjugaison (le *Bescherelle* et *Les verbes logiques* de Dugas et Manseau), *Le Bon Usage*, entre autres (cf. Buvet et al., 2009, p. 118 ; Mathieu-Colas, 2009, p. 138).

formelle de la morphologie des unités lexicales permet de générer l'ensemble des formes fléchies du français. Les tâches de ce moteur de flexion sont les suivantes :

[il] se charge de la collecte des informations dans les tables, de la création des formes fléchies à partir du lemme et du code de flexion approprié, de l'attribution des catégories grammaticales correspondantes (genre, nombre, etc.) et des autres métadonnées (par exemple) rareté, et enfin il centralise l'exécution des modules. (Buvet et al., 2009, p. 220)

Un ensemble d'interfaces de navigation et d'interrogation de Morfetik Flex ont été développés, ainsi que plusieurs outils de maintenance et d'exploitation de cette ressource linguistique. Ainsi, comme nous pouvons le voir sur le schéma de la Figure 25, diverses applications ont été élaborées, telles que TextBox (un outil d'analyse linguistique des textes), Telanaute (un outil de repérage des néologismes d'une langue), GEF (un générateur automatique d'exercices de français dans une plateforme d'e-learning), etc. (cf. Buvet et al. 2009).

Morfetik constitue une ressource évolutive qui s'enrichit de manière continue. Il s'agit du dictionnaire morphologique le plus extensif du français⁴³⁶ (Grezka, Cartier et Mathieu-Colas, 2015).

4.1.1.2. Dictionnaires syntacticosémantiques

Les dictionnaires syntacticosémantiques de l'ancien laboratoire LDI visent la description des emplois des mots. Les données et la configuration de ces dictionnaires sont également basées sur la réflexion théorique de la langue du modèle des classes d'objets. Comme Blanco (1999) le signale,

Les classes d'objets sont intrinsèquement destinées au traitement automatique des langues naturelles. L'élaboration lexicographique d'une classe implique, donc, la mise au point d'un

⁴³⁶ En dehors du laboratoire LDI, d'autres ressources visant la morphologie ont été développées sur de basses théoriques différentes et servant à des finalités différentes. Ainsi, nous trouvons : des ressources de formes fléchies telles que Morphalou (Romary et al., 2004), Lexique 3 (New et Pallier, 2005) et DELAF (Courtois, 2004) ; des ressources et outils réalisant la flexion comme FLEMM (Namer, 2000) ; et des environnements complets, tels que INTEX (Silberztein, 1993a), Unitex (Paumier, 2016) et NooJ (Silberztein, 2005), entre autres. Pour une comparaison de Morfetik avec les ressources linguistiques analogues Morphalou, Lexique 3 et DELAF, voir Buvet et al. (2009).

module de dictionnaire électronique constitué par des entrées présentant des champs standardisés. (p. 7)

Il existe quatre types de dictionnaires, un pour chaque fonction primaire décrite, les prédicats (PRED_DIC), les arguments (ARGU_DIC) et les actualisateurs (ACTU_DIC), et un autre dictionnaire pour la description des noms d'êtres humains (ETHU_DIC). Les informations contenues dans ces dictionnaires sont stockées sous forme de table dans une base de données. Les données sont structurées sous des descripteurs formels, c'est-à-dire les informations sont normalisées et totalement explicites.

La nomenclature des dictionnaires des arguments est constituée de noms élémentaires (simples ou composés). Ils sont classés en traits (concret, locatif, humain, etc.) et en classes d'objets en fonction des prédicats qui les sélectionnent. Ils sont décrits à partir de leur catégorie grammaticale, leur trait syntaxique, leur classe d'objets, leur hyper-classe et le domaine auquel ils appartiennent. La description de la classe d'objets donne lieu à la création de six fichiers différents. Les deux premiers fichiers contiennent la liste des objets de la classe (un fichier pour les unités simples, et un autre pour les unités polylexicales). Ensuite, il faut recenser les opérateurs appropriés, en séparant les verbes, les noms et les adjectifs en trois fichiers différents. Enfin, le dernier fichier comporte des unités polilexicales qui contiennent des éléments de la classe décrite. La finalité de ce dernier fichier est d'aider à la désambiguïsation pour que l'ordinateur puisse interpréter une suite de mots de manière compositionnelle ou non compositionnelle, en fonction des cas.

En ce qui concerne les dictionnaires des prédicats, ils décrivent des noms, des verbes et des adjectifs. Dans ce cas, la microstructure est plus complexe que celle des dictionnaires des arguments, car elle comporte deux types de descripteurs : des descripteurs de définition et des descripteurs de conditions (Buvet, 2009b). La fonction du premier type de descripteurs est d'explicitier les propriétés sémantiques des emplois : la racine prédictive, la classe sémantique, le type sémantique et l'aspect inhérent. Voici un tableau d'exemple de deux emplois du verbe *dévor* :

entrée (emploi)			Définition		
emploi	indice	Racine prédictive	Classe	type	aspect inhérent

<i>Dévor</i>	1	dévor-	<alimentation>	Act	Duratif perfectif
<i>Dévor</i>	2	dévor-	<lecture>	Act	Duratif perfectif

Tableau 12 : *Descripteurs de définition de dévor* (extrait de Buvet, 2009b, p. 44)

Quant aux descripteurs de condition, leur fonction est de vérifier que les emplois prédicatifs ont des propriétés spécifiques qui les caractérisent. Ces caractéristiques spécifiques sont : la construction (schéma prédicatif), la distribution morphosyntaxique et la distribution syntacticosémantique. Le tableau ci-dessous illustre les propriétés de conditions de *dévor* :

entrée (emploi)			Conditions	
emploi	indice	Construction	distribution 1	distribution 2
<i>Dévor</i>	1	X0 V X1	X0 = GN X1=GN	X0=<humain> X1=<aliment>
<i>Dévor</i>	2	X0 V X1	X0 = GN X1=GN	X0=<humain> X1=<texte>

Tableau 13 : *Descripteurs de définition de dévor* (extrait de Buvet, 2009b, p. 44)

Nous observons que les deux emplois de *dévor* ont le même schéma prédicatif et ont le même type de construction en position argumentale. La différence de ces emplois réside dans la nature sémantique de l'argument 1 (premier complément). C'est pourquoi ces deux emplois appartiennent à des classes d'objets différentes.

On notera que ce système d'étiquetage permet l'affectation multiple, pour les cas où un emploi est à la fois prédicat et argument.

Finalement, pour élaborer les dictionnaires des actualisateurs, on fait une première distinction entre l'actualisation argumentale et l'actualisation prédicative. Ensuite, une seconde distinction est faite selon qu'ils portent une signification grammaticale faible ou forte. Par exemple, soit les phrases⁴³⁷ :

(2) a. *Léa a du respect pour l'équipe avec laquelle elle travaille.*

b. *Léa a beaucoup de respect pour l'équipe avec laquelle elle travaille.*

⁴³⁷ Exemples adaptés de Massoussi et Mejri (2015, p. 91).

Dans la phrase (2a) la séquence [vebe support + nom prédicatif] équivaut au verbe respecter et le déterminant a une valeur grammaticale faible. Par contre, dans (2b) l’adverbe *beaucoup*, ayant un aspect intensif, apporte une valeur forte à la séquence.

Les dictionnaires des actualisateurs recensent des éléments tels que des verbes supports, les verbes auxiliaires, les constructions déterminatives des noms, les prépositions et locutions prépositives, etc. Ils portent donc sur des éléments linguistiques très variés et qui demandent un traitement différent. Il s’avère donc très difficile d’établir un format commun pour tous les types d’actualisateurs. Pour ça, d’habitude chaque étude porte sur un seul type d’actualisateur. À titre d’exemple, nous tenons à signaler le travail de G. Gross (2004a) sur les actualisateurs des prédicats nominaux, c’est-à-dire les verbes supports, ou celui de Buvet (2008a) portant sur les actualisateurs déterminatifs.

4.1.2. Grammaires locales

« Une grammaire locale décrit le contexte d’une unité lexicale donnée en tant qu’ensemble de configurations de mots » (Buvet, 2009b, p. 44). Une grammaire locale est représentée sous forme de graphe orienté. Celui-ci contient un nœud initial, des nœuds intermédiaires et un nœud final. Ces nœuds sont reliés par des arcs, en fonction des configurations de mots de la grammaire locale. Les nœuds intermédiaires sont associés à des mots, des lemmes, des catégories grammaticales ou d’autres informations métalinguistiques encodées dans les dictionnaires électroniques.

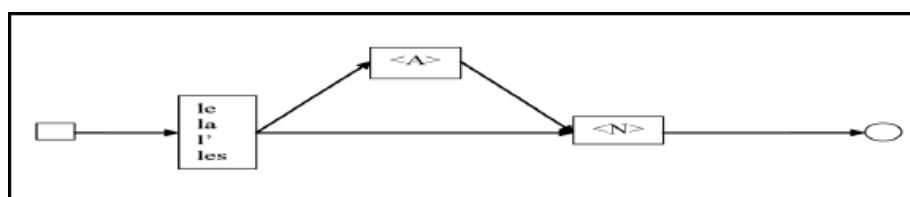


Figure 26 : Graphe de groupe nominal (extrait de Buvet, 2009b, p. 44)

Le graphe se lit de gauche à droite. Dans les nœuds intermédiaires, les items en minuscule sont utilisés pour les formes d’un lemme ; ceux en majuscules, pour les catégories grammaticales. Les chevrons simples (<>) marquent que ces parties doivent faire l’objet d’ajustements grammaticaux (accord, élision, contraction). La figure précédente représente le graphe qui permet d’identifier les groupes nominaux du type *article défini + nom* ou du type *article défini + adjectif + nom*. Puis, il faudrait d’autres graphes, pour spécifier, par exemple, l’utilisation de *l’*.

Un message peut être exprimé de plusieurs façons. Toutes ces possibilités de réalisation discursive sont recueillies dans des grammaires locales. Les grammaires locales sont élaborées à partir de ces données morphologiques, syntaxiques et sémantiques contenues dans les dictionnaires électroniques. Une grammaire locale est un regroupement d'analyses lexicales, sous forme de graphes. Ces graphes, manipulés par un automate à états finis, permettent de décider si une séquence de mots donnée correspond à l'une des configurations de la grammaire locale. Par exemple, un graphe regroupe des synonymes (des unités lexicales simples et polylexicales) ; ensuite, les différentes branches donnent l'ensemble des possibilités de combinatoire⁴³⁸. Soit le graphe suivant :

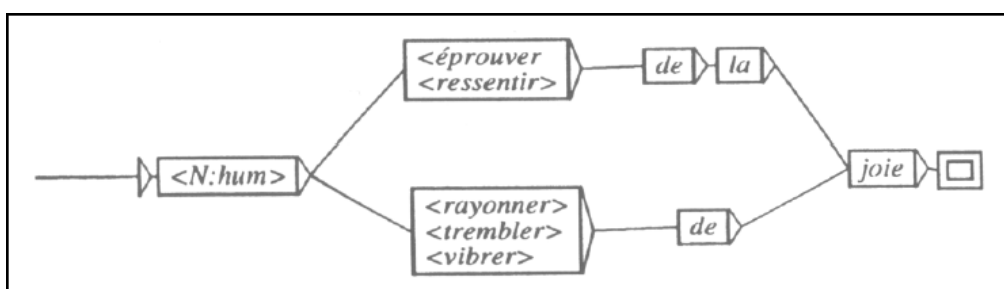


Figure 27 : Graphe de phrases avec le prédicat de < sentiment > joie (extrait de M. Gross, 1995, p. 74)

Dans ce graphe, chacun des parcours correspond à une phrase. Ce graphe comporte des phrases du type *Luc ressent de la joie* ou *Luc rayonne de joie*.

Le transducteur à états finis est un automate qui associe une nouvelle information à l'information reconnue. Cette fonctionnalité est capitale pour l'étiquetage. Le transducteur permet aussi de remplacer les items d'une séquence reconnue par d'autres items, c'est-à-dire il permet toutes sortes de manipulations des phrases (Buvet, 2009b). Les transducteurs et les automates à états finis exploitent les données des dictionnaires électroniques.

4.1.3. Corpus

⁴³⁸ La formalisation des propriétés combinatoires est difficilement compatible avec la structure des dictionnaires. C'est pourquoi elles sont prises en charge dans des grammaires locales (Buvet, 2009b). Pour une étude approfondie des grammaires locales, leurs atouts et les problèmes d'analyse qu'elles posent, voir Constant (2003).

L'élaboration d'outils pour le TAL requiert trois sortes de corpus avec lesquels on travaille à des étapes successives : un corpus d'investigation, un corpus de test et un corpus de validation. Pour éviter des solutions *ad hoc*, il est nécessaire que ces trois corpus portent sur des contenus différents. Nous expliquerons brièvement leurs fonctions à la lumière de l'article de Buvet (2013a).

Le corpus d'investigation sert à reconnaître les phénomènes linguistiques qui seront traités. Ce corpus aide à identifier les ressources lexicales nécessaires au bon fonctionnement des outils d'analyse automatique et il permettra de calibrer les transducteurs à états finis qui ont déjà été élaborés ou, si nécessaire, de concevoir de nouveaux transducteurs.

Ensuite, ces outils calibrés ou nouvellement développés seront testés dans un autre corpus (le corpus de test). Si ces expérimentations donnent lieu à du bruit ou à du silence, c'est dans cette étape qu'on corrigera les outils qui ne sont pas performants.

Enfin, le travail dans le corpus de validation sert à vérifier la qualité des outils élaborés ou paramétrés. En cas d'invalidation, on reprendra le travail en utilisant d'autres corpus d'investigation et de test.

4.2. Étapes tu traitement automatique

Ayant décrit les ressources linguistiques dont le système se sert, nous allons expliquer de manière succincte le fonctionnement de ce système. Selon Issac (2016), il existe « un continuum dans l'analyse des textes qui part du caractère pour progresser jusqu'au sens en passant par des unités plus importantes pouvant être assimilées à des mots » (p. 100). Chaque étape du traitement est fondée sur des choix méthodologiques ou des choix portant sur la nature des données à traiter qui vont nécessairement amener à des résultats différents.

Le système de traitement automatique développé dans le cadre du modèle des classes d'objets commence par une phase de prétraitement, qui précède les trois phases de traitement proprement dit. Enfin, l'objectif est d'obtenir un document entièrement étiqueté et dépourvu d'ambigüités, qui pourraient provoquer du bruit ou du silence dans l'analyse.

4.2.1. Le prétraitement : normalisation des documents

Le traitement automatique des langues demande, dans un premier temps, une représentation du discours dans un format informatiquement manipulable et normalisé (Heitz, 2006). Dans cette étape, les documents du corpus sont normalisés en les convertissant en format XML et en encodage UTF-8. Actuellement, une DTD⁴³⁹ minimale permet d'annoter les titres, sous-titres, sections, sous-sections et paragraphes des textes. Ainsi, cette normalisation permet de baliser les textes de la manière suivante : `<text><title level='1..N'><section level='1..N'><p>` (Buvet et al., 2007, p. 241). À la fin de cette phase, le résultat est une suite de caractères qui respectent les règles de typographie⁴⁴⁰ et où la mise en forme est absente (cf. Cartier et Issac, 2009).

4.2.2. Segmentation en phrases et en mots

Il s'agit de la première phase du traitement proprement dit. La première segmentation effectuée est la segmentation en phrases. À cette étape, la phrase est définie du point de vue graphique : une majuscule indique le début de phrase, et un point, point d'exclamation ou point d'interrogation indiquent la fin. Le corpus est analysé en incorporant une nouvelle balise : `<s>` (pour *sentence* en anglais). On notera que tous les points qui apparaissent dans un texte ne correspondent pas à la fin d'une phrase. Pour éviter ce bruit, des listes d'abréviations et d'acronymes ont été élaborées, ainsi qu'un ensemble d'expressions régulières⁴⁴¹. De cette manière les points qui ne représentent pas de fins de phrase, ne seront pas suivis de la balise `</s>`.

Une partie du corpus (150 000 tokens) a été annotée manuellement pour établir un corpus de référence. Les résultats de la segmentation automatique en phrases sont

⁴³⁹ La DTD (*Document Type Definition*) est la structuration d'un texte. Il s'agit d'un document informatique qui permet de définir un modèle de structuration des données contenues dans un document XML.

⁴⁴⁰ Nous entendons par typographie tout ce qui concerne la manière dont on représente la langue écrite : la représentation graphique des mots, des nombres, des acronymes, etc.

⁴⁴¹ En informatique, une expression régulière est une chaîne de caractères qui décrit, en utilisant une syntaxe précise, un ensemble variable de chaînes de caractères ; par exemple, `^[A-Z]` est une expression régulière qui va permettre de reconnaître toutes les majuscules en début de ligne.

très satisfaisants : le taux de rappel est de 98,7 % et le taux de précision est de 97,6 % (Buvet et al., 2007).

La phase de segmentation en tokens (dite également tokenisation) est très complexe. Cette phase consiste à regrouper un ensemble de caractères pour reconnaître des unités linguistiques diverses : des mots, des acronymes, des abréviations, des entités numériques (6,5 millions, 7 %), des entités numériques à base temporelle (08/09/2018), des entités spécifiques (*url*), des symboles (=). Chacune de ces unités est considérée une unité linguistique minimale (token).

D'un point de vue informatique, un texte est une suite de caractères. La segmentation de la phrase en tokens se fait sur la base de caractères non alphabétiques : l'espace, les signes de ponctuation, les apostrophes, le tiret, les guillemets, etc. Pourtant, certains de ces éléments requièrent des traitements particuliers : ni le tiret ni l'espace ni l'apostrophe ne sont nécessairement des séparateurs, car ils peuvent être présents dans des mots composés (une *voiture-balai*, un *a b c*) ou faire partie de mots simples (*aujourd'hui*). Dans ces cas, il faut élaborer des dictionnaires contenant ces mots, afin de pouvoir reconstruire le token après découpage ou de pouvoir empêcher ce découpage.

La segmentation en tokens se fait au moyen de balises qui déterminent le type d'entité dont il s'agit : un mot (<w>), une information numérique (<num>), un signe de ponctuation (<punct>) ou un symbole (<symb>). Également, certains attributs ont leurs propres attributs ; par exemple, la case *w* a ses attributs de *case* (*case='lowercase, uppercase, titlecase'*), et *num* a un attribut *type* (qui permettra de reconnaître, par exemple, les dates) (Buvet et al., 2007).

4.2.3. Annotation morphosyntaxique

Dans cette phase, les informations du dictionnaire morphologique sont associées aux tokens⁴⁴². Ainsi, chaque forme (token), qui comprend une ou plusieurs entrées dans

⁴⁴² Cette étape et la suivante (annotation syntacticosémantique) ne sont pas obligatoires en TAL. Par exemple, pour les analyses statistiques, le traitement et les données obtenues dans les étapes précédentes sont suffisantes.

le dictionnaire, se voit associer les différents attributs des lemmes correspondants (en fonction de la partie du discours dont il s'agit) : le lemme, la catégorie morphosyntaxique, des informations sur le nombre, le genre, le temps, le mode et la personne. Dans cette phase, la plupart des mots sont ambigus. Par exemple, *porte* comporte toutes les informations liées aux formes verbales (mode : indicatif, subjonctif ; temps : présent, personne\$1 : 1, 3) et à la forme nominale (nom, féminin, singulier) (Buvet et al., 2007). Ces informations associées apparaissent dans le fichier XML sous l'annotation suivante :

```

<w                                     case='lc'>
  <morph lemma='porter' cat='v' mood=''subj' tense='pres' pers='1' />
  <morph lemma='porter' cat='v' mood=''subj' tense='pres' pers='3' />
  <morph lemma='porter' cat='v' mood=''ind' tense='pres' pers='1' />
  <morph lemma='porter' cat='v' mood=''ind' tense='pres' pers='3' />
  <morph lemma='porte' cat='n' nb='plu' gender='fem' />
  porte
</w>

```

Figure 28 : Annotation morphosyntaxique de *porte* (extrait de Buvet et al., 2007, p. 242)

4.2.4. Annotation syntacticosémantique

Après l'annotation morphologique, chaque élément de la phrase est associé à plusieurs étiquettes, c'est-à-dire les tokens ont plusieurs interprétations possibles. La prise en compte du contexte linguistique permettra de rejeter les étiquettes incorrectes. À ce stade, le système fait appel à des grammaires locales, au dictionnaire des arguments et à celui des prédicats. Cette phase est la plus complexe du traitement.

La première levée d'ambiguïté se fait en appliquant des grammaires locales. Celles-ci sont fondées sur la syntaxe lexicale. Ces grammaires locales permettent de faire une analyse syntaxique à partir des éléments lexicaux en contexte. Le principe est le suivant : parmi toutes les étiquettes associées à un token donné, le système gardera celle dont l'existence est validée par une grammaire locale. Ainsi, dans la séquence *une porte*, où *une* est associé à l'étiquette ARTICLE, et *porte* est associé aux étiquettes VERBE ou NOM, le système éliminera l'étiquette VERBE, du fait de son incompatibilité avec l'article indéfini.

Pourtant, l'application de ces grammaires peut ne pas être suffisante. Soit la phrase :

(3) a. *Le secrétaire vole des livres.*⁴⁴³

L'application de grammaires locales donnerait le résultat suivant :

le : article masculin singulier

secrétaire : nom masculin singulier

vole : voler (transitif) 3^e personne singulier indicatif présent

des : article indéfini pluriel

livres : nom masculin pluriel

Comme les noms *secrétaire* (meuble, humain, oiseau) et *livre* (ouvrage, argent) restent ambigus, le système peut générer six analyses différentes :

- a. un oiseau dérobe un ouvrage
- b. un oiseau dérobe de l'argent
- c. un meuble dérobe un ouvrage
- d. un meuble dérobe de l'argent
- e. un humain dérobe un ouvrage
- f. un humain dérobe de l'argent

Un être humain peut facilement éliminer les interprétations (c) et (d), et hésiter quant à (a) et (b), mais l'automate a besoin de plus de données pour être capable d'éliminer n'importe quelle interprétation. Ces données sont compilées dans le dictionnaire des arguments et le dictionnaire des prédicats. L'étape suivante consiste donc à projeter le dictionnaire des arguments sur le fichier XML dans le but d'ajouter des données sémantiques à certains éléments de type mot (<w>).

Ensuite, le système identifie diverses constructions syntaxiques : des groupes nominaux, groupes prépositionnels, etc. La tête de chaque syntagme est identifiée et balisée au moyen de <head>. À cette étape, des séquences relatives à des entités nommées (organisations, personnes, lieux, événements, artefacts) sont identifiées grâce à des dictionnaires spécifiques.

⁴⁴³ Exemple emprunté à Cartier et Issac (2009, p. 149).

En dernier lieu, l'application du dictionnaire des prédicats permet de repérer les structures prédicat-arguments. L'identification des schémas prédicatifs et l'analyse et prise en compte d'un contexte plus large permettront enfin de résoudre les cas d'ambiguïté qui persistaient. Nous illustrons ci-dessous (Figure 23) le XML résultant du processus d'analyse syntacticosémantique de la phrase *la population a subi une hausse*⁴⁴⁴ :

```

<s>
  <gn head='12/12/2003'>
    <w case='tc' ><morph lemma='le' cat='det'>Le</morph></w>
  </gn>
  <pred sem='subir' X0='population' X1='hausse'>
    <gn head='population'>
      ...
    </gn>
    </gv head='subir'>
      ...
    </gv>
    </gn head='hausse'>
      ...
    </gn>
  </pred>
    <w type='punct'>,</w>
    <gadj head='atteindre'>
      ...
    </gadj>
    <punct>.</punct>
</s>

```

Figure 29 : Représentation XML de l'analyse prédictive (extrait de Buvet et al., 2007, p. 243)

Cette phase d'interprétation est fondamentale : elle permet d'associer aux tokens une classe sémantique et une acception, à partir de critères configurationnels et distributionnels.

5. Conclusions

Au tournant du XX^e siècle le traitement automatique des langues naturelles représentait un défi majeur, voire insurmontable pour certains. Les spécificités des langues naturelles (la polysémie, la polymorphie, le figement, la paraphrase, entre autres) semblaient s'opposer diamétralement aux spécificités et besoins du traitement automatique (à savoir l'univocité, la cohérence, la régularité, etc.). Pourtant, ces deux

⁴⁴⁴ Nous avons marqué en rouge la partie correspondante à l'analyse syntacticosémantique.

disciplines se sont tellement bien intégrées qu'elles ont donné naissance à ce qu'on appelle la linguistique informatique.

La recherche en mathématiques et concrètement, le développement informatique ont transformé la manière d'appréhender la description des langues. L'avènement de l'informatique dans le domaine de la linguistique a supposé une vraie révolution dans ses méthodes et ses objectifs. Les approches théoriques actuelles de description de la langue sont fortement déterminées par l'exploitation automatique. Néanmoins, nous avons essayé de montrer que le rôle du linguiste reste fondamental. Nous considérons qu'une position intermédiaire entre l'approche quantitative (ou statistique) et l'approche qualitative reste la meilleure place pour observer, analyser et décrire les faits de langue. Ainsi, nous avons présenté quelques ressources lexicales nées de la collaboration entre linguistes et informaticiens. Le grand intérêt de ces ressources, à part la description fine et pertinente de la langue aux niveaux morphologique, syntaxique et sémantique, tient au fait qu'elles sont exploitables par des logiciels de traitement automatique et, en même temps, elles disposent d'interfaces qui les rendent consultables par l'être humain.

Enfin, nous avons fait une incursion sur les méthodes, les outils et les processus de traitement automatique dans le cadre du modèle des classes d'objets. Nous concevons cette dernière section du Chapitre IV comme le pont ou le passage vers le Chapitre V, dans lequel nous expliquons la méthodologie de recherche suivie pour l'élaboration du dictionnaire des noms prédicatifs des <aides>.

CHAPITRE VI : MÉTHODOLOGIE

Notre méthodologie de recherche est fortement déterminée par le cadre théorique dans lequel on aborde la description de la langue et la visée en TAL. D'une part, comme le souligne Grezka (2006) : « dans le modèle des classes d'objets, l'une des caractéristiques essentielles est l'approche syntacticosémantique du lexique qui interdit de traiter séparément le lexique, la sémantique et la syntaxe ». D'autre part, si l'on veut obtenir une description des noms des <aidesnancières> visant le TAL, il faudra tenir compte des spécificités que celui-ci impose (cf. Chap. V). Ces spécificités constituent un défi et un important avantage en même temps. En effet, la description des faits de langue doit être nécessairement exhaustive et systématique. Ces deux conditions nous obligent à expliciter tout ce qui pourrait être sous-entendu par le cerveau humain, mais elles nous mènent également vers une description ayant un haut degré de finesse.

Ce chapitre est organisé en deux grands volets. Dans un premier temps, nous présentons les principes d'analyse du modèle des classes d'objets. Toute théorie est fondée sur des principes d'analyse qui dérivent directement de ses postulats théoriques et définissent la méthodologie de recherche. Dans un deuxième temps, nous présentons les grandes étapes de notre travail de recherche et les outils linguistiques dont nous avons eu besoin pour décrire les prédicats nominaux des <aides financières>.

1. Principes d'analyse

Dans le modèle des classes d'objets, la description de la langue est fondée sur des principes d'analyse précis, qui ont déjà été esquissés dans la description du cadre théorique⁴⁴⁵. Ces principes sont les règles méthodologiques que nous avons suivies dans le processus d'élaboration du dictionnaire, dans le but de décrire de manière précise les phénomènes considérés. Ils s'avèrent nécessaires à l'élaboration de taxonomies suffisamment cohérentes pour dévoiler les régularités et expliquer le fonctionnement des unités lexicales décrites. À continuation, nous présentons les six principes d'analyse

⁴⁴⁵ Voir Chap. IV *Le modèle des classes d'objets*.

qui sont à la base des recherches lexicographiques dans le cadre du modèle des classes d'objets (cf. Blanco et Buvet, 2000 ; Grezka, 2006 ; Buvet et Grezka, 2007).

1.1. La langue comme métalangue⁴⁴⁶

À la suite de M. Gross (1975) et Harris (1976), nous considérons que la description du fonctionnement d'une langue donnée doit se faire exclusivement au moyen des propriétés linguistiques. Donc pour l'étude des faits de langue les seuls éléments d'explication pris en compte sont d'autres faits de langue. Les phénomènes langagiers sont étudiés indépendamment d'autres phénomènes avec lesquels ils interagissent (p. ex., des phénomènes à caractère psychologique, des aphorismes, etc.). Le recours exclusif à des propriétés linguistiques permet de vérifier de manière objective l'interprétation des données observées et, dans notre cas spécifique, il évite le recours à des descriptions ontologiques, subjectives et arbitraires.

1.2. La phrase élémentaire comme cadre d'analyse

Le principe d'analyse précédent implique de préciser la nature des propriétés linguistiques utilisées dans la description de la langue. Nous avons choisi des propriétés qui rendent compte des relations entre les unités lexicales au niveau de la phrase : des propriétés configurationnelles et combinatoires⁴⁴⁷. Ainsi, ce deuxième principe n'exclut pas d'autres propriétés linguistiques (p. ex., les propriétés sémantiques ou morphologiques), mais celles-ci sont subordonnées aux propriétés mentionnées. De cette façon, les unités lexicales sont décrites en termes d'emplois dans le cadre des phrases où elles apparaissent.

1.3. L'approche syntacticosémantique du lexique

Au lieu de procéder par niveau d'analyse (d'abord, l'étude de la morphologie, ensuite, celle de la syntaxe et enfin, la sémantique), dans le cadre du modèle des classes

⁴⁴⁶ Nous concevons la métalangue à la manière de Rey-Debove (1972) : « La métalangue d'une langue donnée est un sous-système de cette langue qui permet de produire des discours sur la langue en question. Ce sous-système est caractérisé par des règles syntaxiques, morphologiques, phonologiques qui lui sont propres par rapport au système complémentaire du discours sur le monde » (p. 11).

⁴⁴⁷ Voir Chap. VI, section 2.3. *Constitution d'une grille d'analyse*.

d'objets le sens des unités lexicales est défini en fonction de leurs caractéristiques morphosyntaxiques. Ce principe dérive donc du deuxième, car c'est dans le contexte de la phrase qu'il est possible d'étudier le lexique sur les plans morphologique, syntaxique et sémantique conjointement⁴⁴⁸.

1.4. Le souci d'exhaustivité et de systématicité

Le recensement exhaustif des unités lexicales observées dans le contexte de la phrase permet de faire apparaître les phénomènes massifs et les possibles irrégularités. Il s'agit donc de procéder à l'élaboration d'inventaires les plus complets possible et de le faire de manière systématique pour que le système de TAL puisse utiliser ces données.

1.5. La nécessité d'homogénéité

Nous avons regroupé dans une même classe les emplois qui présentent suffisamment de propriétés communes. Afin de respecter le principe antérieur, les regroupements sont effectués à partir de l'observation exhaustive et systématique. Ainsi, nous avons obtenu des classes cohérentes et régulières qui rendent compte du fonctionnement des unités lexicales analysées.

1.6. L'analyse ascendante

Si les classes mentionnées ci-dessus, qui constituent des unités minimales de description, partagent certaines propriétés communes avec d'autres classes, elles peuvent être regroupées dans des hyperclasses. Il en résulte que les taxonomies élaborées dans le modèle des classes d'objets sont ascendantes, contrairement aux taxonomies résultantes des descriptions générales, ontologiques ou intuitives.

Ces principes d'analyse qui sous-tendent l'étude des unités lexicales dans le modèle des classes d'objets sont enchâssés les uns dans les autres : comme la phrase élémentaire est le cadre d'analyse des unités lexicales, l'approche ne peut être que syntacticosémantique ; le souci d'exhaustivité et de systématicité permet une description homogène des phénomènes linguistiques ; les phénomènes massifs mis en

⁴⁴⁸ Rappelons que ces trois niveaux d'analyse sont regroupés dans la notion d'emploi. Les entrées du dictionnaire que nous avons élaboré correspondent à des emplois.

évidence grâce au principe d'exhaustivité de d'homogénéité permettent aussi de procéder à une analyse ascendante.

2. Étapes de travail et outils d'analyse

Le modèle des classes d'objets relève de la linguistique descriptive, c'est-à-dire il vise l'explicitation de tous les emplois des unités lexicales d'une langue donnée, observés en contexte.

Le processus du travail n'a pas été fait de manière simultanée dans les deux langues. Nous avons d'abord analysé les prédicats nominaux des <aides> en français, et ensuite, en espagnol. Pour ce qui est de l'analyse et la description, nous avons suivi les mêmes étapes dans les deux langues. Premièrement, nous avons établi la nomenclature à décrire. Deuxièmement, nous avons étudié en contexte chaque unité lexicale de notre nomenclature à l'aide de corpus déjà constitués⁴⁴⁹ et notamment à l'aide du moteur de recherche Google. Troisièmement, à partir de l'observation des unités lexicales en contexte, nous avons pu élaborer une grille d'analyse qui rend compte des propriétés morphologiques, sémantico-syntaxiques et pragmatiques. Enfin, l'application de la grille d'analyse nous a permis d'identifier et de lister tous les emplois des mots retenus dans la nomenclature, et de regrouper en classes d'objets les emplois qui partagent les mêmes propriétés syntactico-sémantiques. Ainsi, nous avons obtenu deux dictionnaires électroniques de prédicats nominaux : l'un en français et l'autre en espagnol. Rappelons qu'il s'agit de deux dictionnaires d'emplois de mots et que cette notion d'emploi regroupe des propriétés lexicales, syntaxiques et sémantiques. À continuation nous présentons ces trois grandes étapes de notre travail.

2.1. Établissement de la nomenclature

La spécificité de notre objet d'étude réside dans le fait qu'il est très hétérogène. D'une part, nous avons recensé autant des unités monolexicales et que des unités polylexicales. D'autre part, nous avons recensé des emplois généraux et des emplois spécialisés appartenant à des domaines de spécialité différents. En conséquence, les

⁴⁴⁹ Ils sont précisés ultérieurement, dans la section 2.2. *Constitution du corpus*.

sources d'information utilisées pour constituer la nomenclature ont été très variées et multidisciplinaires.

Le recensement des noms prédicatifs de <aide financière> s'est appuyé sur des critères sémantiques, à partir de leur signification telle qu'elle est précisée dans les ouvrages lexicographiques. Nous avons d'abord consulté de nombreux dictionnaires de la langue générale et de la langue spécialisée, ainsi que des ouvrages spécialisés⁴⁵⁰. Comme résultat de cette première phase, nous avons prélevé les données qui constituent une première base lexicale de noms d'<aide financière> et que nous avons développée par la suite de deux façons différentes. En ce qui concerne les mots de la langue générale, nous avons procédé à la recherche des synonymes dans des dictionnaires généraux et dans des dictionnaires de synonymes, afin d'enrichir la nomenclature initialement établie. Quant aux termes spécialisés (comme c'est le cas, par exemple, des noms concrets de certaines allocations), nous nous sommes centrés notamment sur l'exploration du web (des sites d'institutions, d'associations, d'entreprises, de forums spécialisés⁴⁵¹, etc.). Les requêtes sur le web des termes spécialisés se sont avérées plus fructueuses que la consultation d'ouvrages spécialisés imprimés. Premièrement, parce que notre champ d'étude (les aides financières) est en évolution perpétuelle ; de nouvelles aides sont mises en place adaptées à la réalité changeante. Deuxièmement, l'exploitation du web nous a permis de recenser des emplois plus ou moins communs qui étaient passés sous silence dans les dictionnaires et ouvrages spécialisés.

La tâche de recensement (et celle d'analyse syntacticosémantique) des termes a été bien plus compliquée du fait de l'aspect pluridisciplinaire de notre objet d'étude et des connaissances extralinguistiques qui nous étaient imposées pour décrire ces unités lexicales. En effet, la description de ces unités lexicales nécessite non seulement une compétence linguistique, mais aussi une compétence extra-linguistique qui nous

⁴⁵⁰ Nous invitons le lecteur à consulter la section Outils lexicographiques utilisés pour prendre connaissance des ouvrages consultés.

⁴⁵¹ Par exemple, pour le français, nous avons consulté le portail du Sénat et de la Caisse d'Allocations Familiales (entre autres) pour récupérer les types d'allocations qui existent ; le portail de nombreux ministères tels que le Ministère des Affaires étrangères, le Ministère de l'Enseignement supérieur et de la Recherche et de l'Innovation, etc.

permette de bien comprendre le sens de ces termes et leur fonctionnement dans la phrase.

Le caractère changeant de certains de ses termes nous a confrontée à un problème de type méthodologique : il a fallu déterminer si l'on inclurait dans notre nomenclature les noms des aides financières (notamment des allocations) qui n'existent plus ou qui ont changé de nom⁴⁵². Comme il s'agit de noms qui se trouvent toujours dans les archives et sur la toile, il se peut que l'automate doive en faire l'analyse. De plus, ces termes font toujours partie du vocabulaire des locuteurs natifs. Nous avons donc décidé de les recenser et de les décrire. Néanmoins, nous tenons à insister sur le fait que le recensement de noms d'aides appartenant aux discours de spécialité est exhaustif, mais il ne pourra jamais être complet ou définitif.

En outre, nous avons pu constater qu'il existe des variations diatopiques concernant le lexique des domaines étudiés. On notera que le nom des aides allouées par les institutions publiques, par exemple, n'est pas nécessairement le même dans tous les pays hispanophones ou francophones. D'ailleurs, il n'existe pas les mêmes types de revenus d'assistance dans tous les pays. Il a donc fallu trancher : dans ce travail de recherche nous décrivons les noms prédictifs des <aides financières> de l'espagnol de l'Espagne et du français de la France.

2.2. Constitution du corpus

Puisque l'un des principes méthodologiques qui ont guidé notre recherche est de considérer la phrase élémentaire comme le cadre d'analyse des emplois des unités lexicales, et que nous prétendons que la description de la langue soit fondée sur des méthodes strictement linguistiques, l'exploitation du corpus s'est avérée nécessaire. En effet, l'écart entre l'intuition du linguiste et les données récoltées objectivement à partir des textes peut être énorme et systématique (Sinclair, 1991).

Lerat (2002) soutient que « les corpus sont indispensables à la description des emplois des mots spécialisés. Corpus textuels mais aussi corpus lexicographies » (p.

⁴⁵² Par exemple, en France, l'*allocation RMI (revenu minimum d'insertion)*, en vigueur entre 1988 et 2009, a été remplacée par le *RSA (revenu de solidarité active)*, une allocation plus large.

205-204). Ainsi, nous avons utilisé ces deux types de corpus pour recenser les collocations verbales. Premièrement, nous avons fait des recherches dans des dictionnaires généraux⁴⁵³ et dans des dictionnaires de combinaisons de mots⁴⁵⁴ pour relever une première liste de collocations, qui parfois était loin d'être exhaustive (notamment dans le cas des noms utilisés en discours de spécialité). Ensuite, nous avons lancé des requêtes dans des corpus textuels informatisés pour observer le comportement de ces noms prédicatifs en contexte et dans l'espoir e pouvoir repérer des collocations qui ne sont pas recensées dans les ouvrages lexicographiques.

L'accès à des corpus informatisés a favorisé l'intérêt des lexicographes vers les affinités entre les unités lexicales, en les considérant de véritables unités fonctionnelles du discours (Sinclair, 1991). En effet, les possibilités techniques offertes par le support informatique ont permis de mettre en lumière les cooccurrences. Également, le recours à des corpus informatisés a facilité l'étude des relations entre les environnements syntagmatiques des unités lexicales et le sens de celles-ci. Les lexicographes ne portent plus leur intérêt sur le mot isolé, mais sur l'unité lexicale insérée dans un environnement, c'est-à-dire sur ce que nous avons défini comme l'emploi des mots. Ainsi, « le corpus aura donc contribué à faire en sorte que l'unité lexicale soit désormais vue comme un évènement, l'élément d'une situation, d'un cadre, et non plus seulement comme un objet isolé » (Béjoint, 2007, p. 19).

Comme point de départ dans la constitution du corpus, nous avons dû réfléchir au type de corpus dont nous aurions besoin pour trouver des occurrences des noms prédicatifs faisant partie de notre nomenclature. Nous avons été guidée par un souci d'homogénéité et d'hétérogénéité en même temps. D'une part, nous tenions à travailler sur un corpus suffisamment homogène pour pouvoir obtenir des données fiables. Ainsi, nous avons considéré certaines caractéristiques qui garantiraient l'homogénéité du corpus :

⁴⁵³ Pour le français nous avons utilisé le *Petit Robert* en version numérique et le *TLFi*. Pour l'espagnol, nous avons consulté le *Diccionario de uso del español* de María Moliner.

⁴⁵⁴ En français, nous avons consulté le *Dictionnaire des combinaisons de mots* du Robert (Le Fur, 2007) et le dictionnaire des cooccurrences contenu dans l'outil *Antidote* (en différentes versions, de la version 6.4 à la 10). En espagnol, nous avons utilisé les dictionnaires *Redes* (Bosque, 2004) et *Práctico* (Bosque, 2006).

- établir des limites chronologiques (XX^e et XXI^e siècles) ;
- bannir les variations diatopiques ;
- bannir les textes poétiques, car le souci stylistique de celle-ci peut donner lieu à un discours ne correspondant pas à l'usage courant de la langue ;
- bannir les textes traduits, c'est-à-dire ne travailler que sur des textes originaux.

D'autre part, nous tenions également à ce que le corpus soit suffisamment hétérogène pour qu'il soit représentatif du français ou de l'espagnol appartenant au discours général et au discours spécialisé. Ainsi, il s'avérait nécessaire de tenir compte de textes littéraires, journalistiques, administratifs, etc.

La constitution du corpus doit également répondre au souci de pertinence par rapport aux domaines abordés. Comme la nomenclature de notre dictionnaire relève autant des emplois de la langue générale, que des emplois spécialisés, notre étude s'est appuyée sur deux types différents de ressources : de vastes corpus généraux et le web. En ce qui concerne les corpus généraux, nous avons utilisé Frantext⁴⁵⁵, pour le français, CREA⁴⁵⁶ et CORPES XXI⁴⁵⁷, pour l'espagnol, et Sketch Engine⁴⁵⁸ (Kilgarriff et al.,

⁴⁵⁵ Frantext (<https://www.frantext.fr/>) est une base de données, développée au sein du laboratoire ATILF (*Analyse et Traitement de la Langue Française* ; UMR 7118 ; CNRS-Université de Lorraine). Elle comporte 251 millions de mots issus de textes du IX^e au XXI^e siècle (la proportion de textes appartenant au XX^e siècle est quantitativement bien plus importante). Cette base de données contient 90 % de textes littéraires (tous les genres y sont représentés : roman, théâtre, poésie, essai mémoire, journal personnel, etc.) et 10% de textes scientifiques ou techniques. La majorité de ces ouvrages sont français, mais il y en a aussi qui sont issus du monde francophone. Les textes contenus dans cette base de données sont lemmatisés et étiquetés en parties du discours. Ils sont exploitables grâce à une interface de recherche, qui permet d'établir de nombreux critères dans les requêtes.

⁴⁵⁶ Le Corpus de Referencia del Español Actual (CREA ; <https://www.rae.es/banco-de-datos/crea>) est une base de données développée par la Real Academia de la Lengua Española. Il comporte des textes écrits et des transcriptions de productions orales datés entre 1975 et 2004. La partie du CREA composée par des textes écrits contient plus de 154 millions de mots, dont 49 % sont issus de livres (littérature, art, santé, sciences sociales, sciences et technologie, loisirs, etc.), un autre 49 % procèdent de la presse et le 2 % restant sont composés de matériels divers. Quant à l'origine géographique, 50 % de ces productions écrites proviennent d'Espagne, et l'autre 50 % des textes proviennent de pays hispanophones de

2004, 2014) pour les deux langues. Les interfaces de ces corpus nous ont permis de cibler nos requêtes en sélectionnant certains critères ; par exemple, la délimitation chronologique (seulement des textes du XX^e et du XXI^e siècles), la délimitation géographique (des textes produits en France et en Espagne), le genre littéraire (nous avons banni la poésie), etc. L'exploitation de ces corpus a parfois fourni de maigres résultats, notamment dans les domaines de spécialité. En conséquence, nous avons considéré nécessaire d'exploiter le web comme corpus.

L'emploi du web comme corpus, une tendance connue sous le nom *web as corpus*, ne fait pas l'unanimité parmi les linguistes, mais elle a attiré de plus en plus

l'Amérique. Nous avons notamment utilisé la dernière version du CREA lemmatisée et étiquetée en parties du discours (version 0.3, mars 2021) et une autre version antérieure (version 3.2, juin 2008). Il existe une version intermédiaire du CREA (version 0.1, novembre 2015), mais elle est nettement plus réduite et nous ne l'avons guère utilisée.

⁴⁵⁷ Le Corpus del Español del Siglo XXI (CORPES XXI ; <https://www.rae.es/banco-de-datos/corpes-xxi>), développé par la Real Academia de la Lengua Española, est la continuation du CREA, car il recueille des textes produits entre 2004 et 2015. Ce corpus contient près de 80 millions de mots issus de la presse (79 %), des livres (18 %) et du web (3 %). Seulement une minorité des mots provient de productions littéraires (8 %). Force est de constater que nous n'avons pas travaillé sur l'ensemble de cet énorme corpus car seulement 30 % du CORPES XXI provient de textes produits en Espagne. Dans notre recherche, nous avons commencé à utiliser la version 0.83 (juin 2016) et ultérieurement nous avons utilisé les versions du CORPES qui ont été publiées successivement (version 0.9 en juillet 2018 ; version 0.92 en mai 2020 ; version 0.93 en février 2021 ; version 0.94 en juillet 2021). Actuellement, il existe une toute dernière version (1.0), mais sa parution a été trop tardive pour notre recherche (juin 2023).

⁴⁵⁸ Sketch Engine (<https://www.sketchengine.eu/>) est une plateforme de gestion de corpus et un outil d'analyse de textes en ligne largement utilisé en lexicographie. Il a été développé par Lexical Computing Limited (une société de recherche fondée par Adam Kilgarriff en 2003). Sketch Engine prend en charge plus de 90 de langues. Il comprend des corpus en langue générale et du discours de spécialité. Nous avons utilisé les corpus *French Web 2020* (*fr TenTen20* ; 20.9 milliards de mots) et le corpus *Spanish Web 2018* (*es TenTen18* ; 16.9 milliards de mots). Il s'agit de deux corpus annotés de textes collectés du web appartenant à la famille TenTen (cf. Jakubíček et al., 2013). L'intérêt de cet outil tient à sa capacité à extraire de manière automatique les collocations sur la base de mesures statistiques précises et significatives. En outre, Sketch Engine permet de comparer la combinatoire de deux mots de la même langue (Word Sketch Difference). En fait, le postulat de base dans ces calculs est que plus la combinatoire de deux mots est proche, plus ils sont synonymes. Autrement dit, la synonymie est mesurée en termes de similarité de fonctionnement et de combinatoire.

d'adeptes au cours de ces deux dernières décennies. Ceux qui sont contre invoquent le caractère artificiel du discours dans le web, son caractère incorrect (on y trouve beaucoup de fautes et certaines sont très fréquentes), son caractère irrégulier (il y a des domaines qui sont beaucoup plus représentés que d'autres) et sa nature changeante. Par exemple, selon Sinclair (2004) : « The World Wide Web is not a corpus, because its dimensions are unknown and constantly changing, and because it has not been designed from a linguistic perspective ». Pourtant, nous considérons que ni le caractère « illimité » du web ni son caractère changeant ne constituent des entraves à l'utilisation du web en tant que corpus (cf. Berland et Grabar, 2002 ; Colson, 2007, 2010 ; Ferraresi et al., 2010 ; Fletcher, 2013). Le web comme corpus offre beaucoup plus d'avantages que de véritables inconvénients. Le développement du web a bouleversé la notion de corpus, qui, à notre avis, n'est plus aussi rigide qu'elle ne l'était auparavant. La toile fournit une énorme quantité de ressources, qui, avec certaines précautions, sont tout à fait valables et exploitables pour la recherche linguistique.

Nous avons donc décidé d'utiliser le web comme corpus. Si nous concevons le corpus tel qu'il est défini par Sinclair (2004) : « A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research », le corpus web répond à nos critères de recherche et constitue une vraie représentation de la langue. D'abord, comme il a déjà été mentionné, de nombreux domaines de spécialité sur lesquels nous avons travaillé sont sous-représentés dans les corpus généraux. Les occurrences de certains emplois fournies par ces corpus n'ont pas été suffisamment représentatives et parfois aucune occurrence n'a été fournie. Par contre, la toile constitue une énorme base de documents appartenant à de multiples domaines de spécialité. D'autre part, nous considérons que le web constitue un corpus très complet, car il renvoie une image approximative de l'ensemble des discours (cf. Cartier, 2011). Étant donné que l'objet de notre étude est en constante innovation (de nouveaux types d'allocations, de bourses, de crédits, etc. voient le jour de manière continue), le corpus le plus actualisé est sans aucun doute la toile.

En effet, de nombreux sites portent sur des sujets de spécialité dont notre nomenclature relève, ce qui nous permet d'y observer l'usage syntagmatique des unités lexicales qui nous intéressent. Les forums se sont avérés spécialement intéressants dans

le recensement des collocations verbales. Bien évidemment, nous avons été confrontée à des emplois erronés. Le travail de validation des collocations verbales a donc été lent et pénible : nous avons établi un socle à minima de 100 occurrences pour accepter une collocation verbale donnée. Force est de constater que le critère de fréquence n'est pas suffisant, car certaines unités lexicales faisant partie de notre nomenclature ont une fréquence extrêmement basse (p. ex., *étrennes*). Dans les cas où nous avons obtenu peu d'occurrences (en dessous du socle), nous avons été contrainte à observer le contexte dans lequel ces occurrences apparaissaient pour pouvoir vérifier si l'émetteur de ce texte maîtrisait la langue ou s'il commettait de nombreuses erreurs de combinatoire (il pourrait s'agir dans ce dernier cas d'un locuteur qui utilise le français ou l'espagnol sans pour autant le maîtriser). Dans le premier cas, nous avons considéré qu'il s'agit de cooccurrences peu fréquentes, mais correctes⁴⁵⁹ ; par contre, dans le deuxième, nous avons considéré qu'il s'agissait d'une erreur. Donc, les corpus de documents recueillis sur le web doivent être examinés avec soin, car leur qualité et fiabilité sont extrêmement diverses.

Nous avons utilisé le moteur de recherche commercial Google en raison du volume du corpus et de son potentiel d'interrogation (cf. Cartier, 2011). En effet, il permet de lancer des interrogations du type « recherche avancée », dans lesquelles il est possible de sélectionner la langue et le pays d'origine des sites web qui feront l'objet de la requête. Même si le moteur de recherche Google est accessible aux requêtes automatisées, nous avons tenu à repérer et vérifier la combinatoire des prédicats nominaux des <aides financières> de manière manuelle. En effet, nous adhérons à Blanco (2018) lorsqu'il soutient que

pour l'exploitation d'un corpus, aucun analyseur automatique, aucun extracteur de candidats-termes ne peut remplacer (ni du point de vue de la précision ni du point de vue du rappel) la lecture des textes par un expert, ni le repérage "manuel" des unités à traiter par la suite. (p. 22)

Le fait de fonder notre description de la langue sur ce vaste corpus a présenté de nombreux avantages méthodologiques. Premièrement, il nous a permis de vérifier de manière systématique les emplois des unités lexicales retenues dans le processus d'établissement de la nomenclature. Dans certains cas, l'observation du corpus a mis en

⁴⁵⁹ Ces cas échappent souvent à l'analyse dans des corpus « traditionnels » tels que Frantext, CREA, CORDE, voire Sketch Engine.

évidence des emplois ignorés par les dictionnaires. Deuxièmement, il nous a servi à identifier le schéma prédicatif propre à chaque prédicat, à partir de l'observation de l'environnement où apparaissaient les nombreuses occurrences de chaque emploi lexical. Étant donné que l'emploi prédicatif est défini, entre autres, par son schéma d'arguments et que chaque emploi constitue une entrée de notre dictionnaire, l'utilisation du corpus nous a aidée aussi à déterminer la liste d'éléments de la nomenclature. Troisièmement, le travail sur le corpus a permis de soulever la combinatoire restreinte en réduisant au strict minimum la part de la description tirée de notre intuition. Ainsi, nous avons testé (moyennant des recherches sur Google) tous les verbes supports sélectionnés par un nom prédicatif donné en le « croisant » avec d'autres noms prédicatifs susceptibles d'appartenir à la même classe sémantique. Le corpus web nous a donc permis de définir de manière exhaustive l'usage syntagmatique réel des différents emplois des mots.

2.3. Élaboration d'une grille d'analyse

Selon Grezka et Buvet (2007), « la nature d'une grille d'analyse est subordonnée à sa finalité » (p. 234). Notre finalité est de créer une description des noms prédicatifs des <aides financières> qui permette le traitement automatique. Rappelons que les deux grandes difficultés auxquelles le TAL se heurte constituent deux caractéristiques essentielles et massivement présentes dans les langues naturelles : le figement et la polysémie. Nous avons déjà abordé les spécificités du TAL (cf. Chap. V) : il a besoin d'une description exhaustive et systématique des faits de langue.

La tâche de recensement des noms prédicatifs des <aides financières> ayant été largement intuitive, nous avons obtenu une liste très hétérogène de noms. Dans le but de classer ces noms sur la base de leurs propriétés communes, nous avons élaboré une grille d'analyse que nous avons appliquée de manière systématique à tous les noms recensés⁴⁶⁰. Les propriétés qui constituent notre grille d'analyse ont été choisies par leur

⁴⁶⁰ Pour l'élaboration de cette grille d'analyse, nous nous sommes inspirés des travaux développés au sein du laboratoire LDI, comme Grezka (2006) ; Grezka et Buvet (2007) ; G. Gross (2008, 2017), Massoussi et Sfar (2009).

compatibilité avec les principes d'analyse du modèle des classes d'objets exposés ci-dessus. Ainsi, nous avons retenu quatre catégories générales de propriétés linguistiques.

Les propriétés configurationnelles concernent le schéma d'arguments des noms prédicatifs. Elles se subdivisent en propriétés structurelles (le nombre des arguments et leur mode de structuration) et en propriétés distributionnelles (la nature morphosyntaxique des arguments). Rappelons que pour une forme donnée il y a autant d'emplois différents qu'il y a de schémas d'arguments différents⁴⁶¹. La description des schémas d'arguments nous a permis de relever tous les emplois d'une même forme lexicale. Nous tenons à signaler que très souvent ces différents emplois ne sont pas décrits dans les articles des dictionnaires (par exemple, sous la forme de différentes acceptions).

Ces propriétés configurationnelles sont complétées par des propriétés sémantiques, car dans la caractérisation du schéma d'arguments on décrit aussi la nature sémantique de ceux-ci. Les propriétés sémantiques sont au centre du modèle des classes d'objets et c'est clairement ce qui distingue ce modèle de description de la langue du lexique-grammaire. Ainsi, chaque emploi est associé à une classe syntacticosémantique, à un type sémantique (ou trait sémantique) et à un domaine de spécialité. La spécification du domaine de spécialité s'est avérée spécialement intéressante, puisque la nomenclature de notre dictionnaire électronique comporte un grand nombre de termes appartenant au discours de spécialité.

Les propriétés de combinatoire servent à décrire les associations possibles des constituants sur l'axe syntagmatique. Elles se sont avérées très utiles pour désambigüiser les différents emplois d'une forme et la création de classes syntacticosémantiques plus fines. En effet, les deux premiers types de propriétés nous ont permis de faire une approximation à la définition des classes des noms prédicatifs des <aides>, mais ce premier groupement des emplois n'étant pas suffisant, nous l'avons affiné au moyen des verbes supports et des verbes prédicatifs appropriés sélectionnés par ces noms prédicatifs. Alors, nous avons relevé des associations de deux types : soit elles relèvent de la signification grammaticale (l'aspect et les marqueurs d'intensité, du mélioratif ou du péjoratif), soit elles relèvent de la signification

⁴⁶¹ Voir la section 3. *La description des prédicats : la notion d'emploi*, du Chap. IV.

lexicale⁴⁶². L'observation du corpus (notamment les requêtes sur Sketch Engine) ont mis en évidence la forte cooccurrence des noms des <aides financières> avec des verbes ayant un sens de 'demande' (p. ex., *demander, quémander, solliciter une aide*), de 'refus' (p. ex., *refuser / supprimer une aide*), de 'mise en place' (p. ex., *instaurer, mettre en place une aide*) et parfois de 'versement' (p. ex., *verser, payer*).

Enfin, nous avons tenu à observer des propriétés morphologiques spécialement pertinentes pour certains domaines de spécialité. Les noms de nombreuses aides sont souvent associés à un sigle ou une forme ayant souffert une troncation dont l'utilisation est très fréquente (p. ex., *RMI* ou *allocs*). Il est nécessaire donc d'indiquer ces formes associées et de les faire constituer des entrées indépendantes dans notre dictionnaire électronique.

3. Conclusions

Ces outils et méthodes d'analyse sont compatibles avec les principes d'analyse du modèle des classes d'objets. D'abord, tous les critères et les outils utilisés sont strictement linguistiques. Nous avons eu recours à des méthodes intuitives dans la première étape du recensement des termes, mais elles ont toujours été validées par la consultation d'ouvrages lexicographiques et l'observation du corpus. Deuxièmement, chaque emploi qui compose la nomenclature de notre dictionnaire est étudié et décrit dans le cadre de la phrase, comme l'exigent les propriétés configurationnelles (le schéma d'arguments) et les propriétés combinatoires de la grille d'analyse. Ces deux types de propriétés sont complétées par des propriétés sémantiques, ce qui garantit une approche syntacticosémantique du lexique. Quatrièmement, le fait de choisir un corpus web nous a permis de recenser et de décrire les noms prédicatifs des <aides> de façon bien plus exhaustive (en extension et en intension) qu'avec des corpus « traditionnels » (tels que Frantext ou CORPES), qui sont biaisés par nature. En plus, l'application de la grille d'analyse à tous les noms relevés répond à un souci de systématisme et d'exhaustivité. Par ailleurs, elle facilite le regroupement des emplois en classes et sous-classes syntacticosémantiques, c'est-à-dire des ensembles d'unités lexicales qui

⁴⁶² Il s'agit d'une prédication du second ordre.

fonctionnent globalement de la même manière. Enfin, ces classes et sous-classes constituent une taxonomie élaborée sur des bases strictement linguistiques et qui prend en compte l'analyse descendante.

CHAPITRE VII : DESCRIPTION DES PRÉDICATS NOMINAUX D'<AIDE FINANCIÈRE>

L'application systématique de la grille d'analyse aux noms prédicatifs d'<aide financière> recensés a conduit à la constitution de sous-classes sémantiques. Une sous-classe est constituée de différents noms prédicatifs synonymiques ou quasisynonymiques ayant des propriétés homogènes. Ainsi, nous avons obtenu une première classification en onze classes sémantiques. Dans un deuxième temps, quand nous avons décrit l'actualisation verbale de ces sous-classes sémantiques et nous les avons comparées, nous nous sommes rendu compte que certaines sous-classes avaient le même schéma d'arguments et la même combinatoire verbale. Ainsi, il ne nous a pas semblé pertinent de faire cette distinction en onze sous-classes, d'autant plus que la classe <aide financière> constitue déjà une sous-classe de l'hyper-classe <aide> (*soutien, concours, allocation, appui, assistance, coup de main, aide au développement...*). Nous montrons dans le tableau ci-dessous l'évolution du nombre des sous-classes :

Premier classement	Classement final
<aide financière : allocation ponctuelle> (<i>aide à la reprise ou à la création d'entreprise, aide hospitalière</i>)	<aide financière : allocation>
<aide financière : allocation régulière> (<i>allocation d'adoption, chômage</i>)	<aide financière : allègement>
<aide financière : allocation régulière : bourse d'enseignement> (<i>aide à la mobilité internationale, allocation de recherche, bourse</i>)	<aide financière : donation>
<aide financière : allocation viagère> (<i>pension, allocation de vieillesse agricole</i>)	
<aide financière : commission illicite> (<i>bakchich, enveloppe, dessous-de-table</i>)	
<aide financière : compensation> (<i>assurance accident, indemnisation</i>)	
<aide financière : décompte> (<i>abattement, dégrèvement</i>)	
<aide financière : donation >	

(<i>apport, don, étrenne, largesse</i>)
<aide financière : exonération> (<i>exemption, exonération</i>)
<aide financière : prime> (<i>complément, bonus, prime à l'exportation</i>)
<aide financière : subvention> (<i>aide à l'embauche dans les TPE, subvention d'équipement</i>)

Tableau 14 : *Évolution du nombre de sous-classes d'<aides financières>*

L'intérêt de la méthode des classes d'objets réside dans le caractère économique de ces classes : des unités lexicales ayant le même fonctionnement syntacticosémantique sont regroupées dans la même classe pour avoir une description de la langue qui soit économique et performante du point de vue du traitement automatique. Ainsi, nous avons considéré que la distinction entre <aide financière : allocation ponctuelle>, <aide financière : allocation régulière>, <aide financière : allocation régulière : bourse d'enseignement> et <aide financière : allocation viagère> n'était pas fonctionnelle. D'une part, la différence dans leur combinatoire verbale n'était pas significative. De plus, certaines de ces allocations ont changé la modalité de versement et elles ont passé d'être ponctuelles à s'étaler dans le temps pendant quelques mois. Par ailleurs, en Espagne, étant donné que certaines compétences administratives et fiscales ont été conférées aux régions (les *Comunidades autónomas*), une distinction aussi détaillée en petites sous-classes s'avère impraticable. Ces quatre sous-classes ont donc été regroupées dans la classe <aide financière : allocation>.

De plus, la comparaison des propriétés syntacticosémantiques qui caractérisent les classes nous a permis de relever certaines régularités. Les noms prédicatifs appartenant aux classes <aide financière : compensation>, <aide financière : subvention> et <aide financière : prime> et <aide financière : allocation> présentent le même schéma prédicatif et la même combinatoire verbale. Nous les avons donc toutes regroupées dans la classe <aide financière : allocation>. En effet, la distinction que nous avons faite était fondée sur des critères ontologiques. Pourtant, les classes d'objets ne procèdent pas d'une projection du monde sur la langue. Les classes et sous-classes sémantiques doivent être élaborées sur la base de critères strictement linguistiques, d'où l'intérêt de l'élaboration d'une grille d'analyse.

Pour cette raison même, les sous-classes <aide financière : décompte> et <aide financière : exonération> ont été regroupées dans une classe que nous avons appelée <aide financière : dégrèvement>.

En ce qui concerne la classe <aide financière : commission illicite>, elle regroupait des noms tels que *pot-de-vin*, *bakchich*, *arrosage*, *enveloppe*, *dessous-de-table*, etc. Même si le schéma d'arguments correspondait à celui des <aides financières> (N0:Nhum, Ncoll_hum / N1:à Nhum / N2:de N<montant> / N3:pour Vinf) et que des ouvrages lexicographiques tels que le *Petit Robert 2024 (version numérique)* établissent une relation de synonymie entre certains de ces noms et les noms *don*, *libéralité* ou *prime*, nous considérons que du point de vue sémantique une commission illicite n'est pas un type d'aide. Par conséquent, nous avons décidé de supprimer la classe <aide financière : commission illicite>.

Finalement, nous avons obtenu un classement en trois sous-classes : <aide financière : allocation>, <aide financière : donation et <aide financière : dégrèvement>. Nous les décrivons à continuation.

1. <aide financière : allocation>

Cette sous-classe comporte des noms simples (*pension*, *allocation*, *indemnité*, *subside*, etc.), des noms composés (*revenu de solidarité active*, *aide personnalisée au logement*, *rente d'incapacité permanente*, etc.) et des sigles (*APL*, *PPE*, *RSA*, etc.). Elle compte, comme nous venons de le montrer dans les exemples précédents, des noms de la langue générale, du discours de spécialité (notamment du domaine de la protection sociale et du droit du travail) et des noms appartenant à un registre de langue familier (p. ex., *allocs*).

Ces noms sont d'abord caractérisés par leurs propriétés configurationnelles, c'est-à-dire leur schéma d'arguments suivant :

N0	N1	N2	N3
hum, coll_hum	à hum, coll_hum	de <montant>	pour Vinf

Tableau 15 : Schéma d'arguments d'<aide financière : allocation> en français

N0	N1	N2	N3
----	----	----	----

hum, coll_hum	a hum, coll_hum	de <montant>	para Vinf
---------------	-----------------	--------------	-----------

Tableau 16 : Schéma d'arguments d'<aide financière : allocation> en espagnol

La position de premier argument est occupée par un nom caractérisé par le trait humain (*hum*) ou un collectif humain (*coll_hum*, p. ex. : un État, une institution, une organisation, etc.). La position N1 correspond au premier complément, c'est-à-dire la personne (*hum*) ou la collectivité (*coll_hum*) bénéficiaire de cette aide. Comme il s'agit d'aides financières, l'argument N2 diffère des autres sous-classes d'<aide>. Cette position argumentale est occupée par des groupes prépositionnels dénotant un montant d'argent⁴⁶³. Finalement, la position N3, bien qu'elle soit moins fréquemment actualisée, est occupée par un complément prépositionnel contenant un verbe en infinitif. Ce N3 indique le but ou la finalité de l'aide en question.

Nous tenons à signaler que certains noms d'<aides financières : allocations> comportent un cinquième argument, car il s'agit d'aides qui s'étalent sur une sur durée donnée. Leur schéma d'arguments serait donc le suivant :

N0	N1	N2	N3	N4
hum, coll_hum	à hum, coll_hum	de <montant>	pour, pendant, sur <période_tps>	pour Vinf

Tableau 17 : Schéma d'arguments d'<aide financière : allocation> avec 5 arguments en français

N0	N1	N2	N3	N4
hum, coll_hum	a hum, coll_hum	de <montant>	de, durante, por <période_tps>	para Vinf

Tableau 18 : Schéma d'arguments d'<aide financière : allocation> avec 5 arguments en espagnol

Nous montrons ici quelques exemples :

⁴⁶³ Par exemple, dans le cas de la classe <aide morale> (*soutien, secours, protection, assistance, appui*), le schéma d'arguments est le suivant : N0:Nhum, Ncoll_hum/N1:à Nhum/N2:pour Vinf # dans N. C'est-à-dire le N2 de la classe <aide morale> correspond au N3 de la classe <aide financière : allocation>.

- (1) *Sélectionnés sur des critères d'excellence, ces anciens élèves de lycées français du monde ont bénéficié d'une bourse pendant cinq ans afin de poursuivre leurs études supérieures en France.* (27/12/2019 ; <https://www.alfm.fr/fr/news/paris-ceremonie-des-boursiers-excellence-major-2014-2019-199259>)
- (2) *Par exemple quand un étudiant souhaite arrêter l'école pour travailler à la place d'un parent malade, PSE donne une aide financière pendant quelques mois afin que le jeune puisse continuer sa scolarité.* (16/04/2023 ; <https://www.pse.org/education/programmes-aider-familles-enfants-pris-charge>)
- (3) *La prestación económica consiste en un subsidio diario calculado en función de la base reguladora y el origen de la incapacidad, que se abonará durante los días naturales en que el interesado se encuentre en la situación de incapacidad temporal.* (16/04/2023 ; https://www.mites.gob.es/es/guia/texto/guia_14/contenidos/guia_14_29_3.htm)
- (4) *Mediante esta ayuda, los beneficiarios pueden disfrutar de una subvención del 100% de sus cuotas a la Seguridad Social durante los primeros 12 meses de afiliación al Régimen Especial de Trabajadores Autónomos (RETA).* (16/04/2023 ; <https://tomillo.org/actualidad/blog/>)

Concernant l'actualisation par des verbes supports, les noms prédicatifs de la sous-classe <aide financière : allocation> se combinent avec des verbes supports appropriés tels qu'*accorder, allouer, apporter, attribuer, donner, fournir, octroyer, offrir* et *procurer*, et le verbe support à variation diastématique *consentir*, appartenant au domaine de spécialité du droit et des finances :

- (5) *Si vous avez un enfant de moins de 20 ans qui est victime d'un accident, d'une maladie ou d'un handicap et que vous souhaitez réduire ou cesser votre activité, la CAF peut vous **accorder** une **Allocation Journalière de Présence Parentale (AJPP)**.* (02/04/2023 ; <https://www.aide-sociale.fr/aide-financiere-caf/>)

- (6) *Il est fréquent lorsque le président d'une société anonyme prend sa retraite que le conseil d'administration lui **alloue** une **pension de retraite complémentaire**.* (26/12/2020 ; <https://www.lesechos.fr/2001/01/nouvelle-jurisprudence-sur-le-complement-de-retraite-des-anciens-presidents-708513>)
- (7) *La deuxième décision a été de **consentir** une **subvention** de 300 euros en faveur des sinistrés du Liban après la catastrophe de Beyrouth.* (26/12/2020 ; https://www.larep.fr/combleux-45800/actualites/la-solidarite-des-agents-communiaux-recompensee_13835978/)

En espagnol, les verbes supports appropriés à la sous-classe <aide financière : allocation> sont *aportar, asignar, brindar, conceder, dar, ofrecer, otorgar, proporcionar* et *suministrar* :

- (8) *La UE le **ha brindado** una ayuda de 11.000 millones, mientras EEUU ha hecho lo propio con 1.000 millones.* (02/01/2021 ; <https://www.euroxpress.es/index.php/noticias?page=440>)
- (9) *Por ello, la OCDE dio la bienvenida a la decisión del Gobierno de **proporcionar** un **subsidio** de 420 euros a los parados de larga duración.* (02/01/2021 ; https://www.eldiario.es/canariasahora/economia/crisis-dejaramillones-parados_1_5435408.html)
- (10) *La Generalitat **otorga** una **prestación** de 625 euros anuales por hijo a las familias con niños menores de tres años.* (02/01/2021 ; <https://www.salud.mapfre.es/salud-familiar/mujer/el-parto/ayudas-sociales-post-parto/>)

Le verbe *prestar* présente une forte fréquence de cooccurrence avec certains noms d'<aide financière : allocation> :

- (11) *Una biografía revisa el papel político durante la dictadura de Salazar de la legendaria cantante, que **prestó apoyo económico** a los comunistas.* (02/01/2021 ; <https://elpais.com/noticias/amalia-rodriguez/>)
- (12) *Además, el profesor Velarde subrayó que el peso de España en la economía mundial es muy superior al de Grecia -país al que las autoridades europeas y el FMI tuvieron que **prestar** una **ayuda** de emergencia superior a*

los 100.000 millones de euros el pasado mes de mayo-, de manera que si se acentúa la depresión "España arrastraría también a las instituciones alemanas o inglesas que le dieron ayuda". (02/01/2021 ; https://cincodias.elpais.com/cincodias/2010/08/09/economia/1281489821_850215.html)

(13) *Es decir, el progenitor estará obligado a **prestar** la **pensión de alimentos** hasta que el hijo tenga independencia económica, aunque haya alcanzado la mayoría de edad.* (30/03/2023 ; <https://www.salamancahoy.es/salamanca/ciudad/mantener-hijos-justificado-pagar-pension-20230330174249-nt.html>)

(14) *Gimeno define la renta básica como un "sistema mucho más ágil, sencillo y eficiente de **prestar** una **renta mínima** a todas las personas que lo necesiten.* (03/01/2021 ; https://cadenaser.com/programa/2019/02/15/de_buenas_a_primeras/1550191104_989758.html)

Pourtant, nous n'avons pas retenu dans la liste de verbes supports le verbe *prestar*, car les recherches sur le corpus de la combinaison de ce verbe avec certains noms d'aides tels que *subsidio*, *subvención*, *renta* ou *prestación* ont donné un nombre d'occurrences au-dessous du seuil de fiabilité. Par ailleurs, *beca* ne sélectionne pas *prestar* comme verbe support.

Les noms prédicatifs d'<aide financière : allocation> peuvent également apparaître dans des constructions converses à l'aide de verbes supports passifs tels qu'*avoir*, *avoir droit à*, *bénéficier de*, *disposer de*, *encaisser*, *jouir de*, *percevoir*, *profiter de* et *recevoir*, et la variante diastratique du registre de langue familier *toucher* :

(15) *L'employeur qui embauche un salarié en CUI-CIE **bénéficie** d'une **aide** à l'**insertion professionnelle**.* (27/06/2021 ; <https://www.cadremploi.fr/editorial/conseils/droit-du-travail/quest-ce-quun-contrat-aide>)

(16) *Fin 2017 j'étais en arrêt maladie puis en janvier 2019 j'ai été déclaré inapte par la médecine du travail et licenciée en mars 2019 j'ai été depuis au*

chômage ARE jusqu'en juin 2021 et actuellement je touche l'Allocation Spécifique de solidarité. (18/08/2022 ; <https://forum-assures.ameli.fr/questions/2983794-pension-invalidite-touchant-allocation-specifique-solidarite-accord-demande-aah>)

- (17) *Si un institut ou centre de catégorie 2 cesse de recevoir l'appui financier d'un ou de plusieurs États membres parrainants ou de toute autre source de financement, le Directeur général invite l'État membre ou les États membres parrainants à étudier d'autres possibilités de financement dans un délai de six mois.* (18/08/2022 ; https://ich.unesco.org/doc/src/37-C-Resolution_93_FR.pdf)

En espagnol, nous avons recensé les verbes supports passifs *contar con*, *disfrutar de*, *gozar de*, *percibir*, *recibir*, *tener* et *tener derecho a* :

- (18) *Además, las familias pueden contar con la ayuda a la escolaridad de la Comunidad de Madrid (cheque guardería).* (23/08/2022 ; <https://ludus.org.es/es/alegra-centro-libre>)
- (19) *El Real Madrid sigue sin reconocer que buena parte de su supervivencia en posguerra se debe a que en el Madrid republicano un grupo de 'rojos' le mantuvo latente y el Barcelona no desea recordar que aún en los últimos meses del franquismo seguía gozando del apoyo económico del régimen.* (23/08/2022 ; <https://www.marca.com/blogs/tirandoadar/2016/10/06/la-patetica-venganza-contra-samaranch.html>)
- (20) *Tienen derecho a la renta de ciudadanía en los términos y condiciones previstos en esta ley: a) Quienes tengan la condición de ciudadanos de La Rioja, de conformidad con lo establecido en el Estatuto de Autonomía. b) Las personas extranjeras con vecindad administrativa en La Rioja, en los términos previstos en la legislación aplicable.* (21/11/2022 ; <https://www.boe.es/buscar/act.php?id=BOE-A-2017-5627>)

En ce qui concerne l'aspect, les noms prédicatifs de cette classe sont ponctuels. Néanmoins, ils peuvent se combiner avec certains verbes supports aspectuels à valeur inchoative, progressive, terminative, itérative et intensive. Ainsi, des verbes tels que *décrocher*, *gagner*, *obtenir* et *trouver* portent une valeur aspectuelle inchoative :

- (21) *La Chaire COAST de MBS décroche une subvention de recherche d'une valeur de 406 000 euros dans le cadre du projet Horizon Europe ENFASYS.* (20/08/2022 ; <https://www.montpellier-bs.com/?p=82880>)
- (22) *Une épouse qui s'était consacrée à son mari et à ses enfants, dont l'un était malade, a obtenu une prestation de 30 000€.* (18/08/2022 ; <https://cabinet-avocat-daude.fr/lappreciation-de-la-disparite-creee-par-le-divorce-dans-le-cadre-dune-prestation-compensatoire/>)
- (23) *Tous les pays européens essaieront, tour à tour, de trouver un secours financier auprès d'une Union européenne très affaiblie en passe d'être rapidement déstabilisée.* (17/08/2022 ; https://www.contrepoints.org/2013/05/21/124976-leurope-remede-miracle-et-bouc-emissaire#comments_container)

En espagnol, nous avons repéré les verbes supports aspectuels à valeur inchoative *conseguir, ganar, lograr* et *obtener* :

- (24) *Para conseguir la pensión máxima de jubilación que está por encima de los 3.000 euros mensuales es necesario tener 36 años y 6 meses cotizados a la Seguridad Social y que en los últimos 25 años se haya cotizado por la base máxima de cotización. Las bases máximas para este año son 4.495,50 euros mensuales.* (18/08/2022 ; https://www.elconfidencial.com/economia/2023-03-30/requisitos-cobrar-pension-maxima-jubilacion_3601923/)
- (25) *Para ello, Asaenes ha puesto en marcha esta iniciativa de recaudación de fondos con el fin de lograr el apoyo económico que permita hacer realidad este proyecto de integración y mejora de la calidad de vida de las personas con trastorno mental y sus familiares.* (18/08/2022 ; <https://consaludmental.org/sala-prensa/asaenes-crowdfunding-nuevo-espacio/>)
- (26) *Un equipo australiano ha ganado una subvención de 600.000 dólares para mezclar la IA con células cerebrales humanas.* (30/08/2023 ; <https://www.muyinteresante.es/actualidad/61025.html>)

La combinatoire en français avec des verbes supports aspectuels à valeur progressive est bien plus restreinte que celle des verbes supports aspectuels inchoatifs, car nous n'avons repéré qu'un verbe qui puisse se combiner avec tous les noms appartenant à la sous-classe d'<aide financière : allocation>. Il s'agit du verbe *maintenir* :

- (27) *On peut supposer que lorsque les héritiers sont en même temps les enfants de la créancière, ils font le choix de **maintenir** la **prestation** versée à leur mère.* (27/07/2023 ; https://www.justice.gouv.fr/sites/default/files/migrations/portail/art_pix/rapp-ort-prest-compens.pdf)
- (28) *Par ailleurs, **maintenir** une **rente dorée** à un industriel privé, propriétaire d'un journal, Le Figaro, au détriment de l'entreprise publique EADS que financent les contribuables, constitue, semble-t-il, une aberration qu'il sera difficile à supporter jusqu'en 2030.* (27/07/2023 ; https://www.lemonde.fr/idees/article/2012/01/27/une-niche-particuliere_1635556_3232.html)
- (29) *L'État **maintient** le **soutien financier** apporté aux entreprises pendant la crise du Covid-19.* (27/07/2023 ; <https://www.rejoindre-plus-que-pro.fr/soutien-etat-entreprises-covid-19/>)

Le verbe *prolonger* présente une combinatoire assez ample, comme le montrent les exemples ci-dessous :

- (30) *La Caisse nationale d'allocations familiales (Cnaf) a décidé ce mercredi 3 juin de **prolonger** l'**aide financière exceptionnelle** versée aux crèches, encore limitées dans l'accueil des enfants à cause du coronavirus, et d'attribuer une aide supplémentaire aux structures qui rouvrent afin de compenser les surcoûts en équipements sanitaires.* (30/08/2023 ; <https://www.ouest-france.fr/sante/virus/coronavirus/deconfinement-une-aide-financiere-supplementaire-versee-aux-creches-6857033>)
- (31) *Compte tenu du renforcement du protocole sanitaire et face au rebond de l'épidémie, AMELI a décidé de **prolonger** la **subvention** « **Prévention COVID** » afin de continuer à aider les TPE et PME pour protéger la santé*

des salariés et lutter contre la propagation du virus. (30/08/2023 ; <https://www.caheb.fr/actualites/subvention-prevention-covid-les-tpe-pme-et-travailleurs-independants-peuvent-faire-leur-demande-de-subvention-des-le-15-octobre-pgkxqr>)

Pourtant, nous n'avons trouvé aucune occurrence de la combinaison du verbe *prolonger* avec des noms tels que *pension*, *rente* ou *subside*. De plus, la combinaison avec certains noms (p. ex., *subvention*) présentait un nombre d'occurrences au-dessous du seuil de fiabilité.

L'aspect progressif présente une combinatoire verbale faiblement plus ample dans des constructions converses dans lesquelles l'argument N1 occupe la position de sujet et l'argument N0 disparaît. Ainsi, nous avons repéré les verbes supports *conserver*, *garder* et *maintenir* :

(32) *La veuve, en effet, bien souvent ne se remarie pas et vit maritalement pour **conserver** la **pension**.* (27/07/2023 ; https://www.lemonde.fr/archives/article/1953/01/01/la-pension-des-veuves-de-guerre_1970403_1819218.html)

(33) *Alors que 58% d'entre eux déclarent vouloir **garder** le **soutien économique** tel quel, 30% estiment qu'il est insuffisant et souhaitent accroître ces **aides**.* (27/07/2023 ; <https://www.radioclassique.fr/societe/quoi-quil-en-coute-7-francais-sur-10-approuvent/>)

(34) *J'aimerais que quelqu'un m'explique s'il est possible de **maintenir** mes **allocations chômage** si je pars vivre en Angleterre.* (27/07/2023 ; <https://forums.commentcamarche.net/forum/affich-12438608-maintiens-du-chomage-en-angleterre>)

En revanche, en espagnol la différence entre le nombre de verbes supports aspectuels à valeur progressive qui entrent dans des constructions N0 N1 et ceux qui entrent dans des constructions converses (N1 N0) n'est pas significative. Ainsi, pour le premier cas de figure, nous avons repéré les verbes *mantener*, *prolongar*, *prorrogar* et *renovar* :

- (35) *Calviño se abre a **prorrogar** la ayuda al transporte local y autonómico tras el 30 de junio.* (27/07/2023 ; <https://www.elperiodico.com/es/economia/20230621/calvino-abre-prorrogar-ayuda-transporte-publico-88988685>)
- (36) *La Unión Europea ha **prolongado** las ayudas al carbón hasta 2018.* (27/07/2023 ; <https://www.rtve.es/television/20110107/escarabajo-verde-sin-carbon/392739.shtml>)
- (37) *Asimismo, ha anunciado que no se van a **renovar** las subvenciones a personas físicas para el fomento de las condiciones de seguridad y de mejora medioambientales, lo cual supone un ahorro de 12,5 millones de euros, y que se cancela la subvención nominativa a sindicatos para la sociedad de la información.* (26/07/2023 ; <https://www.europapress.es/economia/noticia-cospedal-suspendera-construccion-vivienda-protegida-20111202170957.html>)

Dans le cas des constructions converses, nous avons repéré les verbes *conservar* et *mantener* :

- (38) *Denegación, anulación o suspensión cuando el beneficiario hubiera actuado fraudulentamente para obtener o **conservar** la prestación* (27/07/2023 ; https://www.mites.gob.es/es/guia/texto/guia_14/contenidos/guia_14_29_4.htm)
- (39) *El SEPE ahora permite a los mayores de 52 años **mantener** el subsidio por desempleo aunque consigan trabajo* (20/09/2023 ; <https://www.epe.es/es/activos/20230911/sepe-cobrar-subsidio-desempleo-trabajo-91950618>)

Par ailleurs, des verbes supports tels qu'*annuler*, *couper*, *geler*, *retirer*, *supprimer* et *suspendre* expriment l'aspect terminatif :

- (40) *À gauche, certains candidats rêvent déjà de renationaliser non seulement EDF et l'ex-GDF, mais aussi les autoroutes pour **annuler** la rente concédée il y a vingt ans par la droite aux Vinci et autres Eiffage.*

(25/06/2022 ; https://www.liberation.fr/politique/elections/campagne-presidentielle-retraites-salaires-impots-relocalisations-revenu-universel-le-clivage-gauche-droite-plus-que-jamais-dactualite-20220406_IRLXN2Z3EVFVLB6HFJRAF5VPCU/)

(41) *L'Union européenne va **couper** l'aide au développement aux pays émergents.* (25/06/2022 ; https://www.lemonde.fr/europe/article/2011/12/07/l-ue-va-couper-l-aide-au-developpement-aux-pays-emergents_1614376_3214.html)

(42) *Le Vatican prend des sanctions contre Amnesty International. Il a décidé d'appeler tous les catholiques à **suspendre** leur soutien financier à l'association.* (25/06/2022 ; <https://www.20minutes.fr/monde/164241-20070613-vatican-accuse-amnesty-international-detre-pro-avortement>)

En espagnol, les verbes supports aspectuels à valeur terminative se combinant avec tous les substantifs de la classe <aide financière : allocation> sont *anular, cancelar, congelar, eliminar, quitar, retirar, suprimir* et *suspender* :

(43) *El Govern ha vuelto a **congelar** la Renta Garantizada de Ciudadanía (RGC) en los presupuestos de 2022.* (25/06/2022 ; https://cronicaglobal.elespanol.com/politica/20211112/los-presupuestos-govern-congelan-renta-garantizada-ciudadania/626687358_0.html)

(44) *Dos años más tarde, el Servicio Público de Empleo Estatal (SEPE) le **quitó** el subsidio para mayores de 55 años.* (25/06/2022 ; https://www.infolibre.es/economia/automatismos-ciegos-empleo-mujer-gano-premio-tv-figurantes-ocho-apellidos-catalanes_1_1122968.html)

(45) *El Instituto de la Mujer **ha suprimido** las subvenciones a las universidades para la realización de postgrados de género y actividades relacionadas con la igualdad.* (25/06/2022 ; https://cadenaser.com/ser/2017/01/13/sociedad/1484305544_826511.html)

De plus, en espagnol et en français, nous avons repéré le verbe support aspectuel à valeur terminative *perdre/perder* qui entre dans des constructions passives dans lesquelles le N1 occupe la position de sujet et le N0 soit il disparaît, soit il apparaît en

position de complément, dans un syntagme prépositionnel commençant par la préposition *de* :

- (46) *Malgré ses titres de championne d'Europe juniors 2009-2011-2013, elle a en effet perdu l'appui financier de sa marque en 2012.* (25/06/2022 ; https://www.lepoint.fr/sport/surf-johanne-defay-une-pepite-reunionnaise-sans-sponsor-04-10-2015-1970620_26.php)
- (47) *Algunas mujeres han perdido la pensión compensatoria que habían obtenido tras su divorcio* (25/06/2022 ; <https://aba-abogadas.com/se-puede-perder-el-derecho-a-pension-compensatoria-al-rehacer-la-vida-sentimental/>)

En ce qui concerne la valeur aspectuelle itérative, nous avons repéré les verbes *débloquer*, *reconduire* et *renouveler* :

- (48) *Le séisme survenu mercredi peut-il être un argument pour que la communauté internationale débloque l'aide au développement ?* (07/07/2022 ; https://www.liberation.fr/international/moyen-orient/en-afghanistan-la-crise-va-saggraver-avec-des-consequences-devastatrices-pour-la-survie-des-familles-20220623_XCFDD2LMGZHQJO7XKUCXCUEVYY/)
- (49) *Le conseil propose de reconduire les allocations versées aux 5 adultes pour un montant de 62 427 euros ainsi qu'aux 7 enfants pour un montant de 26 160 euros toujours présents dans la circonscription.* (07/07/2022 ; https://boston.consulfrance.org/IMG/pdf/pv_cc_08122022_site_internet.pdf?8791/d6ccd4a76b259be0d43bedd8a040c79b1363e1f1)
- (50) *Malgré la flambée des prix de l'essence durant l'été, le gouvernement n'a pas prévu de renouveler l'indemnité carburant attribuée à certains ménages français en début d'année.* (30/08/2023 ; <https://www.linternaute.com/auto/conduite/3434681-elisabeth-borne-dit-non-au-retour-des-cheques-carburants/>)

En espagnol, nous avons repéré aussi les verbes *desbloquear*, *reconducir* et *renovar* :

- (51) *Los países de la UE logran **desbloquear** la **ayuda** de 18.000 millones a Ucrania.* (01/09/2023 ; https://www.eldiario.es/internacional/paises-ue-logran-desbloquear-ayuda-18-000-millones-ucrania_1_9783508.html)
- (52) *Por su parte Juan Carlos Pedrosa señalaba que ha negociado con la directora general de Atención Social, Pilar Albert, para **reconducir** la **subvención** de 300.000 euros que la Generalitat concedió al centro.* (01/09/2023 ; <https://elperiodicodevillena.com/el-nuevo-centro-de-dia-de-alzheimer-atendera-a-80-pacientes-de-villena-y-comarca/>)
- (53) *Una vez se ha agotado el paro y el SEPE no te **renueva** el **subsidio por desempleo** al no reunir los requisitos exigidos, existen estas ayudas.* (01/09/2023 ; <https://www.eldia.es/economia/2023/06/08/ayudas-cobrar-52-anos-acaba-paro-dv-88473359.html>)

Enfin, les verbes supports aspectuels intensifs peuvent porter une valeur progressive-augmentative ou progressive-atténuative. En français, la valeur progressive-augmentative est exprimée au moyen de verbes tels qu'*améliorer*, *augmenter* et *majorer* :

- (54) *L'exécutif a promis d'**améliorer** les **pensions** des retraités les plus modestes.* (10/09/2023 ; <https://www.lesechos.fr/economie-france/social/petites-pensions-le-coup-de-pouce-va-se-faire-attendre-pour-nombre-de-retraites-1974361>)
- (55) *Tout au long de ce jour de grève, La Région et Île-de-France Mobilités proposent la gratuité du covoiturage pour les Franciliens et **augmente** la **subvention** accordée aux conducteurs.* (09/09/2023 ; <https://www.iledefrance.fr/rechercher?keys=subvention%20velo&wrapper=format=html&page=0>)
- (56) *Le gouvernement décide de **majorer** les **prestations familiales** à partir du 1^{er} avril.* (09/09/2023 ; <https://bibnum.sciencespo.fr/s/catalogue/ark:/46513/scq0rvh>)

Nous tenons à signaler que nous avons repéré d'autres verbes supports aspectuels intensifs à valeur progressive-augmentative tels qu'*optimiser*, *relever*,

revaloriser, réviser, réviser à la hausse et revoir à la hausse, mais ils ne se combinent pas avec tous les noms de la classe <aides financières : allocation>. Nous avons observé que parfois ces verbes, si habituels dans les discours de spécialité du domaine des finances ou de la protection sociale, ne se combinent pas avec des noms appartenant au discours général tels qu'*appui, secours* ou *soutien*. Pourtant, nous n'avons pas créé une classe indépendante car ces noms n'ont pas un comportement combinatoire homogène avec tous les verbes du discours de spécialité signalés plus haut. Par exemple, *revaloriser* se combine avec *soutien* :

- (57) *Pour la danse, il s'agit tout d'abord de revaloriser le soutien financier aux compagnies conventionnées.* (09/09/2023 ; <https://www.culture.gouv.fr/Presse/Archives-Presses/Archives-Discours-2012-2018/Annee-2016/Discours-d-Audrey-Azoulay-ministre-de-la-Culture-et-de-la-Communication-prononce-a-l-occasion-de-la-presentation-des-mesures-destinees-a-favoriser>)

Par contre, nous n'avons pas relevé de cooccurrences **revaloriser* (*appui, secours*) ou **soutien* (*optimiser, relever, revoir à la hausse*).

En espagnol, la combinatoire avec des verbes supports aspectuels intensifs à valeur progressive-augmentative s'est avérée plus prolifique qu'en français. En effet, nous avons relevé des verbes tels qu'*ampliar, aumentar, elevar, incrementar, mejorar* et *subir* :

- (58) *La respuesta del Ayuntamiento fue clara y les explicaron que ahora mismo, con los presupuestos cerrados, no pueden ampliar la ayuda al Zamora* CF. (11/09/2023 ; <https://mas.laopiniondezamora.es/canales/zamora-cf/content/view/2226/99/>)
- (59) *Nuestro gobierno ha elevado las pensiones de los discapacitados, las pensiones mínimas.* (11/09/2023 ; https://elpais.com/internacional/2012/11/18/actualidad/1353259053_632567.html)
- (60) *De hecho, [el Gobierno de José Luis Rodríguez Zapatero] subió la subvención en más de 300.000 euros entre 2007 y 2008, para estabilizarla desde entonces y hasta 2011 en los citados 15,79 millones de*

euros. (11/09/2023 ; https://www.elconfidencial.com/economia/2013-08-06/rajoy-recorta-a-la-mitad-la-subsencion-a-ugt-y-ccoo-para-sufragar-su-actividad_15632/)

Quant aux verbes supports aspectuels intensifs à valeur progressive-atténuative en français, nous n'avons relevé que trois verbes se combinant avec tous les noms de la classe : *diminuer*, *limiter* et *réduire*. Voici quelques exemples :

(61) *On parle de la nouvelle grille des aides financières du Département aux clubs sportifs , qui doit profiter à toutes les associations, mais implique de **diminuer** les **subsidés** aux « gros », surtout le quatuor BBD, TFC, BPFC et USB.* (14/08/2023 ; <https://www.sudouest.fr/sport/basket/boulazac-basket-dordogne/subventions-aux-clubs-sportifs-perigourdins-il-fallait-reequilibrer-4767010.php>)

(62) *Globalement, les différentes réformes législatives successives en rapport avec le sport professionnel ont eu pour objectif de **limiter** l'**appui financier** apporté par les collectivités.* (14/08/2023 ; <https://www.ccomptes.fr/fr/documents/8990>)

(63) *Pour satisfaire les demandes des institutions financières internationales, les autorités **ont réduit** les **subventions** étatiques et adopté un régime de change flottant, qui a entraîné un bond de l'inflation.* (14/08/2023 ; <https://www.ouest-france.fr/economie/soudan-le-premier-ministre-espere-se-delester-de-la-dette-de-60-milliards-de-dollars-7259490>)

Comme dans le cas des verbes progressifs-augmentatifs, nous avons repéré de nombreux verbes supports intensifs à valeur progressive-atténuative (*baisser*, *minorer*, *plafonner* et *raboter*), mais leur combinatoire avec des noms tels qu'*appui*, *secours* et *soutien* est très irrégulière. D'autres verbes tels que *réviser à la baisse* et *revoir à la baisse* ont une combinatoire beaucoup plus restreinte : *réviser à la baisse* (*aide*, *allocation*, *indemnité*, *prestation*) ; *revoir à la baisse* (*aide*, *allocation*, *indemnité*, *pension*, *prestation*).

En espagnol, la liste de verbes supports aspectuels intensifs à valeur progressive-atténuative se combinant avec des noms prédicatifs de la classe <aide financière :

allocation> est plus ample : *bajar, congelar, cortar, disminuir, rebajar, recortar et reducir*. Nous notons ci-dessous quelques exemples :

(64) *El equipo de gobierno ha aprobado congelar las tasas en 2010, a pesar de que dio luz verde a una subida en el alcantarillado y **bajó** las **subvenciones** destinadas a los cursos de la kultur etxea.* (25/05/2022 ; <https://www.deia.eus/bizkaia/2009/11/15/gorliz-congela-tasas-sube-alcantarillado-5699378.html>)

(65) *La Confederación Asperger España se una al Comité Español de Representantes de Personas con Discapacidad (CERMI) que ha instado al Ministerio de Derechos Sociales y Agenda 2030 a rectificar de inmediato la decisión de **recortar** el **apoyo económico** público al Tercer Sector de Acción Social de ámbito estatal, en el momento en que el país y la sociedad española más necesitan a las redes solidarias como consecuencia de la pandemia.* (25/05/2022 ; https://www.asperger.es/tercer_sector_denuncia_ministerio_derechos_sociales_recortes_2020.html)

(66) *España es el país miembro de la Unión Europea que más **redujo** las **ayudas** públicas al sector pesquero (un 6,1 por ciento), en el período comprendido entre 1994 y 1998, según revela un informe del Ejecutivo comunitario.* (25/05/2022 ; <https://www.lavozdegalicia.es/temas/union-europea/3366>)

Enfin, nous avons observé que les noms prédicatifs de la classe <aides financières : allocations> se combinent souvent avec des verbes prédicatifs que nous avons classés comme verbes de <demande>, de <mise en place>, de <refus> et de <versement>. En ce qui concerne les verbes de <demande>, nous tenons à signaler qu'ils apparaissent dans des constructions converses (N1N0). Nous avons repéré les verbes *demander, réclamer et solliciter* :

(67) *J'ai **demandé** une **subvention** au conseil général qui m'a accordé 148 000 euros.* (03/06/2022 ; https://actu.fr/ile-de-france/melun_77288/gymnase-fermeture-ou-travaux-durgence_7036310.html)

(68) *Entre 1.000 et 2.000 artisans manifestent à Paris pour **réclamer des aides** au gouvernement.* (28/08/2023 ;

<https://www.francebleu.fr/infos/economie-social/prix-de-l-energie-entre-1-000-et-2-000-artisans-manifestent-a-paris-pour-reclamer-des-aides-au-gouvernement-7200509>)

(69) *En application de l'article L. 431-2 du code de la sécurité sociale, les ayants droits disposent d'un délai de deux ans pour **solliciter une rente de survivant.*** (03/06/2022 ;

<https://www.courdecassation.fr/en/decision/5fca6c50668ca6570fe223fd>)

Le travail sur le corpus s'est avéré d'une importance capitale, car il nous a permis d'observer que certains verbes apparaissant dans les ouvrages lexicographiques comme des synonymes (p. ex., *solliciter* – *quémander*) ont une combinatoire distincte : *solliciter* (*une aide, une allocation, une prestation, une rente, un subsidie*) ; *quémander* (*une aide, une allocation, *une prestation, *une rente, un subsidie*). Tel est le cas, aussi du verbe *exiger*, quasi-synonyme de *demander* (*exiger* ayant une valeur plus injonctive ou impérative) : *demander une prestation* ; **exiger une prestation*.

En espagnol, nous avons retenu les verbes de <demande> *demandar, exigir, pedir, reclamar* et *solicitar* :

(70) *El cónyuge puede **demandar la pensión alimenticia** al cónyuge en caso de separación legal, o al ex cónyuge en caso de divorcio, si el primero no es capaz de hacer frente a sus necesidades por razones ajenas a su voluntad (artículo 4:29, apartado 1, del Código Civil).* (10/06/2022 ; https://e-justice.europa.eu/content_maintenance_claims-47-hu-es.do?init=true&member=1)

(71) *Los pensionistas vuelven a **exigir** en Bilbao una **prestación mínima** de 1.080 euros.* (13/09/2022 ;

<https://www.elcorreo.com/economia/pensiones/pensionistas-vuelven-exigir-20220823211910-nt.html>)

(72) *El retraso en la solicitud de la pensión no significa su pérdida para el potencial beneficiario, puesto que la condición de viudo permanece varios años después del fallecimiento del cónyuge, aunque todavía no **haya***

reclamado la pensión (y lo mismo sucede con la pensión de jubilación).
(28/08/2023 ; <https://dialnet.unirioja.es/descarga/articulo/8559435.pdf>)

Nous avons repéré de nombreux verbes de <mise en place> se combinant avec des noms d'<aide financière : allocation> : *créer, instaurer, instituer, introduire* et *voter*. Pourtant, seulement les verbes *introduire, mettre en place* et *voter* se combinent avec tous les noms de la classe :

(73) *En France, le régime de Vichy en 1941 introduit l'allocation aux vieux travailleurs et le minimum vieillesse, mais ce sera finalement après la libération, le 26 avril 1946, que sera créé plus officiellement le système de retraite par répartition, puis les régimes de base, que nous connaissons.*
(14/09/2022 ;

<https://www.opes.fr/retraite/#:~:text=En%20France%2C%20le%20r%C3%A9gime%20de,de%20base%2C%20que%20nous%20connaissons.>)

(74) *Depuis 2016, les entreprises peuvent mettre en place une indemnité kilométrique vélo (IK vélo).* (14/09/2022 ;
<https://www.europe1.fr/economie/venir-au-travail-a-velo-pourquoi-les-entreprises-devraient-encourager-leurs-salaries-a-pedaler-3647867>)

(75) *La région Île-de-France vote l'aide aux départs en vacances.*
(14/09/2022 ; https://www.lepoint.fr/societe/la-region-ile-de-france-vote-l-aide-aux-departs-en-vacances-09-03-2017-2110716_23.php)

En effet, les recherches menées sur le corpus nous ont permis d'observer, par exemple, que le verbe *instituer* ne se combine pas avec des noms tels qu'*appui, soutien* ou *subside*, que le verbe *instaurer* ne se combine pas avec *appui*, ou que le verbe *créer* ne se combine pas avec des noms tels qu'*appui, pension* ou *secours*. Comme nous l'avons déjà signalé (cf. verbes supports aspectuels intensifs à valeur progressive-augmentative), ces noms appartenant au discours général présentent une combinatoire plus restreinte avec des verbes appropriés du discours de spécialité (notamment du domaine de la protection sociale ou des finances).

Par contre, en espagnol nous n'avons pas observé cette sorte de différences de combinatoire entre les noms d'<aide financière : allocation> appartenant à un discours

général ou spécialisé. En effet, nous avons repéré les verbes *aprobar*, *establecer*, *instituir* et *poner en marcha*, qui se combinent avec tous les noms de la classe.

(76) *Recientemente acabamos de **aprobar** un **apoyo** extraordinario de 100 millones de euros para la resiliencia del sector turístico en las Islas Canarias, 5 millones de euros del plan de sostenibilidad para las Islas o para la isla de Lanzarote...* (15/09/2022 ; <https://www.lamoncloa.gob.es/presidente/intervenciones/Paginas/2021/prsp20210811.aspx>)

(77) *Con el objetivo de brindar apoyo económico a este grupo demográfico, el Gobierno de España **ha establecido** el **subsidio para mayores de 52 años**.* (15/09/2022 ; <https://asesora10.com/subsidio-para-mayores-de-52-anos/>)

(78) *En esta ocasión muy poco se puede esperar de un gobierno débil, incapaz de derogar la reforma laboral pese a lo prometido, ineficaz para **instituir** una **renta básica universal** e inútil a la hora de defender a las clases más débiles.* (15/09/2022 ; <https://portaldeandalucia.org/opinion/el-sindicalismo-necesario/>)

(79) *En lado opuesto a la CEOE, el secretario general de UGT, Josep Maria Álvarez, valoró que el Gobierno finalmente **ponga en marcha** la **renta mínima** y recordó que “la emergencia requiere agilidad y atender las demandas perentorias de la ciudadanía”.* (15/09/2022 ; <https://www.lavanguardia.com/economia/20200411/48407808901/el-gobierno-acelera-la-renta-minima-de-430-euros-durante-tres-meses.html>)

Quant aux verbes prédicatifs appropriés de <refus>, nous avons constaté que le verbe *refuser* apparaît dans des constructions NON1 et dans des constructions N1N0 (constructions converses) :

(80) *Son parti souhaite **refuser** le **revenu minimum d'insertion** aux immigrants lors de leurs dix premières années aux Pays-Bas.* (20/09/2022 ; https://www.la-croix.com/Actualite/Monde/La-crise-domine-la-campagne-electorale-aux-Pays-Bas-_NG_-2010-06-07-552683)

(81) *Le CIMADE a donné l'exemple d'une organisation au Soudan ayant pris la décision de **refuser l'appui financier** car elle ne souhaitait pas être contrainte par les exigences administratives des bailleurs de fonds du Nord.* (20/09/2022 ; <https://www.intrac.org/wpcms/wp-content/uploads/2018/11/Praxis-Paper-1-French.pdf>)

Dans le cas de constructions converses (N1N0), nous avons obtenu un grand nombre d'occurrences faisant partie de constructions passives avec *se voir* + *infinitif* :

(82) *M. X... s'est vu **refuser le revenu minimum d'insertion** et s'est trouvé dans une situation de grande précarité.* (20/09/2022 ; <https://www.conseil-etat.fr/CE/CJAS/boc2015/ccas3/a0030016.htm>)

(83) *De nombreuses personnes en situation de handicap présentent un taux d'incapacité permanente entre 50 et 79%, mais **se voient refuser l'allocation adulte handicapé.*** (20/09/2022 ; <https://www.philippe-avocat.fr/fr/actualites/id-115-restriction-allocation-adulte-handicape>)

(84) *La requérante, résidente espagnole appartenant à la Communauté Rom, s'est vu **refuser une pension de réversion** au décès de son mari au motif que, mariée selon le rite propre à cette communauté, elle n'était pas civilement la conjointe du défunt.* (20/09/2022 ; https://juridique.defenseurdesdroits.fr/index.php?lvl=notice_display&id=6165)

À part le verbe *refuser*, nous avons relevé aussi que le verbe *renoncer* à, apparaissant dans des constructions passives, se combine avec tous les noms de la classe <aide financière : allocation> :

(85) *Est-il possible de **renoncer à une rente à vie** pour accident de travail ?* (27/12/2022 ; <https://forum-assures.ameli.fr/questions/3031003-possible-renoncer-rente-vie-accident-travail>)

(86) *Ma collectivité souhaite **renoncer à l'appui financier** d'une action non réalisée ?* (27/12/2022 ; https://www.centre-val-de-loire.developpement-durable.gouv.fr/IMG/pdf/guidepaiementstepcv_v20180409.pdf)

(87) *En cas d'abandon de leur projet : adresser un courrier motivé au secrétariat technique régional en indiquant explicitement **renoncer au soutien financier** en tant que lauréats de ce concours.* (27/12/2022 ; https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000022963231/2010-12-31)

En espagnol, le verbe *rechazar* apparaît aussi dans des constructions NON1 et dans des constructions converses :

(88) *Pregunta sobre si el Gobierno comparte el criterio del Ministerio de Educación y Ciencia de **rechazar** la **subvención** de actividades universitarias si los organizadores y los ponentes están todos circunscritos al ámbito catalán.* (02/01/2023 ; https://www.senado.es/web/composicionorganizacion/senadores/composicion_senado/fichasenador/preguntasescritas/index.html;jsessionid=DvbBfbLT0b_yyy4qh8LvkLpsL9QppGJX5bcv10Q2TkJKN3RQV9Qhb!-678587447?id1=12804&legis=8&id2=684&id3=pre)

(89) *Los beneficiarios tendrán un plazo de 10 días, desde la notificación de la resolución definitiva, para aceptar o **rechazar** la **subvención**.* (02/01/2023 ; https://www.senado.es/web/composicionorganizacion/senadores/composicion_senado/fichasenador/preguntasescritas/index.html;jsessionid=DvbBfbLT0b_yyy4qh8LvkLpsL9QppGJX5bcv10Q2TkJKN3RQV9Qhb!-678587447?id1=12804&legis=8&id2=684&id3=pre)

(90) *La empresa Zinez **rechaza** la **ayuda** de 9.000 euros del Ayuntamiento de San Sebastián, gobernado por Bildu.* (09/06/2023 ; <https://www.elperiodico.com/es/politica/20120628/productora-documental-presos-eta-renuncia-subvencion-1990761>)

À part le verbe *rechazar*, nous avons repéré d'autres verbes de <refus> tels que *denegar* et *negar* :

(91) *Bruselas **denegó** la **ayuda** a España para financiar las vallas en Ceuta y Melilla.* (02/01/2023 ; <https://www.lavanguardia.com/politica/20150618/54432905580/bruselas-denego-ayuda-espana-financiar-vallas-ceuta-melilla.html>)

- (92) *Juzgan en Bilbao a la directora del SEPE denunciada por LAB por denegar el subsidio de paro a expresos de ETA.* (09/06/2023 ; <https://www.europapress.es/nacional/noticia-juzgan-bilbao-directora-sepe-denunciada-lab-denegar-subsidio-paro-expresos-eta-20170206120835.html>)

En ce qui concerne les constructions converses impliquant des verbes de <refus>, de même qu'en français nous avons observé que *refuser* et *renoncer* à se combinent avec tous les noms de la classe <aide financière : allocation>, en espagnol, ce sont les verbes *rechazar* (déjà mentionné) et *renunciar a* qui fonctionnent avec toute la classe :

- (93) *Ante esta situación, muchos trabajadores se preguntan si es posible renunciar temporalmente a la pensión de jubilación para seguir cotizando y solicitarla en un momento más favorable.* (09/06/2023 ; <https://noticiastrabajo.huffingtonpost.es/economia/pensiones/como-renunciar-pension-seguir-cotizando-cobrar-mas/>)
- (94) *Para poder trabajar, la mujer renunció a una Renta Mínima de Inserción (RMI) y ahora no tiene trabajo ni ingresos.* (09/06/2023 ; <https://www.migranodearena.org/reto/alimentar-els-fills>)
- (95) *La productora del documental sobre presos de ETA renuncia a la subvención.* (09/06/2023 ; <https://www.elperiodico.com/es/politica/20120628/productora-documental-presos-eta-renuncia-subvencion-1990761>)

Finalement, nous avons repéré certains verbes prédicatifs que nous avons appelés des verbes de <versement> : *liquidar*, *payer*, *régler* et *verser*. Nous montrons ci-dessous quelques exemples :

- (96) *En ayant jugé que la Caisse n'avait pas commis de faute en refusant de liquidar la pension de retraite de l'assuré à hauteur des montants qui lui étaient dus, prétexte pris de ce que la position que la Caisse avait adoptée avait été confirmée par des cours d'appel, la cour d'appel a violé l'article 1382 ancien du code civil.* (12/05/2023 ; <https://www.courdecassation.fr/decision/619f34d15e6dd569d4273326>)

- (97) *Après les sénateurs, les députés veulent moderniser la loi de 1975 qui conduit souvent les divorcés à **payer** une **rente** à vie à leur ex-conjoint, sans possibilité de révision en cas de chômage, maladie ou décès.* (12/05/2023 ; <https://www.humanite.fr/societe/-/une-bouee-pour-les-naufrages-du-divorce>)
- (98) *La CNBF a refusé de lui **régler** l'**allocation d'invalidité temporaire** qu'elle doit verser à compter du 91ème jour au motif qu'en dépit de ses demandes.* (12/05/2023 ; <https://www.doctrine.fr/d/CA/Paris/2015/R0A9C045BCBDBA40B57B9>)
- (99) *La municipalité a dû suspendre ses allocations mère-enfant et ne pourra verser son **soutien financier** trimestriel aux écoles publiques, qui scolarisent 43% des élèves de la ville et courent donc le risque de mettre temporairement à la porte près de 35 000 élèves.* (12/05/2023 ; <https://www.nouvelobs.com/rue89/rue89-etats-unis/20131011.RUE9407/ferme-pour-cause-de-shutdown-ma-journee-de-guigne-a-washington.html>)

Néanmoins, seulement le verbe *verser* se combine avec tous les noms de la classe. Le verbe *payer* a une combinatoire très ample avec les noms d'<aide financière : allocation>, mais nous n'avons pas trouvé d'occurrences où il se combine avec le nom *appui*. Par contre, les verbes *liquider* et *régler* ont une combinatoire plus restreinte. Nous tenons à rappeler que dans notre dictionnaire électronique, nous n'avons retenu que les verbes se combinant avec tous les noms de la classe, pour éviter que l'automate génère du bruit.

En espagnol, seulement les verbes *abonar*, *ingresar* et *pagar* se combinent avec tous les noms de la classe :

- (100) *Urgen al SEPE a **abonar** el **subsidio** a empleadas de hogar afectadas por la crisis.* (12/05/2023 ; <https://www.galiciae.com/articulo/economia/urgen-sepe-abonar-subsidio-empleadas-hogar-afectadas-crisis/20200807192935082869.html>)
- (101) *Cuándo se cobra el paro en mayo: esta es la fecha en la que el SEPE te va a **ingresar** el **subsidio** por desempleo.* (12/05/2023 ;

<https://www.businessinsider.es/cuando-cobra-paro-mayo-fecha-ingreso-1050441>)

(102) *El Ayuntamiento de Santander tendrá que **pagar** a la Asociación de Vecinos de Cueto la **subvención** de 3.450 euros que le hubiera correspondido para el ejercicio de 2018 y que no le **abonó** en su día alegando que a agrupación no había justificado gastos.* (12/05/2023 ; <https://www.eldiariomontanes.es/santander/condenan-ayuntamiento-pagar-20200319211929-ntvo.html>)

Nous présentons ci-dessous une fiche récapitulative de la combinatoire verbale des noms des <aides financières : allocations> :

	FRANÇAIS	ESPAGNOL
Entrées	Aide à la recherche d'emploi, allocation, allocs, appui, APL, bourse Erasmus, pension, prestation d'accueil du jeune enfant, rente d'accident du travail, secours viager, soutien, subvention d'investissement...	Ayuda al alquiler, beca postdoctoral, pensión alimenticia, prestación de orfandad por violencia contra la mujer, renta, subsidio agrícola, subsidio extraordinario por desempleo...
Schéma d'arguments	N0:Nhum, Ncoll /N1: à Nhum, Ncoll/N2: de <montant>/N3: pour Vinf	N0:Nhum, Ncoll /N1: a Nhum, Ncoll/N2: de <montant>/N3: para Vinf
Verbes supports appropriés	Accorder, allouer, apporter, attribuer, consentir (FIN., DR.), donner, fournir, octroyer, offrir, procurer.	Aportar, asignar, brindar, conceder, dar, ofrecer, otorgar, proporcionar, suministrar.
Verbes supports	Avoir, avoir droit à, bénéficier de, disposer de,	Contar con, disfrutar de, gozar de, percibir, recibir, tener,

passifs (N1N0)	encaisser, jouir de, palper (fam.), percevoir, profiter de, recevoir, toucher (fam.).	tener derecho a.
Verbes supports aspectuels inchoatifs	Décrocher, gagner, obtenir, trouver.	Conseguir, ganar, lograr, obtener.
Verbes supports aspectuels progressifs	Maintenir. Conversif : conserver, garder, maintenir.	Mantener, prolongar, prorrogar, renovar. Conversif : conservar, mantener.
Verbes supports aspectuels terminatifs	Annuler, couper, geler, retirer, supprimer, suspendre. Conversif : perdre.	Anular, cancelar, congelar, eliminar, quitar, retirar, suprimir, suspender Conversif : perder
Verbes supports aspectuels itératifs	Débloquer, reconduire, renouveler.	Desbloquear, reconducir, renovar.
Verbes supports aspectuels intensifs à valeur progressive-augmentative	Améliorer, augmenter, majorer.	Ampliar, aumentar, elevar, incrementar, mejorar, subir.
Verbes supports aspectuels intensifs à valeur progressive-atténuative	Diminuer, limiter, réduire.	Bajar, congelar, cortar, disminuir, rebajar, recortar, reducir
Verbes prédicatifs appropriés de <demande>	Demander, réclamer, solliciter.	Demandar, exigir, pedir, reclamar, solicitar.

Verbes prédicatifs appropriés de <mise en place>	Introduire, mettre en place, voter.	Aprobar, establecer, instituir, poner en marcha.
Verbes prédicatifs appropriés de <refus>	Refuser. Conversif : refuser, renoncer à.	Denegar, negar, rechazar. Conversif : rechazar, renunciar a
Verbes prédicatifs appropriés de <versement>	Verser.	Abonar, ingresar, pagar

Tableau 19: Fiche descriptive de la classe <aide financière : allocation>

2. <aide financière : donation>

La classe <aide financière : donation> comporte un nombre très limité de noms : *don* et *donation* en français, et *donación* et *donativo* en espagnol. Nous n'avons pas considéré des noms du domaine de spécialité du droit tels que *donation-partage* (fr.) *libéralité* (fr.), *legs* (fr.), *liberalidad* (es.) ou *legado* (es.) car ils ne constituent pas des noms d'<aide>, mais d'<héritage>.

Les noms appartenant à cette classe sont caractérisés par le schéma d'arguments suivant :

N0	N1	N2	N3
hum, coll_hum	à hum, coll_hum	de <montant>	pour Vinf

Tableau 20 : Schéma d'arguments d'<aide financière : donation> en français

N0	N1	N2	N3
hum, coll_hum>	a hum, coll_hum	de <montant>	para Vinf

Tableau 21 : Schéma d'arguments d'<aide financière : donation> en espagnol

La position de premier argument est occupée par un nom caractérisé par le trait humain (*hum*) ou un collectif humain (*coll_hum*, p. ex. : une association, une institution,

une organisation, etc.). La position N1 correspond au premier complément, c'est-à-dire la personne (*hum*) ou la collectivité (*coll_hum*) bénéficiaire de cette <donation>. Le N1 apparaît souvent sous la forme d'un groupe prépositionnel commençant par la préposition *à* (fr.) ou *a* (es.), mais nous avons trouvé aussi des locutions prépositionnelles telles que *à faveur de* (exemple numéro 128). L'argument N2 correspond à un groupe prépositionnel dénotant un montant d'argent. Finalement, la position N3 est occupée par un groupe prépositionnel contenant un verbe en infinitif. Dans le schéma d'arguments nous avons indiqué que le groupe prépositionnel commence par la préposition *pour* (fr.) ou *para* (es.). Comme nous pouvons l'observer dans l'exemple 119, le N3 peut aussi commencer par *dans le but de*.

En français, les noms prédicatifs des <aides financières : donations> peuvent être actualisés par des verbes supports tels qu'*accorder, consentir, effectuer, faire, octroyer, offrir* et *réaliser* :

(103) *L'association paroissiale de Saint-Carreuc fait un don de 500 € pour aider le peuple ukrainien.* (16/04/2023 ; <https://www.letelegramme.fr/cotes-d-armor/saint-carreuc-22150/l-association-paroissiale-de-saint-carreuc-fait-un-don-de-500-eur-pour-aider-le-peuple-ukrainien-218968.php>)

(104) *Au cours d'un plan bien ficelé, les deux humoristes sont entrés tout d'abord en contact avec Piers Corbyn via mail pour offrir une donation à sa campagne antivax «Stop New Normal».* (18/08/2023 ; https://www.liberation.fr/international/europe/royaume-uni-le-celebre-militant-antivax-piers-corbyn-pris-au-piege-dans-un-canular-piquant-20210804_WZ2TI4VF5ZHX3EWAD2A3Q3BK2U/)

En espagnol, nous avons repéré les verbes supports *aportar, brindar, conceder, dar, efectuar, entregar, enviar, hacer, hacer entrega de, ofrecer, otorgar* et *realizar* :

(105) *En la Catedral de Valencia se celebra una misa por Franco a petición de un particular que ha dado un donativo de diez euros.* (27/05/2021 ; https://www.lasexta.com/noticias/nacional/la-catedral-de-valencia-acoge-una-misa-por-francisco-franco-tras-la-propuesta-de-un-particular_20160718578d1b026584a8b7b4244549.html)

(106) *A través de este fondo, Sony Corporation hizo una donación de 3 millones de dólares (USD) a ACNUR para ayudar a proteger a las personas refugiadas de la pandemia del coronavirus.* (27/05/2021 ; <https://www.acnur.org/es-es/sony>)

Quant aux verbes supports apparaissant dans des constructions converses (N1N0), en français nous avons trouvé un nombre très limité : *bénéficier de*, *percevoir* et *recevoir*. Nous montrons ci-dessous quelques exemples :

(107) *La Ville de Jarny a perçu un don, sous la forme d'un chèque de 500 €, en provenance d'Israël.* (16/04/2023 ; <https://www.vosgesmatin.fr/insolite/2021/10/03/le-remboursement-d-une-dette-d-enfant-61-ans-plus-tard>)

(108) *Depuis plusieurs semaines, une trentaine de communes du département du Nord affirment avoir reçu des donations considérables, envoyées par un expéditeur anonyme.* (16/04/2023 ; <https://www.lanouvellerepublique.fr/a-la-une/un-donateur-anonyme-envoie-des-centaines-d-euros-a-plusieurs-communes-du-nord>)

(109) *Le Parti conservateur britannique a largement bénéficié des donations des milliardaires russes qui ont trouvé refuge à Londres depuis les années 1990.* (16/04/2023 ; https://www.lemonde.fr/m-le-mag/article/2022/03/10/au-royaume-uni-l-embarrassante-generosite-des-oligarques-russes_6116939_4500055.html)

Par contre, en espagnol, les verbes supports passifs se combinant avec les noms de la classe <aide financière : donation> sont plus nombreux : *recibir*, *percibir*, *recaudar* et *recoger*. Nous montrons ci-dessous quelques exemples :

(110) *De paso, Eric Woerth, que también era tesorero del partido del Presidente, percibió presuntamente donaciones de decenas de miles de euros para financiar la campaña electoral de Sarkozy...* (17/07/2022 ; <https://mondiplo.com/la-corrupcion-de-la-democracia-2939>)

(111) *La Escuela Fundación mésmúsica ha conseguido recaudar un donativo de más de dos mil euros para dar apoyo a la labor de Mallorca Sense Fam a*

través del ‘Concert d’Estiu 2021’, celebrado el pasado sábado en el patio de Can Balaguer. (17/07/2022 ; <https://www.diariodemallorca.es/mallorca/2021/06/22/concert-d-estiu-recauda-mil-54029700.html>)

En ce qui concerne les verbes supports à valeur aspectuelle, ils se sont avérés moins nombreux que dans d’autres classes, comme celle des <aides financières : allocations>. En effet, nous n’avons pas trouvé de combinaisons avec des verbes supports aspectuels à valeur progressive, terminative ni intensive. Pourtant, l’aspect itératif s’est avéré très prolifique. À part le nombre très réduit de noms prédicatifs appartenant à la classe <aide financière : donation>, leur aspect interne ponctuel pourrait expliquer cette combinatoire très restreinte avec certains verbes supports aspectuels.

En français, nous avons seulement repéré un verbe support aspectuel inchoatif. Il s’agit du verbe *obtenir*, qui, rappelons-le, apparaît dans des constructions converses (N1N0):

(112) *La chaire Finance comportementale de l’EM a obtenu un don de 20 000€ pour soutenir ce projet. (17/04/2023 ; <https://www.em-strasbourg.com/fr/actualites/don-de-loee-pour-la-chaire-finance-comportementale>)*

(113) *Le vrai problème n’a pas été d’acheminer la nourriture à ceux qui ont faim mais d’obtenir les donations pour payer la nourriture », a déclaré Pierre Carrasse, directeur de la logistique du PAM. (17/04/2023 ; <https://agence-juridique.com/articles/cession-parts-sociales-sci-conditions-et-formalites/>)*

En espagnol, nous avons trouvé les verbes *obtener* et *conseguir* :

(114) *El hijo mayor del presidente admite ante la Fiscalía haber obtenido donaciones de un exnarco para la campaña de su padre, que no fueron registradas, y haberse quedado con parte del dinero. (05/08/2023 ; <https://elpais.com/america-colombia/2023-08-05/la-financiacion-de-la-campana-de-gustavo-petro-provoca-un-terremoto-politico-en-colombia.html>)*

(115) *Manuel Velázquez era un chaval en 1975, cuando se dedicaba a llamar a las puertas del pueblo para conseguir un donativo con el que organizar los festejos de San Roque.* (17/07/2022 ; <https://huetorvega.ideal.es/actualidad/huetenos-gente-noble-20170816194421-nt.html>)

Quant à l'aspect itératif, en français nous avons trouvé les verbes supports *distribuer* et *répartir* :

(116) *Toutes ces actions ont permis de répartir des dons aux diverses associations œuvrant pour le mieux-être des enfants et adultes en situation de handicap.* (17/04/2023 ; <https://www.lanouvellerepublique.fr/chateauroux/inner-wheel-nouvelle-presidente>)

(117) *Un projet ambitieux mais aussi un message non moins pragmatique à l'égard des associations : « Je préfère financer des projets que de distribuer des donations récurrentes. ».* (17/04/2023 ; <https://www.lanouvellerepublique.fr/indre-et-loire/commune/la-riche/le-racing-a-le-vent-en-poupe>)

L'aspect itératif s'est avéré bien plus prolifique dans des constructions converses. Ainsi, nous avons repéré des verbes tels qu'*accumuler*, *collecter*, *cumuler*, *rassembler*, *récolter* et *recueillir* :

(118) *Grâce à deux campagnes de financement sur le site Gofundme, ils ont pu accumuler des dons de la population pour la reconstruction.* (16/04/2023 ; <https://www.laction.com/article/2021/06/10/une-campagne-pour-soutenir-des-eleveurs-de-sangliers>)

(119) *L'association des Amis du patrimoine culturel de L'Absie a collecté des donations au fil des vingt dernières années, dans le but de restaurer les peintures murales et souhaite remettre à la commune la somme totale de 61.000 € sur cinq tranches de travaux pour l'église.* (17/04/2023 ; <https://www.lanouvellerepublique.fr/deux-sevres/commune/l-absie/eglise-une-aide-des-amis-du-patrimoine>)

- (120) *On est donc à 8 000 euros et on sait que depuis plusieurs années on continue à **cumuler** des **dons** grâce au 36 37 et à des dons d'associations tardifs.* (17/04/2023 ; <https://www.ledauphine.com/societe/2021/01/14/isere-morestel-le-telethon-collecte-8-000-euros-malgre-la-crise-sanitaire>)
- (121) *Le Téléthon, destiné à **rassembler** des **dons** pour financer la recherche sur les maladies génétiques rares, s'achèvera samedi à minuit au terme de 30 heures d'animations diverses.* (17/04/2023 ; https://www.lepoint.fr/societe/telethon-face-a-la-baisse-des-dons-l-afm-sonne-la-mobilisation-08-12-2012-1547277_23.php)
- (122) *Au programme de cette soirée : une vente aux enchères et des lots en tombola pour **récolter** des **donations** auprès de mécènes : entreprises comme particuliers.* (17/04/2023 ; <https://www.cbainfo.fr/actualites/partenariats/cba-soutient-linstitut-sainte-catherine-specialise-en-cancerologie/>)

Rappelons que dans les constructions converses, le N0 soit il disparaît, soit il apparaît dans un groupe prépositionnel, le plus souvent commençant par la préposition *de*. Dans le dernier exemple ci-dessus (122) le N0 est introduit par la locution prépositionnelle *auprès de*.

En espagnol, la combinatoire verbale des verbes supports aspectuels à valeur itérative est similaire à celle du français. En effet, les recherches sur le corpus ont donné plus de cooccurrences de verbes apparaissant dans des structures converses. D'une part, nous avons repéré les verbes *distribuir* et *repartir* :

- (123) *El Retirement and Medical Charitable Trust (RMCT) es un fideicomiso que la Congregación creó en Nueva Zelanda en 2010 para recibir y **distribuir donativos** específicamente destinados a cubrir gastos en favor de sacerdotes y personas consagradas ancianos y enfermos o a otros fines religiosos, caritativos y educativos.* (17/07/2022 ; <https://www.legionariosdecristo.org/direcciongeneral/pandora-papers-legionarios-de-cristo-rmct/>)
- (124) *En 2006 recibió el Premio Humanitario de Acción contra el Hambre y en 2013 se convirtió en la primera famosa "100 Good Deeds Bracelet", una organización sin ánimo de lucro que **reparte donaciones** económicas y*

oportunidades de educación y enseñanza a las mujeres y niñas de África y otros lugares. (17/07/2022 ; <https://www.revistavpc.es/noticias-del-sector/2835-susan-sarandon-se-une-a-las-embajadoras-de-l-oreal-paris.html>)

D'autre part, nous avons repéré les verbes supports passifs à valeur itérative *acumular, coleccionar, recolectar* et *reunir* :

(125) *A primera hora de la mañana del lunes, estos colaboradores montaban una mesa petitoria en la Plaza Fuente Arriba con el objetivo de **recolectar** las **donaciones** de los vecinos para un fin tan loable como la lucha contra el cáncer.* (17/07/2022 ; <http://www.alora.es/Inicio/N.asp?IdArticulo=2547>)

(126) *En varios escritos a los que ha tenido acceso El Confidencial, el Instituto Armado asegura que la asociación del antiguo magistrado, dedicada a la presunta elaboración de denuncias falsas contra distintas autoridades para presentarse como adalid de la lucha contra la corrupción, empleó las cuentas de un abogado del entorno de Esquerra Republicana de Cataluña (ERC) para **reunir donativos**.* (18/07/2023 ; https://www.elconfidencial.com/espana/2023-07-18/red-bulos-supremo-juez-presencia-abogado-erc_3699749/)

Finalement, nous avons observé que les noms des <aides financières : donation> se combinent souvent avec des verbes prédicatifs de <demande>, de <refus> et de <versement>. En ce qui concerne les verbes de <demande>, en français nous avons repéré les verbes *demander* et *solliciter* se combinant avec les noms de la classe <aide financière : donation> :

(127) *Les réseaux sociaux comme Facebook vous donnent aussi la possibilité de créer une cagnotte de financement participatif. Vous pouvez, par exemple, **demander des dons** à votre entourage, à l'occasion de votre anniversaire ou encore de votre mariage.* (16/04/2023 ; <https://www.msf.fr/agir/soutenir-nos-actions/appele-aux-dons>)

(128) *La star de Sex and The City n'a pas hésité à faire une vidéo pour **solliciter des donations** en faveur du Parti démocrate et vanter les mérites d'Obama.* (27/04/2023 ; <https://tvmag.lefigaro.fr/programme-tv/article/people/72350/obama-romney-pour-qui-votent-les-stars-tv-.html>)

En espagnol nous avons trouvé également les verbes *pedir* et *solicitar* :

(129) *Naciones Unidas **pidió** un **donativo** de 108 millones de dólares (75.7426 millones de euros) a comienzos de este mes para ayudar al empobrecido país.* (15/07/2022 ; <https://www.europapress.es/epsocial/igualdad/noticia-gobierno-haiti-onu-estan-abrumados-magnitud-desastre-paso-ultimos-huracanes-20080919075402.html>)

(130) *El Ayuntamiento de Los Llanos **solicita donaciones** de particulares para ayudar a los afectados por la erupción. Pone a disposición una cuenta bancaria para centralizar las donaciones particulares, empresas o cualquier tipo de entidad que se quiera realizar para ayudar a los damnificados por la erupción del volcán de La Palma.* (15/07/2022 ; https://cadenaser.com/emisora/2021/09/20/radio_club_tenerife/1632163724_978694.html)

Quant aux verbes prédicatifs de <refus>, nous avons trouvé les verbes *refuser* (en français) et *rechazar* (en espagnol). En français ce verbe apparaît dans des constructions N0N1 (exemple numéro 131) et N1N0 (exemple numéro 132), mais en espagnol, il n'apparaît que dans des constructions N1N0 (exemples numéros 133 et 134) :

(131) *Les donateurs ont l'autorisation de **refuser** la **donation** à la dernière minute.* (16/04/2023 ; <https://fr.clearharmony.net/articles/a31441-Des-preuves-directes-prouvent-que-les-victimes-des-prelevements-d%E2%80%99organes-sont-des-pratiquants-du-Falun-Gong-Photo.html>)

(132) *Le groupe polonais a accusé Unicef Pologne d'**avoir refusé** un **don** de 2 200 euros pour soutenir les enfants victimes de la guerre en Ukraine, en raison de l'image du groupe.* (21/08/2023 ; https://www.liberation.fr/checknews/est-il-vrai-que-lunicef-a-refuse-un-don-du-groupe-de-metal-sataniste-behemoth-pour-les-enfants-ukrainiens-20230821_P7FWEJWA5ZDABPNLUY7PLU5AOU/)

(133) *El jefe del Gabinete de Ministros argentino, Marcos Peña, **pidió** hoy "desdramatizar" que el papa Francisco **haya rechazado** la **donación** hecha*

por el Gobierno de Argentina a su proyecto educativo Scholas Occurrentes y lo distanció de la relación entre el pontífice y el presidente del país, Mauricio Macri. En declaraciones a radio La Red, Peña aclaró que la propia fundación había pedido al Gobierno la donación de 16 millones de pesos (1,14 millones de dólares). (16/04/2023 ; <https://www.lavanguardia.com/vida/20160613/402480942678/gobierno-argentino-pide-desdramatizar-rechazo-de-donacion-por-parte-del-papa.html>)

(134) *En un comunicado, Fesbal rechaza el donativo e invita al PSOE a donar la cantidad a cualquier otra entidad benéfica.* (17/07/2022 ; <https://www.compromisorse.com/rse/2012/11/05/fesbal-rechaza-donativos-de-partidos-politicos-/>)

En ce qui concerne les verbes prédicatifs de <versement>, en français nous n'avons repéré que le verbe *verser* :

(135) *Des personnes touchées par sa situation lui ont versé des donations et plusieurs avocats lui ont prodigué des conseils gratuits.* (27/04/2023 ; <https://www.gqmagazine.fr/pop-culture/news/articles/des-sud-africains-volent-un-bus-pour-se-rendre-a-un-examen/29627>)

(136) *Crise aux Restos du cœur : Bernard Arnault annonce verser un « don solidaire » de 10 millions d'euros.* (04/09/2023 ; https://www.francetvinfo.fr/sante/alimentation/restos-du-coeur-bernard-arnault-annonce-verser-un-don-solidaire-de-10-millions-d-euros_6042785.html)

En espagnol, la combinatoire s'est avérée légèrement plus riche, car nous avons repéré trois verbes se combinant avec les noms d'<aide financière : donation> : *abonar*, *ingresar* et *pagar*. Nous montrons ci-dessous quelques exemples :

(137) *Los que se sientan incómodos dentro de espacios religiosos o simplemente no quieran abonar una donación recomendada de diez dólares, que no sufran: Saint John induce a un viaje lisérgico sin necesidad de penetrar su interior.* (17/07/2022 ; https://elpais.com/elviajero/2012/07/24/actualidad/1343150659_804512.html)

- (138) *Además, un empresario madrileño que tiene negocios de hostelería en la Costa del Sol le **ha ingresado un donativo** y en adelante le dará un porcentaje de la recaudación de uno de sus establecimientos.* (17/07/2022 ; https://sevilla.abc.es/sevilla/sevi-solidaridad-nino-discapacitado-no-puede-hablar-201610212154_noticia.html?ref=https%3A%2F%2Fwww.google.com%2F)
- (139) *Una catedral de Castilla-La Mancha instala un sistema para que los fieles puedan **pagar donativos** con tarjeta.* (17/07/2022 ; <https://www.eldigitaldealbacete.com/2019/05/28/una-catedral-de-castilla-la-mancha-instala-un-sistema-para-que-los-fieles-puedan-pagar-donativos-con-tarjeta/>)

Finalment, nous recueillons l'actualisation des noms d'<aide financière : donation> dans la fiche ci-dessous :

	FRANÇAIS	ESPAGNOL
Entrées	Don, donation.	Donación, donativo.
Schéma d'arguments	N0:Nhum, Ncoll /N1: à Nhum, Ncoll/N2: de <montant>/N3: pour Vinf	N0:Nhum, Ncoll /N1: a Nhum, Ncoll/N2: de <montant>/N3: para Vinf
Verbes supports appropriés	Accorder, consentir, effectuer, faire, octroyer, offrir, réaliser.	Aportar, brindar, conceder, dar, efectuar, entregar, enviar, hacer, hacer entrega de, ofrecer, otorgar, realizar.
Verbes supports passifs (N1N0)	Bénéficiaire de, percevoir, recevoir.	Recibir, percibir, recaudar, recoger.
Verbes supports aspectuels inchoatifs	Obtenir.	Obtener, conseguir.

Verbes supports aspectuels itératifs	Distribuer, répartir. Conversif : accumuler, collecter, rassembler, recueillir, cumuler, récolter.	Distribuir, repartir. Conversif : acumular, colectar, recolectar, reunir.
Verbes prédictifs appropriés de <demande>	Demander, solliciter.	Pedir, solicitar.
Verbes prédictifs appropriés de <refus>	Refuser. Conversif : refuser.	Rechazar.
Verbes prédictifs appropriés de <versement>	Verser.	Abonar, ingresar, pagar.

Tableau 22 : Fiche descriptive de la classe <aide financière : donation>

3. <aide financière : dégrèvement>

Nous décrivons à présent les noms prédictifs la classe <aide financière : dégrèvement>. Cette classe comporte des unités monolexicales (p. ex., *ristourne*, *abattement* ou *exonération* en français ; *rebaja*, *reducción* ou *exoneración* en espagnol) et des unités polylexicales (p. ex., *allègement fiscal*, *escompte de caisse* ou *exonération d'impôt* en français ; *reducción fiscal*, *descuento por pronto pago* ou *exención tributaria* en espagnol). Comme nous pouvons l'observer dans ce petit échantillon d'exemples que nous venons de donner, cette classe contient de nombreux noms appartenant au discours de spécialité du droit fiscal (p. ex., *allègement fiscal* / *reducción fiscal* ; *exonération d'impôt* / *exención tributaria*) et du commerce (p. ex., *escompte de caisse* / *descuento por pronto pago*), et elle contient également des noms du discours général tels que *rabais*, *réduction* ou *remise* en français, et *rebaja*, *reducción* ou *descuento* en espagnol.

Ces noms ont d'abord été caractérisées par leur schéma d'arguments :

N0	N1	N2	N3
coll_hum	de <montant>, #de N	sur N, #pour N	à hum, coll_hum>, #pour hum, coll_hum , #en faveur de hum, coll_hum

Tableau 23 : Schéma d'arguments d'<aide financière : allègement> en français

N0	N1	N2	N3
coll_hum	de <montant>, #de N	en N, #por N, #para N	a hum, coll_hum

Tableau 24 : Schéma d'arguments d'<aide financière : allègement> en espagnol

Les noms de la classe <aide financière : dégrèvement> sont actualisés, en français, par des verbes supports appropriés tels qu'*accorder*, *consentir*, *faire*, *octroyer*, *offrir* et *pratiquer* :

- (140) Dans une [réponse](https://investir.lesechos.fr/placements/impots/taxe-dhabitation-il-ny-aura-pas-de-degrevement-exceptionnel-pour-les-residences-secondaires-inaccessibles-1931980) publiée le 16 février dernier au Journal Officiel, le gouvernement a refusé d'**appliquer un dégrèvement** exceptionnel de la taxe d'habitation pour les propriétaires qui n'ont pu profiter de leur résidence secondaire en raison du [confinement](https://investir.lesechos.fr/placements/impots/taxe-dhabitation-il-ny-aura-pas-de-degrevement-exceptionnel-pour-les-residences-secondaires-inaccessibles-1931980) du printemps 2020. (25/02/2021 ; <https://investir.lesechos.fr/placements/impots/taxe-dhabitation-il-ny-aura-pas-de-degrevement-exceptionnel-pour-les-residences-secondaires-inaccessibles-1931980>)
- (141) La Ville de Paris **fait un rabais** de 36 millions à des promoteurs immobiliers. (25/02/2021 ; <https://www.marianne.net/politique/la-ville-de-paris-fait-un-rabais-de-36-millions-des-promoteurs-immobiliers>)
- (142) Lorsque l'annonce de réduction de prix est d'un taux uniforme et se rapporte à des produits parfaitement identifiés, le commerçant a la possibilité de **pratiquer un escompte à la caisse**. (25/02/2021 ; <https://www.lesechos.fr/1998/12/les-soldes-dhiver-debutent-le-2-janvier-et-dureront-six-semaines-805937>)

En espagnol aussi, nous avons repéré un nombre élevé de verbes supports appropriés étant sélectionnés par les noms d'<aide financière : dégrèvement> : *aplicar*,

conceder, hacer, ofrecer, otorgar, practicar, proponer et realizar. Nous montrons ci-dessous quelques exemples :

(143) *De nuevo, el presidente del PP criticó a Moncloa por "anunciar el cheque bebé y quitarlo" y por "presentarse a las elecciones diciendo que iba a **hacer** una **desgravación** en el impuesto de renta de 400 euros".* (01/03/2021 ; https://www.diariodesevilla.es/sociedad/Rajoy-cambiara-Ley-Aborto-practiquen_0_527047896.html)

(144) *También [el grupo municipal del PP] ha propuesto **realizar** "una **exención fiscal** a comercios, hosteleros y a las propias comunidades de vecinos afectadas", así como "trabajar con la hostelería en una línea de subvenciones que contribuyan a mejorar la estética de las terrazas una vez se puedan instalar".* (27/12/2022 ; <https://www.eldiadelarioja.es/Noticia/Z4376791F-C621-81F4-B280107E65738659/202212/El-PP-propone-romper-el-contrato-de-obra-de-las-Cien-Tiendas>)

(145) *La incógnita es saber dónde se situará el PSOE de Lambán; si en la línea del valenciano Ximo Puig, el más audaz, que **ofrece** una **rebaja fiscal** en el IRPF a las rentas por debajo de 60.000 euros, o en la del extremeño Guillermo Fernández Vara, más contenido, que se limita a una "bajada histórica" de los precios y tasas públicos.* (21/10/2022 ; <https://www.heraldo.es/noticias/aragon/2022/10/02/aragon-cobro-impuestos-millones-pib-1603080.html>)

Nous souhaitons attirer l'attention sur ce dernier exemple, car il nous semble particulièrement intéressant mais il ne s'agit pas d'un cas rare pour autant. Dans cette phrase, le bénéficiaire de cette réduction d'impôt (*las rentas por debajo de 60.000 euros*) est désigné moyennant un procédé très habituel de métonymie. Ainsi, le nom à trait <humain> est remplacé par le nom *revenu*. Nous comprenons donc que ce bénéficiaire est une personne dont les revenus annuels sont au-dessous de 60.000 euros.

Les noms d'<aide financière : dégrèvement> peuvent également apparaître dans des constructions converses dans lesquelles l'argument NO disparaît. En français, nous

avons repéré les verbes supports passifs *avoir*, *avoir droit à*, *bénéficier de* et *profiter de* :

- (146) *Dans le cas d'un logement neuf très économe en énergie, et si la construction correspond à un niveau de performance énergétique supérieur à celui imposé par la législation, vous pouvez **avoir** une **exonération** pendant cinq ans de la Taxe Foncière de 50 %.* (25/02/2021 ; <https://www.ouest-france.fr/bretagne/betton-35830/la-taxe-fonciere-et-les-nouvelles-exonerations-e663b036-e127-40b1-ae58-415fdc85f5b6>)
- (147) *Près de 100.000 PME **ont bénéficié** d'un **allègement** de cotisation foncière pour 2020.* (25/02/2021 ; <https://www.lesechos.fr/politique-societe/regions/pres-de-100000-pme-ont-beneficie-dun-allegement-de-cotisation-fonciere-pour-2020-1281482>)
- (148) *Grâce à cette opération inédite d'autoconsommation collective, ses habitants **profitent d'une réduction** de leur facture d'énergie, et s'inscrivent dès à présent dans la transition énergétique.* (26/07/2023 ; <https://www.edf.fr/collectivites/le-mag/le-mag-collectivites/territoires-realisation/cap-sur-la-plus-grande-experimentation-d-autoconsommation-collective-dans-des-logements-sociaux-en-france>)

En espagnol, nous avons obtenu les verbes supports passifs *beneficiarse de*, *disfrutar de*, *gozar de*, *percibir*, *recibir* et *tener* :

- (149) *Todas las madres con hijos menores de tres años dadas de alta en la Seguridad Social —tanto asalariadas como autónomas— pueden **beneficiarse de una deducción** de 1.200 euros al año en su declaración de la renta.* (12/04/2023 ; <https://cadenaser.com/nacional/2023/04/12/que-me-puedo-desgravar-en-la-declaracion-de-la-renta-2022-2023-todos-los-gastos-que-puedes-deducirte-cadena-ser/>)
- (150) *Desde que en 2015 el Gobierno puso en marcha las deducciones de 1.200 euros al año (100 euros al mes) por familia numerosa o personas con discapacidad a cargo, los contribuyentes **han recibido una rebaja fiscal** de 1.142 millones de euros, según los presupuestos. La deducción tiene dos modalidades de cobro.* (21/10/2022 ;

https://cincodias.elpais.com/cincodias/2017/02/22/economia/1487752017_886342.html)

- (151) *El presidente del Principado ha subrayado hoy que las empresas familiares de Asturias **tienen** una **bonificación** del 99% en el impuesto de sucesiones.* (09/10/2023 ; <https://actualidad.asturias.es/-/barb%C3%B3n-subraya-que-las-empresas-familiares-de-asturias-tienen-una-bonificaci%C3%B3n-del-99%25-en-el-impuesto-de-sucesiones>)

Nous tenons à signaler que certains des verbes supports appropriés signalés précédemment peuvent également apparaître dans des constructions converses. Tel est le cas, par exemple du verbe *hacer* en espagnol ou le verbe *pratiquer* en français :

- (152) *Es verdad que hay incentivos fiscales, pues quien dona tiene derecho a **hacer una desgravación** de hasta el 25%, pero nunca por una cantidad superior al 10% de los ingresos.* (01/03/2021 ; <https://www.elespejocanario.es/entrevistas/la-red-espanola-de-fundaciones-un-salvavidas-en-tiempos-de-pandemia/>)

- (153) *L'employeur peut, pour une liste précise de professions, **pratiquer** une **déduction** forfaitaire spécifique pour frais professionnels.* (25/02/2021 ; <https://www.urssaf.fr/portail/home/employeur/calculer-les-cotisations/la-base-de-calcul/lassiette-minimale-des-cotisations/la-deduction-forfaitaire-specifi.html>)

Pourtant, nous ne les avons pas incorporés à liste des verbes supports passifs, car ils n'apparaissent en combinaison avec tous les noms de la classe <aide financière : dégrèvement>.

L'actualisation aspectuelle des noms de cette classe s'est avéré très riche dans les deux langes. Ainsi, nous avons repéré des verbes supports aspectuels à valeur inchoative, progressive, terminative, itérative et intensive. En ce qui concerne l'aspect inchoatif, comme nous l'avons déjà signalé dans le cas d'autres classes syntacticosémantiques, nous n'avons repéré que des verbes apparaissant dans des constructions converses : *obtenir* en français, et *conseguir*, *lograr* et *obtener* en espagnol. Nous montrons ci-dessous un exemple pour chaque verbe :

(154) À Cahéroult et à la Sénarderie, les riverains **ont obtenu un abattement** de 1 000 € sur la nouvelle taxe de raccordement. (25/02/2021 ; <https://www.ouest-france.fr/pays-de-la-loire/saint-julien-de-concelles-44450/taxe-de-raccordement-un-elan-de-solidarite-6321495>)

(155) La ACCO cuenta con un programa de clemencia, mediante el cual la entidad y/o directivo que haya participado en un cártel pueden aportar información sobre el mismo, de manera escrita o verbal, a la ACCO con el objetivo de **conseguir una exención total** (o una reducción) de la multa que le correspondería. (01/03/2021 ; https://acco.gencat.cat/web/.content/80_acco/documents/arxius/actuacions/COMUNICAT-WEB-ES.pdf)

(156) Se comprende que el agujero del alfiler por el que ha de pasar la persona física para **lograr la exoneración** de sólo una parte de sus deudas conlleva que sólo un muy reducido grupo de deudores que tengan una composición de su pasivo muy concreto, pueden aventurarse a intentar emprender el viaje previsto en esta llamada ley de segunda oportunidad con la esperanza de llegar a buen puerto habiendo dejado atrás la mochila de la deuda. (01/03/2021 ; https://www.garrigues.com/es_ES/noticia/de-verdad-tenemos-una-ley-de-segunda-oportunidad-0)

(157) **Obtén una rebaja** de hasta 40% en el precio de tu mudanza. (29/07/2023 ; <https://www.mudanzas24.es/servicio-de-mudanza/cajas-mudanza/>)

Quant à l'aspect progressif, en français nous avons repéré les verbes *pérenniser* et *prolonger*, et en espagnol, les verbes *mantener* et *prorrogar* :

(158) Dans son analyse du projet de loi de finances pour 2023, le site Francsjeux révèle que le gouvernement prévoit de **pérenniser l'exonération fiscale** (hors TVA) instaurée en 2014 et appliquée aux grands événements sportifs. (04/10/2022 ; https://www.revue-eps.com/index.php?lg=fr&id_nav=9&id_bloc_liaison=0061&id_fiche=5649)

- (159) *D'une part, il est proposé de **prolonger** le **dégrèvement** existant en faveur des 80 % des ménages les moins favorisés en 2020.* (26/02/2021 ; https://www.senat.fr/amendements/2019-2020/139/Amdt_I-508.html)
- (160) *El Consejo de Ministros aprueba **mantener** la **bonificación** del 50% a los autónomos de Ceuta y Melilla.* (10/01/2023 ; <https://ceutaahora.com/art/12132/el-consejo-de-ministros-aprueba-mantener-la-bonificacion-del-50-a-los-autonomos-de-ceuta-y-melilla>)
- (161) *La Tesorería General de la Seguridad Social, previa solicitud de la empresa, podrá **prorrogar** la **exoneración** reconocida a las empresas por otros 12 meses siempre que resulte acreditado tanto que la empresa sigue cumpliendo los requisitos que determinaron el reconocimiento inicial de la exoneración, como que ha puesto en marcha los compromisos adquiridos en cuanto a la necesaria reinversión en la empresa y el mantenimiento en el empleo de los trabajadores afectados por la suspensión o reducción.* (26/02/2021 ; <https://www.boe.es/buscar/doc.php?id=BOE-A-2014-13249>)

Nous tenons à signaler que parfois ces verbes supports aspectuels progressif introduisent un argument désignant le temps, comme c'est le cas du dernier exemple copié ci-dessus (161) : « por otros 12 meses ».

En ce qui concerne les verbes aspectuels à valeur terminative, en français nous n'avons repéré qu'un seul verbe fonctionnant avec tous les noms de la classe. Il s'agit du verbe *supprimer* :

- (162) *Il faudrait déjà que le gouvernement **supprime** la **ristourne** exceptionnelle sur les carburants au lieu de l'amplifier.* (04/10/2022 ; <https://forums.automobile-propre.com/topic/suivi-des-tarifs-des-superchargeurs-33074/?&page=59>)
- (163) *Au mois d'octobre 2013, la société GBF a indiqué à la société [P] qu'elle entendait **supprimer** l'**escompte** de 2% accordé sur l'achat de ses produits en cas de paiement au comptant.* (04/10/2022 ; <https://justice.pappers.fr/decision/99e7b87c06c7b15314ba8c17fc96758e>)

(164) À la suite des députés, les sénateurs ont abaissé le seuil de l'impôt de solidarité sur la fortune (ISF) à 1,3 million d'euros, abrogé le bouclier fiscal, amélioré le régime d'exonération d'ISF des biens professionnels, modifié le barème des droits de succession perçus en ligne directe et **supprimé** les **réductions d'impôt** sur les donations en fonction de l'âge du donateur. (04/10/2022 ; https://www.lexpress.fr/argent/impots/le-senat-adopte-la-reforme-fiscale-et-renforce-la-taxation-des-contrats-d-assurance-vie_1585275.html)

En revanche, en espagnol, la liste des verbes repérés est plus longue : *anular*, *eliminar*, *quitar*, *retirar* et *suprimir*. Voici quelques exemples :

(165) *Un Gobierno socialista propone **anular** la **reducción fiscal** que tienen los matrimonios por declaración conjunta para fomentar la incorporación de las mujeres al trabajo.* (13/08/2023 ; https://www.infolibre.es/club-infolibre/librepensadores/mano-negra_1_1197188.html)

(166) *El Gobierno baraja **quitar** el **descuento** de 20 céntimos en los combustibles para destinarlo a ayudas a la compra de alimentos.* (17/12/2022 ; <https://elpais.com/economia/2022-12-17/el-gobierno-baraja-quitar-el-descuento-de-20-centimos-en-los-combustibles-para-destinarlo-a-ayudas-a-la-compra-de-alimentos.html>)

(167) *En este nuevo modelo de financiación local, el PSOE propone **suprimir** muchas de las **exenciones** y **bonificaciones**, entre las que se encuentran las que se conceden a los bienes y actividades de entidades religiosas.* (17/12/2022 ; https://www.diariodesevilla.es/espana/PSOE-suprimir-exenciones-fiscales-Iglesia_0_713928924.html)

Parmi les verbes supports aspectuels terminatifs, nous avons repéré des verbes apparaissant dans des constructions conservées. Il s'agit des verbes *perdre* en français et *perder* en espagnol :

(168) *Pour éviter une telle situation et de manière exceptionnelle, il est proposé de permettre aux entreprises de différer les amortissements sans en **perdre** la **déduction fiscale**.* (10/10/2022 ; https://www.assemblee-nationale.fr/dyn/15/textes/115b3145_proposition-loi)

- (169) *Les membres du Conseil constitutionnel pourraient **perdre** une **exonération** partielle de l'impôt sur le revenu dont ils bénéficient.* (10/10/2022 ; https://www.liberation.fr/france/2001/01/11/le-conseil-constitutionnel-pourrait-perdre-l-exoneration_350572/)
- (170) *Coup dur pour les automobilistes français qui viennent juste de **perdre** la **remise** de 20 centimes depuis le 15 novembre dernier qui pouvait encore se reconforter grâce à la ristourne de 10 centimes dans les stations TotalEnergies.* (18/11/2022 ; <https://www.turbo.fr/actualite-automobile/ristourne-totalenergies-la-fin-dans-quelques-jours-une-nouvelle-remise-venir-188138>)
- (171) *En ese caso, debemos identificar al mismo, ya que de no ser así podríamos ver la multa duplicada, incluso triplicada, y **perder** el **descuento** del 50% en caso de que la multa original tuviera esa reducción.* (18/07/2023 ; <https://www.eldia.es/buzzeando/2023/07/18/dgt-no-admite-descuentos-multas-dv-90018404.html>)
- (172) *Su consumo energético le puede hacer **perder** la **deducción** por vivienda habitual.* (18/11/2022 ; <https://www.larazon.es/economia/20210504/ptyfzcdderdj3eqwtgifjpp7m.html>)

En ce qui concerne les verbes supports aspectuels itératifs, nous n'avons repéré que les verbes *cumuler* en français et *acumular* en espagnol. Dans les deux cas, les verbes apparaissent dans des structures converses dans lesquelles le N0 tend à disparaître, comme nous pouvons l'observer dans les exemples suivants :

- (173) *Nombre d'entreprises pourraient **cumuler** des **exonérations** de cotisations sociales actuelles à ce nouveau crédit d'impôt.* (18/11/2022 ; https://www.challenges.fr/monde/competitivite-bernard-thibault-qualifie-d-inacceptable-le-plan-gouvernemental_248920)
- (174) *Par ailleurs, cet amendement permet d'éviter qu'une même personne **cumule** les **déductions fiscales** et contourne, par le biais de plusieurs dons à des partis satellites, le principe du plafond des dons à 7 500 € par parti.*

(18/11/2022 ; https://www.senat.fr/amendements/2011-2012/160/Amdt_38.html)

(175) *Otra de las novedades de la nueva aplicación de Cepsa, es que, ha firmado acuerdos con «cerca de 40 empresas colaboradoras, entre las que destacan Amazon, eDreams, Europcar, Mediamarkt, Pangea y Sprinter» para que los clientes del club GOW puedan **acumular descuentos** de hasta el 10% de sus compras y «conseguir un ahorro anual de más de 300 euros».* (21/04/2023 ; <https://www.elcorreo.com/economia/tu-economia/cepsa-nueva-aplicacion-descuento-precio-gasolina-diesel-ahorro-300-euros-20230421110525-nt.html>)

(176) *La banca **acumula desgravaciones** que le permiten recuperar todo lo que ha pagado en impuestos desde el rescate.* (19/02/2023 ; <https://www.publico.es/economia/banca-acumula-desgravaciones-le-permiten-recuperar-pagado-impuestos-rescate.html>)

L'observation sur corpus nous a permis de repérer des combinaisons des noms d'<aide financière : allègement> avec des verbes supports intensifs à valeur progressive-augmentative et à valeur progressive-atténuative. Quant aux premiers, nous avons trouvé les verbes *accroître* et *augmenter* en français, et *aumentar* et *incrementar* en espagnol :

(177) *Au cours des discussions sur le budget rectificatif 2022, le Gouvernement a annoncé vouloir **accroître** la **remise à la pompe** en la portant à 30 centimes, ce qui pourrait acter implicitement l'abandon de « l'indemnité carburant pour les travailleurs modestes ».* (15/08/2022 ; <https://www.assemblee-nationale.fr/dyn/16/amendements/0017/AN/1096.pdf>)

(178) *Emmanuel Macron a promis d'**augmenter** l'**abattement** sur les successions en ligne directe et en ligne indirecte.* (15/08/2022 ; <https://www.lesechos.fr/economie-france/budget-fiscalite/heritage-cette-baisse-dimpot-dont-le-camp-macron-ne-parle-plus-1413479>)

(179) *Por otra parte, también **hemos aumentado** la **deducción** de 1.200 euros para familias numerosas en 600 euros adicionales por cada hijo a partir del*

tercero (o del segundo en familias numerosas con hijos con discapacidad y en las monoparentales). (04/08/2022 ; <https://www.eleconomista.es/firmas/noticias/9035292/03/18/El-acuerdo-presupuestario-CsPP-buenas-noticias-para-los-espanoles.html>)

(180) *El Gobierno no prevé **aumentar el descuento** en combustible como en Francia.* (01/09/2022 ; https://cadenaser.com/emisora/2017/05/29/radio_ibiza/1496058601_108811.html)

Le nombre de verbes supports intensifs à valeur progressive-atténuative que nous avons repérés est aussi très limité. En français, nous avons trouvé les verbes *limiter* et *plafonner* se combinant avec tous les noms de la classe :

(181) *Le Conseil d'analyse économique préconise de **limiter l'allègement** de charges aux salaires inférieurs à 1,6 fois le smic.* (15/08/2022 ; https://www.bfmtv.com/economie/emploi/emploi-quel-bilan-pour-hollande-et-sa-boite-a-outils_AN-201305300147.html)

(182) *Le gouvernement **aurait pu plafonner la réduction d'impôt** à 500 euros par exemple, plutôt que de supprimer cet avantage fiscal certain et largement plébiscité.* (15/08/2022 ; <https://www.optionreel.com/130-2/>)

En espagnol, nous avons obtenu les verbes *limitar*, *rebajar* et *recortar* :

(183) *Parte del Gobierno quiere **limitar el descuento** a los carburantes para las rentas más altas.* (15/08/2022 ; <https://www.eleconomista.es/economia/noticias/11795972/06/22/Parte-del-Gobierno-quiere-limitar-el-descuento-a-los-carburantes-en-las-rentas-mas-altas.html>)

(184) *ERC pide **recortar la desgravación fiscal** de las declaraciones conjuntas de IRPF de más de 60.000 euros.* (15/08/2022 ; <https://www.negocios.com/erc-pide-recortar-la-desgravacion-fiscal-de-las-declaraciones-conjuntas-de-irpf-de-mas-de-60-000-euros/>)

(185) *Mientras que Garamendi reclamaba no **rebajar las exoneraciones** en las cuotas a los trabajadores que siguieran en ERTE, Escrivá defendía la*

necesidad de incentivar la reincorporación de empleados a la actividad, con mayores descuentos en su cotización. (15/08/2022 ; https://www.lavozdegalicia.es/noticia/economia/2021/05/28/prorroga-erte-supondra-850-millones-cuotas-verano/0003_202105G28P29991.htm)

Nous tenons à signaler que nous avons observé très peu de cooccurrences du verbe *rebajar* avec des noms tels que *rebaja* ou *rebaja fiscal*. Nous comprenons que dans ces cas, il s'agit d'éviter la cacophonie provoquée par la répétition, c'est-à-dire il s'agit plutôt de raisons stylistiques et non pas vraiment de combinatoire. Pour ça, nous avons décidé de garder ce verbe dans la liste de verbes supports intensifs à valeur progressive-atténuative se combinant avec des noms d'<aide financière : dégrèvement>. D'ailleurs, nous avons pu en trouver quelques exemples :

(186) *La última de ellas ha sido **rebajar** la **rebaja** de la tasa de referencia para créditos a un año hasta el 3,85%, como medida de alivio del impacto causado por el coronavirus.* (15/08/2022 ; <https://www.fundssociety.com/es/noticias/markets/economia-china-el-signo-de-esperanza-que-esperan-en-europa-y-estados-unidos/>)

(187) *El Ayuntamiento de Madrid **ha rebajado** la **rebaja fiscal** del 5% al 3,5% en el caso de que el contribuyente realice el pago de este impuesto en dos plazos: junio y noviembre.* (15/08/2022 ; <https://www.pisos.com/aldia/tag/pago-fraccionado/>)

L'observation sur corpus nous a permis d'identifier quelques verbes prédicatifs se combinant habituellement avec des noms d'<aide financière : allègement>, tels que des verbes de <demande>, de <mise en place> et de <refus>. Les verbes de <demande> sont les plus nombreux. Ainsi, en français nous avons repéré les verbes *demandar*, *exigir*, *reclamar* et *solliciter* :

(188) *Si vous avez une résidence dans un autre pays et que vous pouvez le prouver, vous pouvez **demandar** une **exemption** d'impôt sur le revenu d'intérêts dans un certain nombre de pays.* (02/01/2023 ; <https://libredetat.com/6-ventajas-pour-le-touriste-perpetuel/>)

(189) *Une retraitée **exige** une **exonération** de la CSG [Contribution Sociale Généralisée] sur les retraites de moins de 2.000 € tout de suite.*

(02/01/2023 ; <https://www.nicematin.com/faits-de-societe/classe-moyenne-qui-sappauvrit-pouvoir-dachat-ric-voila-ce-que-contient-le-cahier-de-doleances-des-citoyens-de-hyeres-289547>)

(190) À l'occasion de sa conférence de presse mensuelle, Pierre Gattaz vient de **réclamer** des **allègements** de cotisations sociales patronales sur « la totalité du spectre des salaires ». (02/01/2023 ; <https://www.pcf84danielecasanova.fr/top/19>)

(191) Le fait pour un fournisseur d'accorder ou pour un acheteur de **solliciter** un **rabais**, une **remise** ou une **ristourne** en méconnaissance des dispositions du précédent alinéa engage la responsabilité de son auteur et l'oblige à réparer le préjudice causé. (02/01/2023 ; https://www.senat.fr/amendements/2009-2010/437/Amdt_533.html)

En espagnol, les verbes de <demande> se combinant avec tous les noms de la classe sont les mêmes qu'en français, à savoir *exigir*, *pedir*, *solicitar* et *reclamar* :

(192) Argentina propone un plan para **exigir** una **reducción** de intereses en la deuda externa. (13/02/2023 ; https://elpais.com/diario/1987/08/16/economia/556063208_850215.html)

(193) Mediante Institut Valencià de Competitivitat Empresarial (IVACE), es posible **pedir** una **reducción fiscal** por aquellas instalaciones de autoconsumo o energías renovables que se realicen en cualquier momento del año. (13/02/2023 ; <https://www.papernest.es/blog/reduce-el-consumo-y-tu-tarifa-de-luz-descubre-la-energia-fotovoltaica/>)

(194) Aún recuerdo que una vez fui a **solicitar** la **rebaja** económica de la licencia de apertura por una subvención que concedían a emprendedores jóvenes. (13/02/2023 ; <https://www.ceac.es/blog/es-facil-ser-emprendedor>)

(195) Aunque el producto considerado se excluyó de la lista de productos promovidos en la última modificación de la Ley de promoción de las inversiones, una empresa que fabrica el producto considerado todavía puede **reclamar** una **exención fiscal** por un período de cinco años. (13/02/2023 ; <https://www.boe.es/buscar/doc.php?id=DOUE-L-2000-80490>)

Les verbes de <mise en place> se combinant avec tous les noms d'<aide financière : allègement> ont été moins nombreux. En français, nous avons repéré la paire *instaurer* et *instituer*, et en espagnol, *establecer* et *implantar* :

- (196) *La proposition d'instaurer une déduction fiscale pour investissement se heurte à des objections techniques.* (21/08/2022 ; <https://www.senat.fr/questions/base/2000/qSEQ001229910.html>)
- (197) *Dans un entretien au Parisien, le PDG du groupe Patrick Pouyanné affirme que Total pourrait, comme il l'avait fait en 2022, instaurer un rabais à la pompe si le prix de l'essence s'envolait encore une fois.* (08/02/2023 ; https://rnc.bfmtv.com/actualites/economie/total-bat-son-record-de-benefice-et-envisage-une-nouvelle-reduction-sur-le-prix-du-carburant_AV-202302080249.html)
- (198) *En outre, les collectivités territoriales peuvent instituer des exonérations de taxe foncière sur les propriétés non bâties en faveur des terrains, agricoles ou non, plantés en oliviers ou des terrains nouvellement plantés en noyers.* (21/08/2022 ; <https://questions.assemblee-nationale.fr/q12/12-95492QE.htm>)
- (199) *Votre supérieur hiérarchique souhaite instituer une remise pour les meilleurs clients de l'entreprise afin de les fidéliser.* (21/08/2022 ; http://pop27.free.fr/AG1/ressources/Exercices_Chap1.pdf)
- (200) *La Generalitat establece una deducción del 10% para adquirir bicicletas y vehículos eléctricos de movilidad personal.* (13/02/2023 ; <https://www.20minutos.es/noticia/4382312/0/la-generalitat-establece-una-deducccion-del-10-para-adquirir-bicicletas-y-vehiculos-electricos-de-movilidad-personal/>)
- (201) *La movilización, convocada para el 12 de diciembre a las 9h, responde a la decisión del Gobierno de vetar las enmiendas a la Ley de los Presupuestos Generales del Estado para 2023 que presentaron los grupos parlamentarios de UP, PP, ERC y EH Bildu para establecer una exención fiscal para las indemnizaciones previstas en el fondo de compensación para las víctimas del amianto, aprobado recientemente en la Ley 21/2022, de 19*

de octubre. (13/02/2023 ; <https://es.ccoo.cat/noticias/ccoo-convoca-una-concentracion-a-las-puertas-de-la-subdelegacion-del-gobierno-de-tarragona-para-exigir-la-exencion-tributaria-a-las-indemnizaciones-del-fondo-para-las-victimas-damiant/>)

(202) *La última medida para combatirla se recoge en un informe presentado en la capital conquense y consistiría en **implantar reducciones fiscales** en esas zonas para ayudar a repoblar sus territorios.* (13/02/2023 ; <https://www.cmmedia.es/noticias/castilla-la-mancha/reducciones-fiscales-irpf-50-ciento-frenar-despoblacion.html>)

(203) *Las medidas aplicadas por el gobierno, que **implantó descuentos** del 20% en las rentas durante el primer confinamiento, ofreció coberturas para gastos entre julio y septiembre y condonó la deuda a los arrendatarios a partir de septiembre (responsabilizándose de entre el 60% y el 80% de la renta según la tipología de propietario) han ayudado al mantenimiento del sector.* (13/02/2023 ; https://www.ascana.es/wp-content/uploads/2021/07/2020_Ascana_Informe.pdf)

Quant aux verbes prédicatifs de <refus> se combinant avec des noms d'<aide financière : allègement>, nous avons observé une différence significative entre le français et l'espagnol. En français, nous avons repéré plus de verbes de <refus> apparaissant dans des constructions converses, c'est-à-dire des constructions dans lesquelles le N3 (l'éventuel bénéficiaire) occupe la position de sujet. Ainsi, dans les constructions actives, nous avons trouvé seulement le verbe *refuser* :

(204) *Elle soutient que : le service ne peut lui **refuser** un **abattement** sur son bénéfice sur le fondement de l'article 302 nonies du code général des impôts, qui ne vient sanctionner que les retards dans les déclarations de chiffre d'affaires, et non les retards dans les déclarations de résultats.* (11/02/2023 ; <https://www.legifrance.gouv.fr/ceta/id/CETATEXT000031647807>)

(205) *L'administration peut **refuser** une **remise** gracieuse, mais elle doit prendre en compte la situation financière du contribuable et tout autre élément pertinent relatif à la situation du contribuable.* (18/04/2023 ;

<https://www.legifiscal.fr/jurisprudences-fiscales/814-demande-remise-gracieuse-absence-insolvabilite.html>)

Dans des constructions converses, nous avons repéré les verbes *refuser* et *renoncer* à :

(206) *Le péché originel a été de **refuser** une **remise** de dette au départ, car la dette était déjà excessive.* (11/02/2023 ; <https://www.lefigaro.fr/vox/monde/2015/07/10/31002-20150710ARTFIG00267-grece-un-tour-d-horizon.php>)

(207) *Une des idées émises au cours des discussions a été que, peut-être l'Eglise catholique japonaise devrait **renoncer** à l'exemption d'impôt que lui garantit la loi, pour regagner une certaine crédibilité dans la société.* (11/02/2023 ; <https://missionsetrangeres.com/eglises-asie/1997-12-01-leglise-cherche-a-vaincre-la-mefiance-de-plus-en/>)

De plus, la plupart des cooccurrences trouvées avec le verbe *refuser* faisaient partie de construction passives du type *se voir* + *infinitif* :

(208) *La Cour suprême annule le plan d'effacement partiel des emprunts étudiants de Biden, 43 millions d'emprunteurs **se voient refuser** un **allègement** de leur dette.* (03/7/2023 ; <https://www.wsws.org/fr/articles/2023/07/03/fwcm-j03.html>)

(209) *Il appelle à cet égard son attention sur la situation d'une personne âgée qui s'est vu **refuser** la **réduction d'impôt** pour les travaux de couverture de sa maison, au motif que l'administration fiscale considérait cette maison comme résidence secondaire.* (11/02/2023 ; <https://questions.assemblee-nationale.fr/q11/11-58138QE.htm>)

Par contre, en espagnol, les verbes de <refus> apparaissant dans des constructions converses sont moins nombreux que ceux qui apparaissent dans des constructions où le N0 occupe la position de sujet (construction active). En effet, nous avons repéré trois verbes apparaissant dans des constructions actives : *denegar*, *negar*, *rechazar*. Nous l'illustrons à l'aide de quelques exemples :

- (210) Sin embargo, la Agencia Tributaria **le ha denegado** la **deducción** del 50 por ciento del combustible utilizado en dicho vehículo, así como la **deducción** de la cuota soportada en la adquisición del ordenador portátil dedicado única y exclusivamente a la actividad. (28/02/2023 ; <https://espaciopymes.com/noticias/denegacion-de-la-deducción-de-combustible-en-coche-de-empresa/>)
- (211) La Agencia Tributaria **ha negado** el **descuento** de 20 céntimos a la flota de atuneros porque repostan en puertos extranjeros, al igual que tampoco se benefician aquellos transportistas que cargan sus depósitos en otros países europeos, según explica el técnico José María Mollinedo. (28/02/2023 ; <https://www.eitb.eus/es/television/programas/en-jake/videos/detalle/8835144/video-el-descuento-del-carburante-se-aplica-solo-cuando-se-reposta-en-territorio-espanol/>)
- (212) El Tribunal Supremo le dice a Hacienda que no puede **rechazar** la **exención** en IRPF de las rentas obtenidas en el extranjero por ser consejero. (28/02/2023 ; <https://www.economista.es/legal/noticias/11884608/07/22/El-Tribunal-Supremo-le-dice-a-Hacienda-que-no-puede-rechazar-la-exencion-en-IRPF-de-las-rentas-obtenidas-en-el-extranjero-por-ser-consejero-.html>)

Nous tenons à signaler que ces verbes apparaissent souvent dans des constructions impersonnelles avec le pronom *se* :

- (213) El Tribunal Constitucional ha amparado a una familia numerosa a la que **se le había negado** una **reducción fiscal** que le correspondía por no tener el título de familia numerosa que expide la Comunidad de Madrid, aunque sí el Libro de Familia. (28/02/2023 ; <https://infovaticana.com/2015/05/21/el-tc-apoya-a-una-familia-numerosa-a-la-que-se-nego-un-beneficio-fiscal/>)
- (214) Apunta en este caso a Josefina, una vecina de Pino Montano que tiene 426 euros de ingresos y paga un alquiler de 356, pese a lo que **se le ha denegado** una **reducción** de las rentas «alegando que la normativa no lo permite», algo que es «falso». (28/02/2023 ;

https://www.teinteresa.es/teinteresa/Espadas-Zoido-sepulcral-alquileres-VPO_0_1002501223.html)

(215) Además, se **rechaza la exención** de una parte de aquella indemnización, toda vez que se opone una clara jurisprudencia frente al pronunciamiento de la Audiencia Nacional de 8/03/2017 invocado. (28/02/2023 ; <https://serviciostelematicosexhacienda.gob.es/TEAC/DYCTEA/criterio.aspx?id=00/02766/2019/00/0/2&q=s%3D1%26rn%3D%26ra%3D%26fd%3D%26fh%3D%26u%3D%26n%3D%26p%3D%26c1%3D%26c2%3D%26c3%3D%26tc%3D%26tr%3D%26tp%3D%26tf%3D%26c%3D%26pg%3D2>)

En ce qui concerne les verbes apparaissant dans des constructions converses, nous n'avons repéré que le verbe *renunciar* à :

(216) Las y los contribuyentes podrán **renunciar a la exoneración** prevista en este artículo mediante el envío de una comunicación al Departamento Foral de Hacienda y Finanzas. (28/02/2023 ; <https://www.gipuzkoa.eus/es/web/ogasuna/ticketbai/exoneraciones>)

(217) ¿Cómo puede ser que los inversores prefieran los dividendos en efectivo sobre los dividendos en acciones de tal manera que estén dispuestos a **renunciar a una desgravación fiscal masiva** para ellos y la empresa? (28/02/2023 ; https://www.blogscapitalbolsa.com/article/17798/los_inversores_sufren_de_una_irracional_aversion_al_arrepentimiento.html)

Nous présentons ci-dessous une fiche récapitulative de la combinatoire verbale de la classe <aide financière : allègement> :

	FRANÇAIS	ESPAGNOL
Entrées	Abattement, allègement, escompte de caisse, exonération fiscale, ristourne...	Reducción, reducción fiscal, descuento por pronto pago, exención fiscal, rebaja...
Schéma	N0: Ncoll / N1: de	N0: Ncoll / N1: de <montant>,

d'arguments	<montant>, #de N / N2: sur N, #pour N / N3 : à Nhum, Ncoll, #pour Nhum, Ncoll, #en faveur de Nhum, Ncoll	#de N / N2: en N, #por N, #para N / N3 : a Nhum, Ncoll,
Verbes supports appropriés	Accorder, consentir, faire, octroyer, offrir, pratiquer.	Aplicar, conceder, hacer, ofrecer, otorgar, practicar, proponer, realizar.
Verbes supports passifs (N1N0)	Avoir, avoir droit à, bénéficier de, profiter de.	Beneficiarse de, disfrutar de, gozar de, percibir, recibir, tener.
Verbes supports aspectuels inchoatifs	Obtenir.	Conseguir, lograr, obtener.
Verbes supports aspectuels progressifs	Pérenniser, prolonger.	Mantener, prorrogar.
Verbes supports aspectuels terminatifs	Supprimer. Conversif : perdre	Anular, eliminar, quitar, retirar et suprimir. Conversif : perder
Verbes supports aspectuels itératifs	Cumuler.	Acumular.
Verbes supports aspectuels intensifs à valeur progressive-augmentative	Accroître, augmenter.	Aumentar, incrementar.
Verbes supports aspectuels intensifs à valeur progressive-	Limiter, plafonner.	Limitar, rebajar, recortar.

atténuative		
Verbes prédicatifs appropriés de <demande>	Demander, exiger, réclamer, solliciter.	Exigir, pedir, solicitar, reclamar
Verbes prédicatifs appropriés de <mise en place>	Instaurer, instituer.	Establecer, implantar.
Verbes prédicatifs appropriés de <refus>	Refuser. Conversif : refuser, renoncer à.	Denegar, negar, rechazar. Conversif : renunciar a.

Tableau 25 : Fiche descriptive de la classe <aide financière : dégrèvement>

CHAPITRE VIII : CONCLUSIONS GÉNÉRALES

L'objectif de notre travail a été de réaliser une étude linguistique des prédicats nominaux d'<aide financière> en appliquant la théorie des classes d'objets. Nous cherchions à obtenir une description fine et exhaustive qui puisse être utilisée par l'automate sans générer du bruit ni du silence. Finalement, les classes syntacticosémantiques ne se sont pas avérées aussi fines que prévu et la description du lexique des classes obtenues ne peut pas être aussi exhaustive qu'on l'aurait souhaité.

En premier, le premier classement en classes et sous-classes sémantiques a été fait à partir de la combinatoire des noms prédictifs avec des verbes supports aspectuels. Ainsi, nous avons créé de nombreuses sous-classes. Par exemple, dans le cas de la classe <aide financière : allocation>, nous avons distingué <aide financière : allocation ponctuelle>, <aide financière : allocation régulière>, <aide financière : allocation viagère>. En effet, les noms d'<aide financière : allocation ponctuelle> ne se combinent pas avec des verbes supports aspectuels inchoatifs, progressifs ou terminatifs. Pourtant, nous avons observé qu'en Espagne une allocation donnée pouvait être ponctuelle ou régulière en fonction de la région. C'est le cas de la *beca Erasmus*, car dans certaines régions elle est allouée dans un seul versement (<allocation ponctuelle>), dans d'autres régions le versement s'effectue en deux temps, et dans d'autres le versement est mensuel. De plus, nous avons également observé que la modalité de versement pouvait changer au long des années. Tel est le cas, par exemple, du nom *bonificación*. La modalité d'une bonification donnée peut changer au long des années et passer donc de la classe <aide financière : allègement> (une réduction d'un taux à un pourcentage donné sur un montant à payer) à la classe <aide financière : allocation> (une somme d'argent fixe qui est remboursée ou allouée au bénéficiaire). Ainsi, si nous voulions assurer une certaine stabilité et validité à notre classification, la distinction en différentes sous-classes ne pouvait pas être aussi fine. Il a donc fallu supprimer ces classes fondées sur les différences aspectuelles.

Deuxièmement, il a été nécessaire de mettre de côté de nombreux verbes supports, car sinon, il serait impossible de regrouper différentes unités lexicales dans une même classe. Prenons comme exemple les noms *allègement* et *dégrèvement*, appartenant à la classe <aide financière : allègement>. Le *Petit Robert 2024 (version*

numérique) les présente comme étant des synonymes. D'ailleurs, nous montrons ci-dessous la définition d'*allègement* dans ce dictionnaire pour prouver que ces deux unités lexicales sont très proches du point de vue sémantique :

allègement ou **allégement** [alɛʒmɑ̃] nom masculin

ÉTYM. *alegemant* 1177 ◇ de *alléger*



① Diminution de poids, de charge. *Donner de l'allègement à un bateau.*

② Fait ou moyen d'alléger (ce qui constitue une charge trop lourde à supporter). *Demander l'allègement des programmes scolaires. Un allègement fiscal.* → **dégrèvement**. « *Ce poids est donc resté sans allègement sur ma conscience* » (Rousseau). *La vie « n'est pas supportable sans de grands allègements »* (Flaubert).

Figure 30 : Définition d'*allègement* (extrait du *Petit Robert*, 2024 – version numérique)

Pourtant, ces deux unités lexicales ne présentent pas exactement la même combinatoire. Ainsi, *dégrèvement* sélectionne les verbes *imputer* et *accroître*, mais *allègement* ne se combine pas avec ces verbes.

Pour ne citer qu'un autre exemple, dans le cas de la classe <aide financière : allocation>, nous avons repéré au total 106 verbes (des verbes supports et des verbes prédicatifs appropriés) se combinant avec quelques noms de cette classe. Seulement 54 de ces verbes se combinent avec tous les noms de la classe.

Troisièmement, après avoir mené cette étude, nous nous demandons quel serait l'intérêt des classes comprenant très peu de noms, comme dans le cas de la classe <aide financière : donation>. En effet, cette classe ne contient que deux noms en espagnol (*donativo* et *donación*) et deux autres en français (*don* et *donation*). Dans le processus d'élaboration des classes, dans un premier temps, cette classe contenait d'autres synonymes tels que *libéralité*, *aumône*, *charité*, *étrennes*, *générosité* et *largesse* en français, et *liberalidad*, *limosna*, *caridad*, *aguinaldo* et *dádiva* en espagnol. En français, nous sommes partis d'un ensemble de 34 verbes (13 verbes supports appropriés, 3 verbes supports passifs, 13 verbes supports aspectuels et 5 verbes prédicatifs appropriés) ; en espagnol, nous avons testé 44 verbes (11 verbes supports appropriés, 7 verbes supports passifs, 11 verbes supports aspectuels et 12 verbes prédicatifs appropriés). Nous avons vérifié la combinatoire de ces verbes avec tous les noms susceptibles d'appartenir à cette classe. Cette analyse a donné comme résultat qu'en

français seulement les verbes supports *accorder* et *accepter* sont sélectionnés par tous les noms de la classe, et en espagnol, les verbes *conceder*, *dar*, *entregar* et *recibir*. Nous avons observé que *don* et *donation* (en français), et *donativo* et *donación* (en espagnol) ont une combinatoire très similaire, tandis que les autres noms présentent une combinatoire très irrégulière. Il a donc fallu écarter de cette classe tous ces autres noms qui, étant des synonymes ou quasi-synonymes, ont une combinatoire très disparate. Il a donc été impossible également de créer une autre classe spécifique pour ces noms que nous avons supprimés de la classe <aide financière : donation>.

Cependant, nous avons obtenu une ressource qui vient enrichir la description de la combinatoire des noms d'<aide financière>. L'atout des classes syntacticosémantiques réside dans leur efficacité pour le traitement automatique et dans le fait qu'elles procurent une description économique de la langue. Les classes étant des généralisations, elles permettent d'économiser les efforts et d'obtenir une description du lexique d'une langue plus rapidement (tout en étant une tâche de longue haleine). Autrement dit, la tâche de décrire une langue dans le cadre de la méthode des classes d'objets s'avère longue, mais envisageable.

En somme, la méthode des classes d'objets fournit une description riche, efficace et économique de la langue, mais elle ne nous semble pas exhaustive, du moins avec les paramètres proposés par cette approche. C'est pourquoi nous laissons ce champ ouvert pour de futures recherches que nous comptons mener. Nous considérons que pour répondre au souci d'exhaustivité, il serait nécessaire de décrire chaque unité lexicale de manière individualisée.

Ainsi, la recherche que nous avons effectuée nous a menée à faire une réflexion profonde sur les restrictions de sélection et sur l'importance du lexique dans la configuration même de la langue. Les lexies constituent une partie primordiale de la langue. Nous avons observé que les règles de grammaire et de syntaxe doivent être formulées en fonction des lexies. Autrement dit, le lexique d'une langue prime sur sa grammaire.

BIBLIOGRAPHIE

- Abeillé, A. (1995). The flexibility of French idioms: a representation with lexicalized tree adjoining grammar. Dans M. Everaert, D.-J. van der Linden, A. Schenk et R. Schreuder (éds.), *Idioms: Structural and psychological perspectives* (p. 15-42). Psychology Press.
- Abeillé, A. et Barrier, N. (2004). Enriching a French treebank. Dans M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa et R. Silva (éds.), *Proceedings of the fourth international conference on Language Resources and Evaluation (LREC 2004)* (p. 2233-2236). European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2004/>
- Abeillé, A., Clement, L. et Liégeois, J. (2019). Un corpus arboré pour le français. *TAL*, 60, 19-43.
- Abeillé, A., Clement, L. et Toussnel, F. (2003) Building a treebank for French. Dans A. Abeillé (éd.), *Treebanks: building and using parsed corpora* (p. 165–187). Springer.
- Abney, S. P. (1991). Parsing by chunks. Dans R. Berwick, S. Abney et C. Tenny (éds.), *Principle-based parsing* (p. 257-278). Springer.
- Aikins, B. T. S. (1996). Bilingual dictionaries: past, present and future. Dans M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström et C. Rödger Pappmehl (dirs.), *Euralex'96 Proceedings* (p. 515-546). Göteborg University.
- Aitchison, J. (2003). *Words in the mind: an introduction to the mental lexicon* (3^e éd.). Blackwell, Oxford.
- Alarcos Llorach, E. (1994). *Gramática de la lengua española* (8^e éd). Espasa Calpe.
- Alexandrova, O. et Ter-Misanova, S. (1987). *English syntax: collocation, colligation and discourse*. Izd-vo Moskovskogo universiteta.
- Alonso, A. et Henríquez Ureña, P. (1938). *Gramática Castellana*. Losada.
- Alonso Ramos, M. (1994-1995). Hacia una definición del concepto de colocación: de J. R. Firth a I. A. Mel'čuk. *Revista de lexicografía*, 1, 9-8.

- Alonso Ramos, M. (1998). *Étude sémantico-syntaxique des constructions à verbe support* [thèse de doctorat, Université de Montréal].
<https://papyrus.bib.umontreal.ca/xmlui/handle/1866/6777>
- Alonso Ramos, M. (2001a). Construction d'une base de données des collocations bilingue français-espagnol. *Langages*, 143, 5-27.
- Alonso Ramos, M. (2001b). Constructions à verbe support dans des langues SOV. *Bulletin de la Société de linguistique de Paris*, 96(1), 79-106.
- Alonso Ramos, M. (2001c). Détermination, incorporation et phraséologie dans les constructions à verbe support. Dans X. Blanco, P.-A. Buvet et Z. Gavriilidou (éds.), *Détermination et Formalisation. Lingvisticae Investigationes Supplementa*, 23 (p. 51-65). John Benjamins Publishing Company.
- Alonso Ramos, M. (2004b). *Las construcciones con verbo de apoyo*. Visor Libros.
- Alonso Ramos, M. et Wanner, L. (2007). Collocation chains: how to deal with them? Dans K. Gerdes, T. Reuther et L. Wanner (éds.), *Proceedings of the Third International Conference on Meaning-Text Theory* (p. 11-20). Wiener Slawistischer Almanach.
- Altenberg, B. et Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2), 173-195.
- Anscombe, J.-C. (2011). Figement, idiomatité et matrices lexicales. Dans J.-C. Anscombe et S. Mejri (dirs.), *Le figement linguistique : la parole entravée* (p. 17-40). Honoré Champion.
- Anscombe, J.-C. (2019). Figement, lexique et matrices lexicales. *Cahiers de lexicologie*, 114, 119-147.
- Anscombe, J.-C. et Mejri, S. (dirs.) (2011). *Le figement linguistique : la parole entravée*. Honoré Champion.
- Antidote Web* [en ligne]. Druide informatique. <https://antidote.app>.
- Apresjan, J. (1973). Regular Polysemy. *Linguistics*, 142, 5-32.
- Assal, A. (1994). La métaphorisation terminologique. *Terminologie et traduction*, 2, 235-242.

- ATILF (2019). Morphalou. ORTOLANG (Open Resources and TOols for LANGuage).
www.ortolang.fr.
- Azarova, I., Mitrofanova, O., Sinopalnikova, A., Yavorskaya, M. et Oparin, I. (2002). Russnet: building a lexical database for the Russian language. Dans *Proceedings of the Workshop on WordNet structures and standardization, and how these affect WordNet application and evaluation. Las Palmas de Gran Canaria* (p. 60–64). <http://www.lrec-conf.org/proceedings/lrec2002/pdf/ws10.pdf#page=66>
- Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, 9, 5-16.
- Badia, T. (1994). *Aspectes del sintagma nominal en català des de la perspectiva de la traducció automàtica*. Publicacions de l'Abadia de Montserrat.
- Badia, T. et Colominas, C. (1997). *The Predicate-Argument Structure*. Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Baker, C., Ellsworth, M., Petruck, M. et Swayamdipta, S. (2018). Frame semantics across languages: towards a multilingual FrameNet. *Proceedings of the 27th international conference on Computational Linguistics: tutorial abstracts*, (p. 9–12). Association for Computational Linguistics. <https://framenet2.icsi.berkeley.edu/docs/ComboBib2.pdf>
- Baker, C., Fillmore, C. et Lowe, J. (1998). The Berkeley FrameNet project. Dans *COLING-ACL'98: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics* (p. 86–90). Université de Montreal. <https://aclanthology.org/C98-1013.pdf>
- Balboni, P. (1982). Le microlingue : considerazione teorische. *Scuola e lingue moderne*, 20, 107-111.
- Balliu, C. (2018). Traduire un texte médical, une tension entre mots et termes. Dans I. Sfar et P.-A. Buvet (dirs.), *La phraséologie entre fixité et congruence. Hommage à Salah Mejri* (p. 9-19). L'Harmattan.
- Bally, C. (1909/1951). *Traité de stylistique française* (vol. 1). Librairie Georg. & Cie./Librairie Klincksieck.

- Balvet, A., Barque, L., Condette, M. H., Haas, P., Huyghe, R., Marin, R. et Merlo, A. (2011a). La ressource Nomage. Confronter les attentes théoriques aux observations du comportement linguistique des nominalisations en corpus. *Traitement Automatique des Langues*, 52(3), 129-152.
- Balvet, A., Barque, L., Condette, M. H., Haas, P., Huyghe, R., Marín, R. et Merlo, A. (2011b). Nomage: an electronic lexicon of French deverbal nouns based on a semantically annotated corpus. Dans *Proceedings of the first international Workshop on Lexical Resources, WoLeR* (p. 8-15). http://alpage.inria.fr/~sagot/pub/WoLeR_2011_proceedings.pdf
- Balvet, A., Barque, L., Condette, M.-H., Haas, P., Huyghe, R., Jugnet, A., Marín, R. et Merlot, A. (2010). *Le projet Nomage. Description des noms déverbaux du français*. STL, Université de Lille 3. <https://hal.science/hal-02190190/document>
- Baránov, A. N. et Dobrovol'skij, D. O. (1998). Idiomaticidad e idiomatismos. Dans J. Luque Durán et A. Pamies Bertrán (éds.), *Léxico y fraseología* (p. 19-42). Granada Lingvistica y Método Ediciones.
- Bárdosi, V. (1989). Un ange passe : contribution à l'étymologie d'une locution. Dans G. Gréciano (éd.), *Europhras 88 : phraséologie contrastive. Actes du Colloque International Klingenthal-Strasbourg, 12-16 mai 1988* (p. 7-16). Université des Sciences Humaines.
- Bardovi-Harlig, K. (2012). Formulas, routines, and conventional expressions in pragmatics research. *Annual Review of Applied Linguistics*, 32, 206-227.
- Baroni, M., Bernardini, S., Ferraresi, A. et Zanchetta, E. (2009). The WaCky Wide Web: a collection of very large linguistically processed Web-crawled corpora. *Journal of Language Resources and Evaluation*, 43(3), 209-226.
- Baroni, M., Murphy, B., Barbu, E. et Poesio M. (2010). Strudel: a corpus-based semantic model based on properties and types. *Cognitive Science*, 34(2), 222-254.
- Barque, L. et Polguère, A. (2009). Structuration et balisage sémantique des définitions du *Trésor de la Langue Française informatisé* (TLFi). Dans D. Beck, K. Gerdes, J. Milicevic et A. Polguère, *Proceedings of the Fourth International Conference*

on *Meaning-Text Theory (MTT'09)* (p. 35-45).
<http://olst.ling.umontreal.ca/pdf/ProceedingsMTT09.pdf>

- Barque, L., Nasr, A. et Polguère A. (2010). From the definitions of the *Trésor de la Langue Française* to a semantic database of the French language. Dans A. Dykstra et T. Schoonheim (éds), *Proceedings of the 14th Euralex International Congress* (p.245-252). <https://hal.science/hal-00509289v1/document>
- Barrios Rodríguez, M.A. (2007). Dictionarios combinatorios del español: diferencias y semejanzas entre "Redes" y "Práctico". *RedEle. Red electrónica de Didáctica del Español como Lengua Extranjera*, 11, 1 - 14.
- Bartsch, S. (2004). *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Narr.
- Bauer, L. (1983). *English word-formation*. Cambridge University Press.
- Beauchesne, J. (2001). *Dictionnaire des cooccurrences*. Guérin.
- Beccaria, G. L. (1973). *I linguaggi settoriali in Italia*. Bompiani.
- Béguelin, M.J. (2000). *De la phrase aux énoncés : grammaire scolaire et descriptions linguistiques*. De Boeck-Duculot.
- Béjoint, H. (2007). Informatique et lexicographie de corpus : les nouveaux dictionnaires. *Revue française de linguistique appliquée*, 23, 7-23.
- Beliakov, V. et Mejri, S. (éds.). (2015). *Stéréotypie et figement. À l'origine du sens*. Presses Universitaires du Midi.
- Bello, A. (1847/2004). *Gramática de la lengua castellana destinada al uso de los americanos* (édition de 2004). Edaf.
- Bennet, P. (1997). Compositionnalité et figement des locutions : étude comparative du français et de l'anglais. Dans P. Fiala, P. Lafon et M.-F. Piguet (éds.), *La locution : entre lexique, syntaxe et pragmatique. Identification en corpus, traitement, apprentissage* (p. 11-77). Klincksieck,
- Benson, M. (1985). Lexical combinability. *Papers in Linguistics*, 18(1-4), 3-15.

- Benson, M., Benson, E. et Ilson, R. (1986). *The BBI combinatory dictionary of English. A guide to word combinations*. John Benjamins Publishing Company.
- Benveniste, E. (1966/1974). *Problèmes de linguistique générale* (vol. 1 et vol. 2). Gallimard.
- Benveniste, E. (1967). Fondements syntaxiques de la composition nominale. Dans E. Benveniste (éd.), *Problèmes de linguistique générale* (vol. 2, 145-176). Gallimard.
- Berland, S. et Grabar, N. (2002). Assistance automatique pour l'homogénéisation d'un corpus Web de spécialité. Dans A. Morin et P. Sébillot, *Actes des 6èmes Journées internationales d'analyse statistique des données textuelles (JADT 2002)* (p. 125–136).
- Berruto, G. (1997). *Sociolinguistica dell'italiano contemporaneo*. Nuova Italia Scientifica.
- Bestgen, Y. (2017). Évaluation de mesures d'association pour les bigrammes et les trigrammes au moyen du test exact de Fisher. Dans I. Eshkol et J.-Y. Antoine (éds.), *Actes des 24^e Conférence sur Traitement Automatique des Langues Naturelles* (vol. 2, p. 10-18). ATALA
- Biber, D. (2009). A corpus-driven approach to formulaic language: multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14, 381-417.
- Biber, D., Conrad, S., et Reppen, R. (1998/2006). *Corpus linguistics. Investigating language structure and use*. Cambridge University Press.
- Bierwisch, M. (1990). Event nominalizations: proposals and problems. *Acta Linguistica Hungarica*, 40(1-2), 19-84.
- Bittar, A. (2010). *Building a TimeBank for French: a reference corpus annotated according to the ISOTimeML standard* [thèse de doctorat, Université Paris-Diderot].
- Bittar, A., Amsili, P. et Denis, P. (2011). French TimeBank: un corpus de référence sur la temporalité en français. Dans *TALN 2011 - Traitement Automatique des Langues Naturelles* (vol. 1, p. 259-270). Laboratoire d'Informatique de Robotique et de Microélectronique.

- Bittar, A., Amsili, P., Denis, P. et Danlos, L. (2011). French TimeBank: an ISO-TimeML annotated reference corpus. Dans *Proceedings of the 49th annual meeting of the Association for Computational Linguistics: human language technologies (ACL'11)* (p. 130-134). The Association for Computational Linguistics. <https://aclanthology.org/P11-2023.pdf>
- Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A. et Fellbaum, C. (2006). Introducing the Arabic WordNet project. Dans P. Sojka, K.-S. Choi, C. Fellbaum et P. Vossen, *Proceedings of the 3rd Global WordNet Conference, Jeju Island, Korea, South Jeju, January 22-26, 2006* (p. 295-300). <https://research.vu.nl/ws/files/151566101/arabic>
- Blanco Escoda, X (1998). Les déterminants nominaux figés. Perspective contrastive espagnol-français. Dans S. Mejri, G. Gross, A. Clas et T. Baccouche (éds.), *Le figement lexical. Actes des Premières Rencontres Linguistiques Méditerranéennes, Tunis, 17-18 septembre 1998* (p. 19-34).
- Blanco Escoda, X. (1997). Un dictionnaire électronique des déterminants nominaux en espagnol, *BULAG, numéro spécial Actes du Colloque FRACTAL 97* (p. 45-54). Université de Franche-Comté.
- Blanco Escoda, X. (1999). *Lexicographie bilingüe français-espagnol et classes d'objets*. Servei de Publicacions Universitat Autònoma de Barcelona.
- Blanco Escoda, X. (2000). Verbos soporte y clases de predicados en español. *Lingüística española actual*, 22(1), 99-118.
- Blanco Escoda, X. (2013). Équivalents de traduction pour les pragmatèmes dans la lexicographie bilingue français-espagnol. *Lexicographica*, 29, 5-28.
- Blanco Escoda, X. et Buvet, P.-A. (1998). El tratamiento de los determinantes en procesamiento automático del lenguaje natural. *Terminologie et traduction*, 3, 162-180.
- Blanco Escoda, X. et Buvet, P.-A. (1999). À propos de la traduction automatique des déterminants de l'espagnol et du français. *Meta : Journal des traducteurs/Meta: Translators' Journal*, 44(4), 525-545.

- Blanco Escoda, X. et Buvet, P.-A. (2000). De l'analyse syntactico-sémantique du lexique à la traduction automatique. *BULAG*, 25, 69-87.
- Blanco Escoda, X. et Buvet, P.-A. (2004). Verbes supports et significations grammaticales : implications pour la traduction espagnol-français. *Lingvisticae Investigationes*, 27(2), 327-342.
- Blanco Escoda, X. et Catena Rodulfo, À. (2015). Équivalents de traduction polylexicaux dans un dictionnaire d'adjectifs espagnol-français. *Estudios de traducción*, 5, 109-118.
- Blanco Escoda, X. et Mejri, S. (2018). *Les pragmatèmes*. Classiques Garnier.
- Blanco Escoda, X., Buvet, P.-A. et Gavriilidou, Z. (éds.). (2001). *Détermination et Formalisation. Lingvisticae Investigationes Supplementa*, 23. John Benjamins Publishing Company.
- Blumenthal, P. et Hausmann, F. J. (éds.). (2006). *Collocations, corpus, dictionnaires [= Langue française*, 150].
- Bojanowsky, P., Grave, E., Joulin A. et Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics (TACL)*, 5, 135–146.
- Bolly, C. (2008). *Les unités phraséologiques : un phénomène linguistique complexe ?* [thèse de doctorat, Université Catholique de Louvain]. https://dial.uclouvain.be/pr/boreal/object/boreal:19625/datastream/PDF_08/view
- Bolly, C. (2010). Flou phraséologique, quasi-grammaticalisation et pseudo marqueurs de discours: un no man's land entre syntaxe et discours ? *Linx*, 62-63, 11-38.
- Bolly, C. (2011). *Phraséologie et collocations. Approche sur corpus en français L1 et L2*. Peter Lang.
- Boons, J.-P., Guillet, A. et Leclère, C. (1976). *La structure des phrases simples en français : classes de constructions intransitives*. Librairie Droz.
- Boons, J.-P., Guillet, A. et Leclère, C. (1976b). *La structure des phrases simples en français : classes de constructions transitives*. Librairie Droz.

- Bosque, I. (1983). Clases de nombres comunes. Dans *Serta philologica. F. Lázaro Carreter: natalem diem sexagesimum celebranti dicata* (vol. 1, 75-88). Cátedra.
- Bosque, I. (1990). *Las categorías gramaticales. Reacciones y diferencias*. Editorial Síntesis.
- Bosque, I. (2001a). On the weight of light verb predicates. Dans J. Herschenson, K. Zagana et E. Mallén (éds.), *Features and Interfaces in Romance* (p. 23-38). John Benjamins Publishing Company.
- Bosque, I. (2001b). Sobre el concepto de 'colocación' y sus límites. *Lingüística española actual*, 23(1), 9-40.
- Bosque, I. (dir.). (2004) *Redes. Diccionario combinatorio del español contemporáneo*. SM.
- Bosque, I. (dir.). (2006). *Diccionario combinatorio práctico del español contemporáneo. Las palabras en su contexto*. SM.
- Bosque, I. et Demonte, V. (dirs.). (1999). *Gramática descriptiva de la lengua española* (vol. 1). Espasa Calpe.
- Bourigault, D. et Slodzian, M. (1999). Pour une terminologie textuelle. *Terminologies nouvelles*, 19, 29-32.
- Bouveret, M. et Fillmore, C. (2008). Matching verbo-nominal constructions in FrameNet with lexical functions in MTT. Dans P. Bataner et J. de Cesaris (éds.), *De Lexicographia. Actes del I Symposium Internacional de Lexicografía* (p. 297-308). IULA et Edicions Petició.
- Braasch, A. (1988). Zur lexikographischen kodifizierung von phrasemen in einsprachigen deutschen wörterbücher aus der sicht eines ausländischen wörterbuchbenutzers. Dans K. Hyldgaard-Jensen et A. Zettersten (éds.), *Symposium on Lexicography II. Proceedings of the Second International Symposium on Lexicography, May, 16-17, 1984 at the University of Copenhagen* (p. 83-100). Max Niemeyer.
- Bréal, M. (1897/1976). *Essai de sémantique (Science des significations)*. Hachette.
- Brill, E. (1994). Some advances in rule-based part of speech tagging. Dans B. Hayes-Roth et R. E Korf, *Proceedings of the Twelfth National Conference on Artificial*

- Intelligence (AAAI-94), Seattle, WA, USA, July 21 - August 4* (p. 722-727). AAAI Press / The MIT Press.
- Budin, G. (2001). A critical evaluation of the state-of-the-art of terminology theory. *Itf Journal*, 12(1-2), 7-23.
- Buitrago Jiménez, A. (2002). *Diccionario de dichos y frases hechas*. Espasa-Calpe.
- Burger, H. (1973). *Idiomatik des Deutschen*. Max Niemeyer.
- Burger, H. (1982). Klassifikation: Kriterien. Probleme, Terminologie. Dans H. Burger, A. H. Buhofer, A. Sialm, B. Eriksson et J. Häusermann, *Handbuch der Phraseologie* (p. 20-60). Walter de Gruyter.
- Burger, H. (1998). *Phraseologie. Eine einföhrung am beispiel des Deutschen*. Erich Schmidt.
- Burger, H. (2007). Semantic aspects of phrasemes. Dans H. Burger, D. Dobrovolskij, P. Kühn et N. R. Norrick (éds.), *Phraseologie / Phraseology. Ein internationales handbuch zeitgenössischer forschung / An international handbook of contemporary research* (p. 90-109). Walter de Gruyter.
- Burger, H., Dobrovolskij, D., Kühn, P. et Norrick, N. R. (éds.) (2007). *Phraseologie / Phraseology. Ein internationales handbuch zeitgenössischer forschung / An international handbook of contemporary research*. Walter de Gruyter.
- Buvet, P.-A. (1993). *Les déterminants nominaux quantifieurs* [thèse de doctorat, Université Sorbonne Paris Nord].
- Buvet, P.-A. (1996). Levées d'ambiguïté et domaines d'emploi. *Bulletin de linguistique appliquée et générale (BULAG)*, 21, 63-75.
- Buvet, P.-A. (1998). Détermination et classes d'objets. *Langages*, 131, 91-102.
- Buvet, P.-A. (2002). Analyse de compléments du nom en termes de classes d'objets. *Le français moderne*, 70(2), 187-210.
- Buvet, P.-A. (2008a). Détermination et figement au regard de la traduction. *Meta : Journal des traducteurs/Meta : Translators' Journal*, 53(2), 333-364.

- Buvet, P.-A. (2008b). Vers l'élaboration d'un dictionnaire unique des prédicats du français : DEESSE Dictionnaire Électronique Syntactico-Sémantique. *Cahiers du CENTAL* 5, 23-42.
- Buvet, P.-A. (2009a). Des mots aux emplois: la représentation lexicographique des prédicats. *Le français moderne* 77, 83-96.
- Buvet, P.-A. (2009b). Quelles procédures d'étiquetage pour la gestion de l'information textuelle électronique. *L'Information Grammaticale*, 122, 40-48.
- Buvet, P.-A. (2009c). *Remarques sur la détermination du français* [mémoire d'habilitation à diriger des recherches, Université Sorbonne Paris Nord].
- Buvet, P.-A. (2013a). Aide à la rédaction : le système TAEMA. De l'intérêt d'exploiter des ressources lexicales en linguistique informatique. *Studi de lingüística*, 3, 187-204.
- Buvet, P.-A. (2013b). Collocation, restriction de sélection et prédication. *Cahiers de lexicologie*, 102, 169-184.
- Buvet, P.-A. (dir.). (2002). *La détermination au regard de la diversité lexicale. Langages*, 145.
- Buvet, P.-A. et Blanco, X. (2000). De l'analyse syntactico-sémantique du lexique à la traduction automatique. *Bulletin de linguistique appliquée et générale (BULAG)*, 25, 69-87.
- Buvet, P.-A. et Grezka, A. (2007). Élaboration d'outils méthodologiques pour décrire les prédicats du français. *Lingvisticæ Investigationes*, 30(2), 217-245.
- Buvet, P.-A. et Grezka, A. (2009). Les dictionnaires électroniques du modèle des classes d'objets. *Langages*, 176, 63-79.
- Buvet, P.-A. et Issac, F. (2006). TAEMA : Traitement Automatique de l'Écriture de Mots Affectifs. Dans P. Mertens, C. Fairon, A. Dister et P. Watrin (éds.), *Verbum ex machina. Actes de la 13^e conférence sur le traitement automatique des langues naturelles (TALN 2006)* (vol. 1, p. 856-867). UCL Presses Universitaires de Louvain.
- Buvet, P.-A., Cartier, E., Issac, F. et Mejri, S. (2007). Dictionnaires électroniques et étiquetage syntactico-sémantique. Dans N. Hathout et Ph. Muller (éds.), *Actes*

des 14e journées sur le Traitement Automatique des Langues Naturelles (p. 239-248). IRIT Press.

- Buvet, P.-A., Cartier, E., Issac, F., Mathieu-Colas, M., Mejri, S. et Madiouni, Y. (2009). Morfetik, ressource lexicale pour le TAL. Dans A. Nazarenko et T. Poibeau, *Actes de la 16^e conférence sur le traitement automatique des langues naturelles (TALN 2009). Articles courts* (p. 217-226). ATALA
- Buvet, P.-A., Girardin, Ch., Gross, G. et Groud, C. (2005). Les prédicats d'<affect>. *Revue de linguistique et de didactique des langues (Lidil)*, 32, 123-143.
- Cabré, M. T. (1993). *La terminología: teoría, metodología, aplicaciones* (traduit par C. Tebé). Antártida/Empúries.
- Cabré, M. T. (1998/1999). Do we need an autonomous theory of terms? *Terminology*, 5(1), 5-19.
- Cabré, M. T. (1999). *La terminología. Representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Institut Universitari de Lingüística aplicada, Universitat Pompeu Fabra
- Cabré, M. T. (2000). Terminologie et linguistique: la théorie des portes. *Terminologies nouvelles. Terminologie et diversité culturelle*, 21, 10-15.
- Cabré, M. T. (2003). Theories of terminology: their description, prescription and explanation. *Terminology*, 9(2), 163-200.
- Cabré, M. T. (2004). ¿Lenguajes especializados o lenguajes para propósitos específicos? *Revista Foro Hispánico*, 26, 19-34.
- Cabré, M. T. (2005). Lexicographie versus terminographie : comment les technologies déplacent leur affrontement. Dans H. Béjoint et F. Maniez (éds.), *De la mesure dans les termes* (p. 189-211). Presses Universitaires de Lyon.
- Calero Vaquera, M. L. (1986). *Historia de la gramática española (1847-1920)*. Gredos.
- Candito, M., Amsili, P., Barque, L., Benamara, F., de Chalendar, G., Djemaa, M., Haas, P., Huyghe, R., Yannick Mathieu, Y., Muller, P., Sagot, B. et Vieu, L. (2014). Developing a French FrameNet: methodology and first results. Dans N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, J. Odiijk et S. Piperidis (éds.), *Proceedings of the Ninth International Conference*

- on Language Resources and Evaluation (LREC'14)* (p. 1372-1379). European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2014/index.html>
- Cantera, J. (2007). *Diccionario de fraseología española: locuciones, idiotismos, modismos y frases hechas usuales en español*. Abada.
- Cardiot, P. et Habert, B. (1997). Aux sources de la polysémie nominale. *Langue française*, 113, 3-11.
- Carneado Moré, Z. (1983). *Consideraciones sobre la fraseografía*. Dans Z. Carneado et A. M. Tristá (éds.), *Estudios de fraseología* (p. 39-46). Academia de Ciencias de Cuba.
- Carneado Moré, Z. et Tristá Pérez, A. M. (1983). *Estudios de fraseología*. Academia de las Ciencias de Cuba.
- Carter, R. (1989). *Vocabulary: applied linguistic perspectives*. Allen and Unwin.
- Cartier, E. (2009). Étiquetage sémantique des textes : état des lieux, éléments de modélisation. *L'Information Grammaticale*, 122, (19-29).
- Cartier, E. (2011). Néologie et description linguistique pour le TAL. *Langages*, 183, 105-117.
- Cartier, E. et Issac, F. (2009). Mot et traitement automatique des langues. *Le français moderne*, 77(1), 145-160.
- Casares, J. (1950/1992). *Introducción a la lexicografía moderna*. Consejo Superior de Investigaciones Científicas.
- Català Guitart, D. (2000). Structuration syntactico-sémantique du lexique de la bourse. Dans M. L. Casal Silva, G. Conde Tarrío, J. Lago Garabatos, L. Pino Serrano et N. Rodríguez Pedreira. (éds.), *La lingüística francesa en España camino del s. XXI* (p. 253-261). Arrecife Editorial.
- Català Guitart, D. (2003). *Les adverbos compostos. Approches contrastives en linguistique appliquée* [thèse de doctorat, Universitat Autònoma de Barcelona]. <https://www.tdx.cat/handle/10803/4924#page=1>
- Cavalla, C. (2009). La phraséologie en classe de FLE. *Les Langues Modernes*, 1, 1-12. <https://www.aplv-languesmodernes.org/spip.php?article2292>

- Cavalla, C. (2018). Lexique transdisciplinaire et enseignement aux étudiants allophones. Dans A. Tutin et M.-P. Jacques (dirs.), *Lexique transversal et formules discursives des sciences humaines* (p. 191-214). ISTE Editions.
- Cavalla, C. et Crozier, E. (2005). *Émotions - Sentiments*. PUG.
- Cavalla, C. et Grossmann, F. (2005). Caractéristiques sémantiques de quelques "Noms scientifiques" dans l'article de recherche en français. Dans T. Kin (dir.), *Akademisk prosa*, (p. 47-59). Skrifter fra KIAP Romansk institutt, Universitetet i Bergen.
- Cavalla, C. et Legallois, D. (2020). Caractériser et identifier les unités phraséologiques pour leur enseignement. Dans *Action Didactique. Revue internationale de didactique du français*, 6, 12-30. <http://www.univ-bejaia.dz/action-didactique/pdf/ad6/CavallaLegallois.pdf>
- Cavalla, C. et Loiseau, M. (2014). Scientext comme corpus pour l'enseignement. Dans F. Grossmann et A. Tutin (dirs.), *L'écrit scientifique : du lexique au discours. Autour de Scientext* (p. 163-180). Presses universitaires de Rennes.
- Cavalla, C., et Legallois, D. (2020). Caractériser et identifier les unités phraséologiques pour leur enseignement. *Action Didactique*, 6, 12-30.
- Čermak, F. (1988). On the substance of idioms. *Folia Lingüística*, 22(3-4), 413-438.
- Čermak, F. (2001). Substance of idioms: perennial problems, lack of data or theory?, *IJL*, 14(1), 1-20.
- Chafe, W. L. (1968). Idiomaticity as an anomaly in the chomskyan paradigm. *Foundations of Language*, 4, 109-127.
- Charaudeau, P. (1992). *Grammaire du sens et de l'expression*. Hachette.
- Charaudeau, P. et Maingueneau, D. (dir.) (2002). *Dictionnaire d'analyse du discours*. Le Seuil.
- Choi, H.-S. (2022). État de l'art: liage de ressources lexicales du français. Dans Y. Estève, T. Jiménez, T. Parcollet et M. Z. Boito (éds.), *Actes de la 29^e conférence sur le Traitement Automatique des Langues Naturelles. Volume 2 : 24^e Rencontres Etudiants Chercheurs en Informatique pour le TAL (RECITAL)* (p. 55-68). <https://aclanthology.org/2022.jeptalnrecital-recital.0.pdf>

- Choi, H.-S., Fort, K., Guillaume, B. et Constant, M. (2023). Des ressources lexicales du français et de leur utilisation en TAL : étude des actes de TALN. Dans C. Servan et A. Vilmat (éds.), *Actes de CORIA-TALN 2023* (vol. 2, p. 23-36). https://coria-taln-2023.sciencesconf.org/data/proceedings_TALN_courts.pdf
- Church, K. W. et Hanks, P. (1990). Word association norms, mutual information and lexicography. *Computational Linguistics*, 16(1), 22–29.
- Clas, A. et Gross, G. (1998). Classes de figement des locutions verbales. Dans S. Mejri (éd.), *Le figement lexical. Actes de la 1re RLM, Tunis, septembre 1998* (p. 11-18). CERES.
- Clear, J. (1994). I can't see the sense in a large corpus. Dans F. Kiefer, G. Kiss et J. Pajzs, *Papers in computational lexicography. Actes de Complex'99* (p. 33-48). Linguistic Institute Hungarian Academy of Sciences.
- Clément, L., Sagot, B. et Lang, B. (2004). Morphology based automatic acquisition of large-coverage lexica. Dans M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa et R. Silva (éds.), *Proceedings of the fourth international conference on Language Resources and Evaluation (LREC 2004)* (p. 1841-1844). European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2004/>
- Colominas, C. (2001). *La representació semàntica de les construccions de suport des d'una perspectiva multilingual* [thèse de doctorat, Universitat Autònoma de Barcelona]. <https://www.tdx.cat/handle/10803/5252#page=1>
- Colson, J.-P. (1992). Ébauche d'une didactique des expressions idiomatiques en langue étrangère. *Terminologie et Traduction*, 2/3, 165-179.
- Colson, J.-P. (2003). Corpus linguistics and phraseological statistics: a few hypotheses and examples. Dans H. Burger, A. Häcki Buhofer, A. et G. Gréciano (éds.), *Flut von Texten – Vielfalt der kulturen. Ascona 2001 zu methodologie und kulturspezifik der phras eologie* (p. 47-59). Schneider Verlag Hohengehren.
- Colson, J.-P. (2007). The World Wide Web as a corpus for set phrases. Dans H. Burger, D. Dobrovolskij, P. Kühn et N. R. Norrick (éds.), *Phraseologie / Phraseology. Ein internationales handbuch zeitgenössischer forschung / An international handbook of contemporary research* (vol. 2, p. 1071-1077). Walter de Gruyter.

- Colson, J.-P. (2008). Cross-linguistic phraseological studies: an overview. Dans S. Granger et F. Meunier (éds.), *Phraseology. An interdisciplinary perspective* (p. 191-206). John Benjamins Publishing Company.
- Colson, J.-P. (2010). La traduction spécialisée basée sur les corpus : une expérience dans le domaine de l'informatique. *Synergies Tunisie*, 2, 115-123.
- Colson, J.-P. (2012). Phraséologie contrastive et linguistique de corpus : outils récents et perspectives. Dans A. Pamies, J. M. Pazos Breña et L. Luque Nadal (éds.), *Phraseology and discourse: cross linguistic and corpus-based approaches [Phraséologie und Parömiologie, 29]* (p. 1-9). Schneider Verlag Hohengehren.
- Conca i Martínez, M. (1999). Relacions d'inclusió entre unitats fraseològiques. *Paremia*, 8, 137-141.
- Conca i Martínez, M. et Guía, J. (2000). La idiomaticitat de les unitats fraseològiques (Ufs) en relació amb el sistema conceptual metafòric d'una llengua i cultura. Aplicació diacrònica y contrastiva. *Cahiers du P.R.O.E.M.I.O.*, 3, 31-48.
- Condamines, A. (1995). Terminology. New needs, new perspectives. *Terminology*, 2(2), 219-238.
- Conklin, K. et Schmitt, N. (2008). Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.
- Constant, M. (2003). *Grammaires locales pour l'analyse automatique de textes : méthodes de construction et outils de gestion* [thèse de doctorat Université Paris-Est Marne-la-Vallée].
- Cori, M. (2008). Des méthodes de traitement automatique aux linguistiques fondées sur les corpus. *Langages*, 171, 95-110.
- Corpas Pastor, G. (1995). *Un estudio paralelo de los sistemas fraseológicos del inglés y del español* [thèse de doctorat, Universidad Complutense de Madrid].
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Gredos.
- Corpas Pastor, G. (1998). Criterios generales de clasificación del universo fraseológico de las lenguas con ejemplos en español y en inglés. Dans M. Alvar Ezquerro et

- G. Corpas Pastor (coords.), *Diccionarios, frases, palabras* (p. 157-187). Universidad de Málaga.
- Corpas Pastor, G. (2000). *Las lenguas de Europa: estudios de fraseología, fraseografía y traducción*. Editorial Comares.
- Corpas Pastor, G. (2001a). Corrientes actuales de la investigación fraseológica en Europa. *Euskera: Euskaltzaindiaren lan eta agiriak*, 46(1), 21-48.
- Corpas Pastor, G. (2003). *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Iberoamericana.
- Corpas Pastor, G. (2013). Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas. Dans I. Olza et E. Manero (éds.), *Fraseopragmática* (p. 335-374). Frank & Timme.
- Corpas Pastor, G. et Colson J.-P. (éd.) (2020). *Computational phraseology*. John Benjamins Publishing Company.
- Corpas Pastor, G. et Mitkov, R. (éds.). (2019). *Computational and corpus-based phraseology: third international conference, Europhras 2019, Malaga, Spain, September 25–27, 2019, Proceedings*. Springer.
- Corpas Pastor, G., Bautista Zambrana, M.R. et Hidalgo-Ternero, C.M. (éds.). (2021). *Sistemas fraseológicos en contraste: Enfoques computacionales y de corpus*. Comares.
- Cortès C. (éd.). (2003). *La métaphore. Du discours général aux discours spécialisés*. Université Paris 7 Denis Diderot.
- Coseriu, E. (1964/1986). *Principios de semántica estructural*. Gredos.
- Coseriu, E. (1966). Structure lexicale et enseignement du vocabulaire. *Actes du premier colloque international de linguistique appliquée* (p. 175-217). AILA.
- Cotta Ramusino, P. et Mollica, F. (2020). Contrastive phraseology : preliminary remarks. Dans P. Cotta Ramusino et F. Mollica (éds.), *Phraseology: languages and cultures in comparison* (p. 1-10). Cambridge Scholars Publishing.
- Cotta Ramusino, P., Mollica, F. et Schafroth, E.(dirs.) (2014). *FRAME: Fraseologia multilingue elettronica*. Università degli Studi. <http://www.fraseologia.it> (06/09/2022)

- Coulmas, F. (1981). *Conversational routine: explorations in standardized communication situations and prepatterned speech* (vol. 96). Gruyter.
- Coulmas, F. (1985). Lexikalisierung von syntagmen. Dans Ch. Schwarze et D. Wunderlich (éds.), *Handbuch der Lexikologie* (p. 250-268). Athenäum.
- Courtois, B. (1990), Un système de dictionnaires électroniques pour les mots simples du français. *Langue française*, 87, 11-22.
- Courtois, B. (2004). Dictionnaires électroniques DELAF anglais et français, Syntax, Lexis and Lexicon-Grammar. Dans C. Leclère, E. Laporte, M. Piot et M. Silberztein (éds.), *Papers in honour of Maurice Gross*, *Linguisticae Investigationes Supplementa 24* (p. 113–133). John Benjamins Publishing Company.
- Courtois, B. et Silberztein, M. (1990). Dictionnaires électroniques du français. *Langue française*, 87, 3-4.
- Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2(3), 223-235.
- Cowie, A. P. (1985). Collocational dictionaries - A comparative view [conférence présentée au *Fourth Anglo-Soviet English Studies Seminar*, Manchester].
- Cowie, A. P. (1989). *Multiword lexical units and communicative language teaching*. Communication présentée au International Colloquium on 'Vocabulary and Applied Linguistics', Lyon, France.
- Cowie, A. P. (1991). Multiword units in newspaper language. *Cahiers de l'institut de Linguistique de Louvain*, 17(1-3), 101-106.
- Cowie, A. P. (1994). Phraseology. Dans R. E. Asher (éd.), *The encyclopedia of language and linguistics* (p. 3168-3171). Oxford University Press.
- Cowie, A. P. (1998a). Introduction. Dans A. P. Cowie (éd.), *Phraseology: theory, analysis and applications* (p. 1-20). Oxford University Press.
- Cowie, A. P. (1998b). Phraseological dictionaries: some East-West comparisons. Dans A. P. Cowie (éd.), *Phraseology: Theory, Analysis and Applications* (p. 209-228). Oxford University Press.

- Cowie, A. P. (1998c). *Phraseology: theory, analysis and applications*. Oxford University Press.
- Cowie, A. P. (2001). Speech formulae in English: problems of analysis and dictionary treatment. *GAGL: Groninger Arbeiten Zur Germanistischen Linguistik*, 44, 1-12
- Cowie, A. P., Mackin, R. et McCaig, I. R. (1983). *Oxford dictionary of current idiomatic English*. Oxford University Press.
- Croft, W. et Cruse D. A. (2004). *Cognitive linguistics*. Cambridge University Press.
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge University Press.
- Cruse, D. A. (1995). Polysemie and related phenomena from a Cognitive Linguistic viewpoint. Dans P. Saint-Dizier et E. Viegas (dir.), *Computational Lexical Semantics* (p. 33-49). Cambridge University Press.
- Culioli, A. (1990). *Pour une linguistique de l'énonciation*. Ophrys.
- Dagneaux, E., Denness, S. et Granger, S. (1998). Computer-aided error analysis. *System*, 26, 163-174.
- Daille, B. (1994). Combined approach for terminology extraction: lexical statistics and linguistic filtering. Dans A. Wilson et T. McEnery, *UCREL Technical Paper*, 5. <https://ucrel.lancs.ac.uk/papers/techpaper/vol5.pdf>
- Daladier, A. (1978). Problèmes d'analyse d'un type de nominalisation en français et de certains groupes nominaux complexes [thèse de 3^e cycle, Université Paris 7].
- Daneš, F. (1966). The relation of centre and periphery as a language universal. *Travaux Linguistiques de Prague*, 2, 9–21.
- Danlos, L. (1994). Coder des informations monolingues sur les noms pour éviter des règles bilingües sensibles au contexte. *Langages*, 116, 95-110.
- Danlos, L., Nakamura T. et Pradet, Q. (2014). Vers la création d'un Verbnets du français. Dans *Actes de la 21^e conférence sur le Traitement Automatique des Langues Naturelles (TALN'2014)*, Atelier FondamenTAL. <http://www.aclweb.org/anthology/W14-6407>
- Danlos, L., Pradet, Q., Barque, L., Nakamura T. et Constant, M. (2016). Un Verbnets du français. *Traitement automatique des langues*, 57(1), 33-58.

- de Beaugrande, R. (1987). Special purpose language and linguistic theory. *ALSED-LSP Newsletter*, 10(2), 2-10.
- de Boer, C. (1922). *Essais de syntaxe du français moderne*. P. Noordoff.
- de Hoyos Puente, J. C. (2002). *El vocabulario de la bolsa: tratamiento informático* [thèse de doctorat, Universidad de Salamanca].
- de Miguel, E. (2008). Construcciones con verbos de apoyo en español. De cómo entran los nombres en la órbita de los verbos. Dans I. Olza, M. Casado Velarde et R. González Ruix (éds.), *Actas del XXXVII Simposio de la SEL* (p 567-578). Servicio de Publicaciones de la Universidad de Navarra.
- Delaunay, B. et Laurent, N. (2012). *Bescherelle. La conjugaison pour tous*. Hatier.
- Dendien J. et Pierrel, J.-M. (2003). Le Trésor de la Langue Française informatisé. Un exemple d'informatisation d'un dictionnaire de langue de référence. *Traitement Automatique des langues*, 44(2), 11-37.
- Dendien, J. (1996). Le projet d'informatisation du TLF. Dans D. Piotrowski, *Lexicographie et Informatique : autour de l'informatisation du Trésor de la Langue Française* (p. 25-34).
- Depecker, L. (2005). Contribution de la terminologie à la linguistique. *Langages*, 157, 6-13.
- Dictionnaire du français contemporain* (1975). Larousse.
- Dictionnaire Larousse en ligne*. <http://www.larousse.fr/dictionnaires/francais>
- Dobrovol'skij D. O. et Piirainen, E. (2005). *Figurative language. Cross-cultural and cross-linguistic perspectives*. Elsevier.
- Dobrovol'skij D. O. et Piirainen, E. (2018). Conventional figurative language theory and idiom motivation. *Yearbook of phraseology*, 9(1), 5-30.
- Dobrovol'skij, D. O. (1988). *Phraseologie als Objekt der Universalienlinguistik*. VEB Verlag Enzyklopädie.
- Dobrovol'skij, D. O. et Filipenko, T. (2007). Russian phraseology. Dans H. Burger, D. Dobrovol'skij, P. Kühn et N. R. Norrick (éds.), *Phraseologie / Phraseology. Ein*

internationales handbuch zeitgenössischer forschung / An international handbook of contemporary research (p. 714-727). Walter de Gruyter.

- Dostie, G. (2019). Paramètres pour définir et classer les phrases préfabriquées : *La vengeance est un plat qui se mange froid. Bon appétit ! Cahiers de lexicologie*, 114, 27-61.
- Dostie, G. (2021). Vers une modélisation lexicographique des propriétés sémantico-pragmatiques des locutions-phrases génériques et situationnelles *La nuit porte conseil* et *Le chat est sortidu sac. Lexique*, 29, 15-38
- Dostie, G. et Tutin, A. (2019). La phrase préfabriquée dans le paysage phraséologique. Introduction. *Cahiers de lexicologie*, 114, 11-25.
- Dostie, G. et Tutin, A. (dirs.). (2023). *La phraséologie dans les interactions orales et écrites*. John Benjamins Publishing Company.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547-619.
- Drouin, P. (2007). Identification automatique du lexique scientifique transdisciplinaire. *Revue française de linguistique appliquée*, 12(2), 45-64.
- Dubois, J. (1964). Distribution, ensemble et marque dans le lexique. *Cahiers de lexicologie*, 4, 5-16.
- Dubois, J. et Dubois-Charlier, F. (1990). Incomparabilité des dictionnaires. *Langue française*, 87, 5-10.
- Dubois, J. et Dubois-Charlier, F. (1997). *Les verbes français*. Larousse-Bordas.
- Dubois, J., Giacomo, M. Guespin, L., Marcellesi, C., Marcellesi, J.-B. et Mével, J. P. (2012). *Dictionnaire de linguistique et des sciences du langage*. Larousse.
- Dugas, A. et Manseau, H. (1996). *Les verbes logiques*. Les Éditions Logiques.
- Duneton, C. et Claval, S. (1990). *Le bouquet des expressions imagées*. Seuil.
- Dunning, T. (1993), Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61-74.
- Durrant, P. et Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL* 47, 157-177.

- Edmonds, A. (2013a), Parfaitement sensé – La préférence sémantique et l’enseignement des collocations en L2. Dans C. García-Debanco Claudine, C. Masseron et Ch. Ronveaux (éds.), *Enseigner le lexique* (p. 131-151). Presses Universitaires de Namur.
- Edmonds, A. (2013b). Une approche psycholinguistique des phénomènes phraséologiques : le cas des expressions conventionnelles. *Langages*, 189, 121-138.
- Edmonds, A. (2016). Le rôle du contexte dans le traitement des expressions conventionnelles. *Cahiers de lexicologie*, 108, 107-123.
- Ellis, N. C., Simpson-Vlach, R. et Maynard, C. (2008). Formulaic language in native and second-language speakers: psycholinguistics, corpus linguistics, and TESOL. *Tesol Quarterly*, 42, 375-396.
- Emorine, M. (1992). *Formalisation syntaxique des constructions à verbe support en français et en espagnol* [thèse de doctorat, Université de Clermont-Ferrand 2].
- Escandell, V. (1995). *Los complementos del nombre*. Arco Libros.
- Ettinger, S. (1992). Techniques d’apprentissage des expressions idiomatiques. Dans G. Dorion, F.-J. Meissner, J. Riesz et U. Wielandt (éds.), *Le français aujourd’hui - une langue à comprendre - französisch heute. Mélanges offerts à Jürgen Olbert* (p. 98-109). Diesterweg.
- Ettinger, S. (2008). Alcances e límites da fraseodidáctica. Dez preguntas clave sobre o estado actual da investigación. *Cadernos de fraseoloxía galega*, 10, 95-127.
- Everaert, M., van der Linden, E.-J., Schenk, A. et Schreuder, R. (1995). *Idioms: structural and psychological perspectives*. Lawrence Erlbaum Associates Inc. Publishers.
- Evert, S. (2004). *The statistics of word cooccurrences: word pairs and collocations* [thèse de doctorat, Universität Stuttgart]. <https://www.stephanie-evert.de/PUB/Evert2004phd.pdf>
- Evert, S. (2008). Corpora and collocations. Dans A. Lüdeling et M. Kytö (éds.), *Corpus linguistics. An international handbook* (p. 1212–1248). Mouton de Gruyter.

- Falkert, A. (2013). À propos de l'existence d'une compétence phraséologique chez l'apprenant de FLE. Une étude expérimentale. Dans C. Garcia-Debanc, C. Masseron et C. Ronveaux (dirs.), *Enseigner le lexique* (p. 153-172). Presses universitaires de Namur.
- Fauconnier, J.-P. (2015). *French Word Embeddings*. <http://fauconnier.github.io>
- Fauconnier, J.-P. (2022). *Word Embeddings for French*. https://sparknlp.org/2022/02/01/word2vec_wac_200_fr.html
- Felber, H. (1984). *Terminology manual*. Unesco : International Information Centre for Terminology / Infoterm.
- Fellbaum, C. (1993). The determiner in English idioms. Dans C. Cacciari et P. Tabossi (éds.), *Idioms: processing, structure, and interpretation* (p. 271-295). Lawrence Erlbaum Associates.
- Fellbaum, C. (1998). *WordNet: an electronic lexical database*. MIT Press.
- Fellbaum, C. (2006). WordNet(s). Dans K. Brown, *Encyclopedia of Language and Linguistics* (2^e éd., p. 665-670). Elsevier.
- Fernández-Montraveta, A. M. et Vázquez, G. (2010). La construcción del WordNet 3.0 en español. Dans M. A. Castillo Carballo et J. M. García Platero, *La lexicografía en su dimensión teórica* (p. 201-220). Servicio de Publicaciones de la Universidad de Málaga.
- Fernández-Montraveta, A. M., Vázquez, G. et Fellbaum C. (2008). The Spanish version of WordNet 3.0. Dans A. Storrer, A. Geyken, A. Slebert et K.M. Würzner (éds.), *Text resources and lexical knowledge* (p. 175-182). Mouton de Gruyter.
- Fernando, C. et Flavell, R. (1981). *On idiom: Critical views and perspectives*. University of Exeter.
- Ferraresi, A., Bernardini, S., Picci, G., et Baroni, M. (2010). Web corpora for bilingual lexicography: a pilot study of English/French collocation extraction and translation. Dans R. Xiao, *Using corpora in contrastive and translation studies* (p. 337-362). Cambridge Scholars Publishing.

- Fiala, P., Habert, B., Lafon, P. et Pineira, C. (1978). Des mots aux syntagmes. Figements et variation dans la Résolution du congrès de la CGT de 1978. *Mots*, 14, 47-88.
- Fiala, P., Lafon, P. et Piguet, M.-F. (éds.). (1997). *La locution : entre lexique, syntaxe et pragmatique. Identification en corpus, traitement, apprentissage*. Klincksieck.
- Fillmore, C. (1968). *The case for case*. Dans E. Bach et R. Harns, (éds.), *Universals in linguistic theory*. Holt, Rinehard & Winston.
- Fillmore, C. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1), 20-32.
- Fillmore, C. (1982). Frame semantics. Dans *Linguistics in the morning calm* (p. 111-137). Hanshin Publishing Co.
- Fillmore, C. (1988). The mechanisms of construction grammar. Dans S. Axmaker, A. Jaisser et H. Singmaster (éds.), *Proceedings of the fourteenth annual meeting of the Berkeley Linguistics Society* (p. 35-55). Berkeley Linguistics Society.
- Fillmore, C., Johnson, C. et Petruk, M. (2003). Background to FrameNet. *International Journal of Lexicography*, 16, 235-250.
- Fillmore, C., Kay, P. et O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language* 64(3), 501-538.
- Fillmore, C., Ruppenhofer, J. et Baker, C. F. (2004). Framenet and representing the link between semantic and syntactic relations. *Frontiers in linguistics*, 1, 19-59.
- Firth, J. R. (1951). *Models of meaning*. Bobbs-Merrill.
- Firth, J. R. (1957a). A synopsis of linguistic theory. 1930–1955. Dans *Studies in linguistic analysis. (Special Volume of the Philological Society)* (p. 1-32). Blackwell. Réédité dans F. R. Palmer (éd.) (1968), *Selected papers of J. R. Firth. 1952-1959*. Longman.
- Firth, J. R. (1957b). *Papers in linguistics. 1934-1951*. Oxford University Press.
- Fišer, D. et Sagot, B. (2008). Combining multiple resources to build reliable WordNets. Dans P. Sojka, A. Horák, I. Kopeček et K. Pala (éds.), *Text, Speech and Dialogue. 11th International Conference, TSD 2008. Czech Republic, September 2008. Proceedings* (p. 61-68). Springer.

- Flaux, N. et Van de Velde, D. (2000). *Les noms en français : esquisse de classement*. Ophrys.
- Flaux, N., Glatigny, M. et Samain, D. (éds.). (1996). *Les noms abstraits. Histoire et théories*. Presses Universitaires du Septentrion.
- Fleischer, W. (1982/1997). *Phraseologie der deutschen gegenwartssprache*. Max Niemeyer Verlag.
- Fleischer, W. (1992). Konvergenz und divergenz von wortbildung und phraseologisierung. Dans J. Korhonen (éd.), *Phraseologie und wortbildungsaspekte der lexikonerweiterung* (p. 53-65). Max Niemeyer Verlag.
- Fletcher, W. H. (2013). Corpus analysis of the world wide web. Dans C. A. Chapelle, *The encyclopedia of applied linguistics* (p. 1339-1347). Wiley-Blackwell.
- Fónagy, I. (1982). *Situation et signification*. John Benjamins Publishing Company.
- Fónagy, I. (1997). Figement et changement sémantiques. Dans M. Martins-Baltar (coord.), *La locution, entre langue et usages*. ENS Éditions.
- Francis, G. (1991). Nominal group heads and clause structure. *Word*, 42(2), 144-156.
- Francis, G., Hunston, S. et Manning, E. (1996). *Collins Cobuild Grammar Patterns 1: verbs*. HarperCollins.
- François, J. et Mejri, S. (éds) (2006). *Composition syntaxique et figement lexical*. Presses Universitaires de Caen.
- François, J. et Sénéchal, M. (2006). Le sémantisme propre des cadres prédicatifs et la polysémie des verbes de production de parole. *Cahiers du CRISCO*, 21, 49-69.
- François, J., Le Pesant, D. et Leeman, D. (2007). Présentation de la classification des verbes français de Jean Dubois et Françoise Dubois-Charlier. *Langue française*, 153(1), 3-19.
- Fraser, B. (1970). Idioms within a transformational grammar. *Foundations of language*, 6, 22-42.
- Frege, G. (1867/1979). *Begriffsschrift*, a formula language, modeled upon that of arithmetic, for pure thought. Dans J. van Heijenoort (éd.), *From Frege to Gödel*:

- a source book in mathematical logic, 1879-1931* (p. 1-82). Harvard University Press.
- Frege, G. (1884/1970). *Les fondements de l'arithmétique : recherche logico-mathématique sur le concept de nombre* (traduit par C. Imbert). Éditions du Seuil.
- Frege, G. (1971/1984). Introducción a la lógica (traduit par L. M. Poveda). *Teorema: Revista internacional de filosofía*, 14(3-4), 497-510.
- Frenck-Mestre, Ch. et Prince P. (1999). La compréhension écrite d'une deuxième langue : lexicque et syntaxe. *Les langues modernes*, 3, 58-59.
- Fuchs, C. (1988). Représentation linguistique de la polysémie grammaticale. *T.A. Informations*, 29(1-2), 7-20.
- Fuchs, C. (1996). *Les ambiguïtés du français*. Ophrys.
- Fuchs, C. et Habert. B. (2004). Le traitement automatique des langues : des modèles aux ressources. *Le français moderne*, 72 (1), 1-13.
- Futagi, Y., Deane, P., Chodorow, M. et Tetreault, J. (2008). A computational approach to detecting collocation errors in the writing of non-native speakers of English. *Computer assisted language learning*, 21(4), 353–367.
- Gaatone, D. (1984). La locution ou les poids de la diachronie dans la synchronie. *Le Moyen Français*, 14-15, 70-81.
- Gaatone, D. (1997). La locution : Analyse interne et analyse globale. Dans M. Martins-Baltar (éd.), *La locution entre langage et usages* (p. 165-177). ENS Éditions.
- Gader, N., Lux-Pogodalla, V. et Polguère, A. (2012). Hand-crafting a lexical network with a knowledge-based graph editor. Dans M. Zock et R. Rapp (éds.), *Proceedings of the Third workshop on cognitive aspects of the lexicon (CogALex III)* (p. 109–125). The COLING 2012 Organizing Committee. <https://hal.science/hal-00768798v1/document>
- Gader, N., Ollinger, S. et Polguère, A. (2014). One lexicon, two structures: so what gives? Dans H. Orav, C. Fellbaum et P. Vossen (dirs.), *Proceedings of the Seventh Global Wordnet Conference (GWC2014)* (p. 163–171). <https://aclanthology.org/W14-0122.pdf>

- Gala, N. (2013). Ressources lexicales mono- et multilingues : une évolution historique au fil des pratiques et des usages. Dans N. Gala et M. Zock (dirs.), Ressources lexicales : contenu, évaluation, utilisation, évaluation. *Linguisticae Investigationes Supplementa* (p. 1–42). John Benjamins Publishing Company.
- Galisson, R. (1984a). *Dictionnaire de compréhension et de production des expressions imagées*. CLE International.
- Galisson, R. (1984b). *Les expressions imagées*. Clé International.
- García-Page Sánchez, M. (1990). Sobre implicaciones lingüísticas. Solidaridad léxica y expresión fija. *Estudios humanísticos. Filología*, 12, 215-227.
- García-Page Sánchez, M. (1998a). Expresión fija y sinonimia. Dans G. Wotjak (éd.), *Estudios de fraseología y fraseografía del español actual* (p. 83-95). Vervuert/Iberoamericana.
- García-Page Sánchez, M. (2004). De los fines y confines de la fraseología. Dans M. González Calvo, J. Terrón González et J. C. Martín Camacho (éds.), *VII Jornadas de Metodología y didáctica de la lengua española: las unidades fraseológicas*. Universidad de Extremadura (p. 23-79). Universidad de Extremadura.
- García-Page Sánchez, M. (2008). *Introducción a la fraseología española. Estudio de las locuciones*. Antropos.
- Gardent, C. et Cerisaraï, C. (2010). Semi-automatic Propbanking for French. Dans M. Dichinson, K. Müürisepp et M. Passarotti, *Proceedings of the ninth international workshop on Treebanks and Linguistic Theories* (p. 67–78). Northern European Association for Language Technology (NEALT).
- Gates, E. (1988). The treatment of multiword lexemes in some current dictionaries of English. Dans M. Snell-Hornby (éd.), *ZüriLEX'86 Proceedings. Papers read at the EURALEX International Congress, University of Zürich, 9-14 September 1986* (p. 99-106). Francke.
- Gaudin, F. (1993). *Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles*. Presses de l'Université de Rouen.

- Geeraerts, D. (2003). The interaction of metaphor and metonymy in composite expressions. Dans R. Driven et R. Pörlings (éds.), *Metaphor and Metonymy in Comparison and Contrast* (p. 435-465). Walter de Gruyter.
- Gibbs, R. W. (1995). Idiomaticity and human cognition. Dans M. Everaert, D.-J. van der Linden, A. Schenk et R. Schreuder (éds.), *Idioms: structural and psychological perspectives* (p. 97-116). Psychology Press.
- Gibbs, R. W. et Nayak, N. P. (1989). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology*, 21, 100-138.
- Gili Gaya, S. (1964). El lenguaje de la ciencia y de la técnica. *Presente y futuro de la lengua española* (vol. 2, p. 269-76). Ediciones Cultura Hispánica.
- Giry-Schneider, J. (1978b). *Les nominalisations en français. L'opérateur faire dans le lexique*. Librairie Droz.
- Giry-Schneider, J. (1987). *Les prédicats nominaux en français : les phrases simples à verbe support*. Librairie Droz.
- Gläser, R. (1986a). A plea for phraseo-stylistics. *Trends in linguistics*, 32, 41-52.
- Gläser, R. (1986b). *Phraseologie der englischen sprache*. Max Niemeyer Verlag.
- Gläser, R. (1988). The grading of idiomaticity as a presupposition for a taxonomy of idioms. Dans W. Hüllen et R. Schulze (éds.), *Understanding the Lexicon* (p. 264-279). Max Niemeyer Verlag.
- Godard, D. et Jayez, J. (1996). Types nominaux et anaphores : le cas des objets et des événements. Dans W. De Mulder, L. Tasmowski-De Ryck D. Ryck et C. Vettters (éds.), *Anaphores temporelles et (in-)cohérence* (p. 41-58). Brill.
- Godart-Wendling, B, Idefonse, F., Pariente, J.-C. et Rosier-Catach, I. (1998) Penser le principe de compositionnalité : éléments de réflexion historiques et épistémologiques. *T.A.L.* 39(1), 9-34.
- Goldberg, A. (1995). *Constructions – A Construction Grammar approach to argument structure*. Chicago University Press.
- Goldberg, A. (1998). Patterns of experience in patterns of language. Dans M. Tomasello (éd), *The new psychology of language – Cognitive and functional approaches to language structure* (p. 203-19). LEA.

- Gómez Asencio, J. J. (1981). *Gramática y categorías verbales en la tradición española (1771-1847)*. Ediciones Universidad de Salamanca.
- Gómez de Enterría, J. (2006). Últimos enfoques en la enseñanza-aprendizaje del español con fines profesionales. Dans A. M. Cestero Mancera (éd.), *Lingüística aplicada a la enseñanza del español como lengua extranjera: desarrollos recientes* (p. 47-60). Universidad de Alcalá.
- Gómez Fernández, A. (2011a). Enseñar a inferir: las colocaciones de una lengua de especialidad mediante la intercomprensión. *Intercomprehension. Learning, teaching, research* (p. 65-83). Narr Francke Attempto Verlag.
- Gómez Fernández, A. (2011b). Inferencias léxicas y colocaciones en la intercomprensión plurilingüe. Dans M. Sanz et J. Verdegal (éds.), *Construcción de identidades y cultura del debate en los estudios en lengua francesa* (p. 204-216). Publicacions de la Universitat Jaume I.
- Gómez Fernández, A. (2020). Le phrasème comme garant de progression textuelle. Dans L. Fesenmeier et I. Novakova (éds.), *Phraséologie et stylistique de la langue littéraire. Approches interdisciplinaires* (p. 199-209). Peter Lang.
- Gómez Fernández, A. (2021a). *J'ai de sérieuses raisons de croire...* à la phraséodidactique plurilingue. *PHRISIS - Rivista di studi fraseologici e paremiologici*, 5, 61-75.
- Gómez Fernández, A. (2021b). Noms composés et autres phrasèmes : fonctionnement discursif dans le domaine spécialisé du sport. *Roczniki Humanistyczne*, 69(8), 39-55.
- Gómez Fernández, A. et Uzcanga Vivar, I. (2004). Expresiones de la intensidad adjetival en francés y español. Dans J. Suso López et R. López Carrillo (coord.), *Le français face aux défis actuels. Histoire, langue et culture* (p. 55-64). Universidad de Granada, APFUE - GILEC.
- Gómez Fernández, A. et Uzcanga Vivar, I. (2010a). Eurom.com.text : étiquetage sémantique. *Synergies Espagne*, 3, 171-186.
- Gómez Fernández, A. et Uzcanga Vivar, I. (2010b). Pour une nouvelle approche de la lexicographie bilingue français → espagnol. Dans F. Neveu, V. Muni Toke, J.

Durand, T. Klingler, L. Mondada et S. Prévost (éds.), *Congrès mondial de linguistique française 2010* (p. 981-989).
https://www.linguistiquefrancaise.org/articles/cmlf/abs/2010/01/cmlf2010_000244/cmlf2010_000244.html

Gómez Fernández, A. et Uzcanga Vivar, I. (2008). De l'exploitation des genres textuels et types discursifs. *Les Langues Modernes*, 1/2008, 1-16.

Gómez González-Jover, A. (2005). *Terminografía, lenguajes profesionales y mediación interlingüística: aplicación metodológica al léxico especializado de la industria del calzado y las industrias afines* [thèse de doctorat, Universidad de Alicante].
<http://rua.ua.es/dspace/handle/10045/760>

González Hernández, A. T. (2016). Criterios para la comprensión de la competencia fraseológica en FLE. Dans T. Gonzalo Santos, M. V. Rodríguez Navarro, A. T. González Hernández y J. M. Pérez Velasco (éds.), *Texto, género y discurso en el ámbito francófono* (p. 155-168). Ediciones Universidad de Salamanca.

González Rey, M. I. (2002). *La phraséologie du français*. Presses Universitaires du Mirail.

González Rey, M. I. (2006). A Fraseodidáctica eo Marco europeo común de referencia paras as Linguas. *Cadernos de fraseoloxía galega*, 8, 123-145.

González Rey, M. I. (2007). *La didactique du français idiomatique*. InterCommunications/E.M.E.

González Rey, M. I. (2012). De la didáctica de la fraseología a la fraseodidáctica. *Paremia*, 21, 67-84.

González Rey, M. I. (2016). *PHRASÉOTEXT - Le français idiomatique. Méthode ded phraséodidactique (français - français)*. Servizo de Publicacións e Intercambio Científico da Universidade de Santiago de Compostela.

González Rey, M. I. (2019). Proceso de elaboración de un método en fraseodidáctica del FLE: de la concepción a la experimentación. Dans C. Crida Álvarez et A. Alessandro (éds.), *Innovación en fraseodidáctica. Tendencias, enfoques y perspectivas* (p. 25-43). Peter Lang.

- González Rey, M. I. (2021). *La nouvelle phraséologie du français*. Presses Universitaires du Midi.
- González Rey, M. I. (éd.). (2014). *Outils et méthodes d'apprentissage en phraséodidactique*. EME & InterCommunications.
- González Royo, C. et Mogorrón Huerta, P. (éds.). (2010). *Fraseología contrastiva: lexicografía, traducción y análisis de corpus*. Universidad de Alicante.
- Gottlob, F. (1984/1970). *Estudios sobre semántica*. Orbis.
- Grabarczyk, Z. (1988). Statistics and LSP didactics. Dans H. Czap et C. Galanski (éds), *Terminology and Knowledge Engineering* (p. 204-206). INDEKS Verlag.
- Grammaire de l'Académie française* (1932). Firmin-Didot.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. Dans A. P. Cowie (éd.), *Phraseology. Theory, analysis and applications* (p. 145-160). Clarendon Press.
- Granger, S. (2003). The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly*, 37(3), 538-546.
- Granger, S. et Meunier, F. (éds.). (2008). *Phraseology. An interdisciplinary perspective*. John Benjamins Publishing Company.
- Granger, S. et Paquot, M. (2008). Disentangling the phraseological web. Dans S. Granger et F. Meunier (éds.), *Phraseology: an interdisciplinary perspective* (p. 27-49). John Benjamins Publishing Company.
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A. et Mikolov, T. (2018). Learning word vectors for 157 languages. Dans N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis et T. Tokunaga (éds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (p. 3483-3487). European Language Resources Association (ELRA). <https://aclanthology.org/volumes/L18-1/>
- Gray, B. et Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, 18(1), 109-136,

- Gréciano, G. (1983). *Signification et dénotation en allemand. La sémantique des expressions idiomatiques*. Recherches Linguistiques n° IX, Université de Metz-Klinsieck.
- Gréciano, G. (1984). Pour un apprentissage des unités phraséologiques. *Nouveaux cahiers d'allemand*, 2, 95-113.
- Gréciano, G. (1989b). Von der struktur zur kultur. Entwicklungstendenzen im deutsch-französischen phraseologievergleich. *Zeitschrift für Germanistik*, 10(5), 517–527.
- Gréciano, G. (1991). Valence, version intégrée. *L'Information Grammaticale*, 50, 13-18.
- Gréciano, G. (1995). Fachphraseologie. Dans R. Métrich et E. Faucher (éds.), *Rand und band – Abgrenzung und verknüpfung als grundtendenzen des Deutschen* (p. 183-195). Gunter Narr Verlag.
- Gréciano, G. (1997). Collocations rythmologiques. *Meta : Journal des traducteurs/Meta : Translators' Journal*, 42(1), 33-44.
- Gréciano, G. (1999). Préfabriqués linguistiques, leur performance cognitive et communicative (traduit de l'allemand par l'auteur). http://www.inst.at/studies/s_0103_f.htm
- Gréciano, G. (2000a). Qui se ressemble s'assemble : locutions, particules et compères. Dans G. Gréciano (éd.), *Micro- et macrolexèmes et leur figement discursif. Études de linguistique comparée français-allemand* (p. 363-373). Peeters.
- Gréciano, G. (2000c). Phraséologie, ses co(n)textes et ses contrastes. *Paremia*, 9, 91-102.
- Gréciano, G. (2018). Le soleil se lève à l'Est / Die Sonne geht im Osten auf / Ex oriente lux. Évolutions de la recherche phraséologique. Dans O. Soutet, I. Sfar et S. Mejri (dirs.), *La phraséologie contrastive* (p. 15-23). Honoré Champion.
- Gréciano, G. (éd.) (1989a). *Europhras 88 : phraséologie contrastive. Actes du colloque international Klingenthal-Strasbourg, 12-16 mai 1988*. Université des Sciences Humaines.
- Gréciano, G. (éd.). (2000b). *Micro- et macrolexèmes et leur figement discursif. Études de linguistique comparée français-allemand*. Peeters.

- Greimas, A. J. (1960). Idiotismes, proverbes et dictons. *Cahiers de lexicologie*, 2, 41-61.
- Greimas, A. J. (1970). Les proverbes et les dictons. *Du sens. Essais sémiotiques*. Éditions du Seuil.
- Grevisse, M. (1980). *Le bon usage* (11^e éd.). Duculot.
- Grezka, A. (2006). *Les prédicats de perception. Traitement de la polysémie (Les sens des sens)* [thèse de doctorat, Université Sorbonne Paris Nord].
- Grezka, A. et Buvet, P-A. (2007). Élaboration d'outils méthodologiques pour décrire les prédicats du français. *Linguisticae Investigationes*, 30(2), 217-245.
- Grezka, A., Cartier, E., et Mathieu-Colas, M. (2015). Dictionnaires morphologiques du français contemporain : présentation de Morfetik, éléments d'un modèle pour le TAL. Dans J.-M. Lecarpentier et N. Lucas (éds.), *Actes de la 22^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2015)* (p. 466-472). ATALA. <https://hal.archives-ouvertes.fr/hal-01443197/>
- Grimshaw, J. (1990). *Argument Structure*. MIT Press.
- Grimshaw, J. et Williams, E. (1993). Nominalization and predicative propositional phrases. Dans J. Pustejovsky (éd.), *Semantics and Lexicon* (p. 97-106). Kluwer Academic Publishers.
- Grobelak, L. (1990). *Dictionnaire collocationnel du français général*. Państwowe Wydawnictwo Naukowe.
- Gross, G. (1982). Un cas de constructions inverses : donner et recevoir. *Linguisticae Investigationes*, 6(2), 1-45.
- Gross, G. (1986). *Typologie des noms composés : le lexique électronique des noms composés du français* [rapport ATP, CNRS, Université Sorbonne Paris Nord].
- Gross, G. (1988). Degré de figement des noms composés. *Langages*, 90, 57-72.
- Gross, G. (1989). *Les constructions converses du français*. Librairie Droz.
- Gross, G. (1992). Forme d'un dictionnaire électronique. Dans A. Class et H. Safar (éds.), *L'environnement traductionnel*. Presses Universitaires du Québec.
- Gross, G. (1993a). Les passifs nominaux. *Langages*, 109, 103-125.

- Gross, G. (1993b). Trois applications de la notion de verbe support. *L'Information Grammaticale*, 59, 16-22.
- Gross, G. (1994a). Classes d'objets et description des verbes. *Langages* 115, 15-30.
- Gross, G. (1994b). Un outil pour le FLE : les classes d'objets. *Actes du colloque du FLE de mai 1992* (p. 169-186). Presses Universitaires de Lille.
- Gross, G. (1995). Une sémantique nouvelle pour la traduction automatique : les classes d'objets. *La tribune des industries de la langue et de l'information électronique*, 17-18-19, 16-19.
- Gross, G. (1996a). *Les expressions figées en français : noms composés et autres locutions*. Ophrys.
- Gross, G. (1996b). Prédicats nominaux et compatibilité aspectuelle. *Langages*, 121, 54-72.
- Gross, G. (1997). Du bon usage de la notion de locution. Dans M. Martins-Baltar (éd.), *La locution entre langage et usages* (p. 201-223). ENS Éditions.
- Gross, G. (1998). Pour une typologie des prédicats nominaux. *Studia Romanica Upsaliensis*, 56, 221-230.
- Gross, G. (1999). Sémantique lexicale et connecteurs. *Langages*, 136, 76-84.
- Gross, G. (2002). Locutions conjonctives et détermination. *Langages*, 36, 9-20.
- Gross, G. (2004a). Pour un Bescherelle des prédicats nominaux. *Linguisticae Investigationes*, 27(2), 343-358.
- Gross, G. (2004b). Réflexions sur le traitement automatique des langues. Dans G. Prunelle, C. Fairon et A. Dister (éds.), *Le poids des mots : actes des 7^{es} journées internationales d'analyse statistique des données textuelles*, Louvain-la-Neuve, 10-12 mars 2004 (vol. 1, p. 545-556). Presses Universitaires de Louvain.
- Gross, G. (2005a). Les circonstancielles sont des complétives. Dans F. Lambert et H. Nølke (éds.), *La syntaxe au cœur de la grammaire, Recueil offert en hommage pour le 60^e anniversaire de Claude Muller* (p. 121-127). Presses Universitaires de Rennes.

- Gross, G. (2005b). Pour un Bescherelle des prédicats nominaux. *Lingvisticae Investigationes*, 27(2), 343-358.
- Gross, G. (2006). Sur le statut des locutions prépositives. *Modèles linguistiques* 27-1(53), 35-50.
- Gross, G. (2008). Les classes d'objets. *Lalies*, 28, 111-165.
- Gross, G. (2009). *Les verbes supports entre temps et aspect* [conférence]. Colloque international : Supports et prédicats non verbaux dans les langues du monde, Paris, France.
- Gross, G. (2010a). La notion d'*emploi* dans une grammaire de prédicats. *Cahiers de lexicologie* 96, 97-115.
- Gross, G. (2010b). Sur la notion de contexte. *Meta : Journal des traducteurs/Meta : Translators' Journal*, 55(1), 187-197.
- Gross, G. (2012). *Manuel d'analyse linguistique : approche sémantico-syntaxique du lexique*. Presses Universitaires du Septentrion.
- Gross, G. (2013). Unité d'analyse linguistique. *Cahiers de lexicologie*, 102, 27-38.
- Gross, G. (2014). Traitement automatique de la polysémie. Dans Y. Morimoto (éd.), *Léxico, didáctica y nuevas tecnologías* (p. 39-56). Universidade da Coruña.
- Gross, G. (2015). Traitement automatique de la polysémie. *Studia Romanica Posnaniensia*, 42 (1), 15-33.
- Gross, G. (2017). Petit historique de la notion de verbes supports. *Cahiers de lexicologie et lexicographie*, 111, 121-144.
- Gross, G. et Clas, A. (1997). Synonymie, polysémie et classes d'objets. *Meta : Journal des traducteurs/Meta : Translators' Journal*, 42(1), 147-154.
- Gross, G. et Mathieu-Colas, M. (2001). Description de la langue de la médecine. *Meta*, 46(1), 68-81.
- Gross, G. et Vivès, R. (1986). Les constructions nominales et l'élaboration d'un lexique-grammaire. *Langue française*, 69, 5-27.
- Gross, M. (1971). Grammaire transformationnelle et enseignement du français. *Langue française*, 11, 4-14.

- Gross, M. (1975). *Méthodes en syntaxe*. Hermann.
- Gross, M. (1977). *Grammaire transformationnelle du français. Syntaxe du nom*. Larousse.
- Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages* 63, 7-52.
- Gross, M. (1982). Une classification des phrases "figées" du français. *Revue québécoise de linguistique*, 11(2), 151-185.
- Gross, M. (1988). Les limites de la phrase figée. *Langages*, 90, 7-22.
- Gross, M. (1993). Les phrases figées en français. *L'information grammaticale*, 59, 36-41.
- Gross, M. (1995). Une grammaire locale de l'expression des sentiments. *Langue française*, 105, 70-87.
- Gross, M. (1998). La fonction sémantique des verbes supports. *Revue Internationale de Linguistique Française*, 37, 23-35.
- Grossmann, F. et Tutin, A. (éds.) (2003). *Les collocations : analyse et traitement*. Werelt.
- Grunig, B.-N. (1997). La locution comme défi aux théories linguistiques : une solution d'ordre mémoriel ? Dans M. Martins-Baltar (éd.), *La locution entre langue et usages* (p. 225-240). END
- Guillet, G. (1993). Le lexique des verbes : description et organisation. *L'Information Grammaticale*, 59, 23-35.
- Guimier, C. et Oueslati, L. (2006). Le degré de figement des constructions < verbe + adjectif invarié >. Dans J. François et S. Mejri (éds.), *Composition syntaxique et Figement lexical* (p. 17-39). Presses Universitaires de Caen.
- Guiraud, P. (1961). *Les locutions françaises*. PUF.
- Gülich, E. et Krafft, U. (1997). Le rôle du préfabriqué dans les processus de production discursive. Dans M. Martins-Baltar (éd.), *La locution entre langue et usages* (p. 241-276). ENS Éditions.
- Gutiérrez Rodilla, B. M. (2005). *El lenguaje de las ciencias*. Gredos.

- Haas, P. et Huyghe, R. (2010). Les propriétés aspectuelles des noms d'activités. Dans E. Moline et C. Vettters (éds.), *Temps, aspect et modalité en français* (p. 103-118). Brill.
- Haas, P., Huyghe, R. et Marín, R. (2008). Du verbe au nom : calques et décalages aspectuels. Dans J. Durand, B. Habert et B. Laks (éds.), *Congrès Mondial de Linguistique Française - CMLF'08* (p. 2039–2053). <https://u-paris.hal.science/hal-00975632/document>
- Habert, B., Nazarenko, A. et Salem, A. (1997). *Les linguistiques de corpus*. Armand Colin.
- Haensch, G. (1987). Les llengües d'especialitat o 'tecnolecte'. *Revista de Llengua i Dret*, 10, 7-31.
- Haensch, G., Wolf, L., Ettinger, S et Werner R. (1982). *La lexicografía: de la lingüística teórica a la lexicografía práctica*. Gredos.
- Halliday, M. A. K. (1961). Categories of the Theory of Grammar. *Word*, 17, 241-292.
- Halliday, M. A. K. (1966). Lexis is a linguistic level. Dans C. E. Bazell et al. (dir.), *In memory of John Firth* (p. 148-162). Longmans.
- Halliday, M. A. K. et Hasan, R. (1976). *Cohesion in English*. Longman.
- Harris, Z. S. (1964/1970). Elementary transformations. Dans *Papers in Structural and Transformational Linguistics* (p. 428-532). Springer
- Harris, Z. S. (1971). *Structures mathématiques du langage*. Dunod.
- Harris, Z. S. (1976). *Notes du cours de syntaxe* (traduit par M. Gross). Le Seuil.
- Häusermann, J. (1977). *Phraseologie. Hauptprobleme der deutschen phraseologie auf der basis sowjetischer forschungsergebnisse*. Max Niemeyer Verlag.
- Hausmann, F. J. (1979). Un dictionnaire des collocations est-il possible ? *Travaux de linguistique et littérature*, 17(1), 187-195.
- Hausmann, F. J. (1984). Wortschatzlernen ist kollokationslernen. Zum lehren u. lernen franzoesischer wortverbindungen. *Praxis des neusprachlichen Unterrichts*, 31(4), 395-407.

- Hausmann, F. J. (1985). Kollokationen im Deutschen wörterbuch: ein beitrag zur theorie des lexikographischen beispiels. Dans H. Bergenholtz et J. Mugdan (éds.), *Lexikographie und grammatik* (p. 118-129). Niemeyer.
- Hausmann, F. J. (1989). Le dictionnaire de collocations. Dans J. F. Hausmann, O. Reichmann, H. E. Wiegand et L. Zgusta (éds.), *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. An international Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie* (vol. 1, p. 1010–1019). Walter de Gruyter.
- Hausmann, F. J. (1991). Collocations in monolingual and bilingual English dictionaries. Dans V. Ivir et D. Kalogjera (éds.), *Languages in contact and contrast. Essays in contact linguistics* (p. 225-236). Walter de Gruyter.
- Hausmann, F. J. (1997). Tout est idiomatique dans les langues. Dans M. Martins-Baltar (éd.), *La locution entre langage et usages* (p. 277-290). ENS Éditions.
- Hausmann, F. J. (1998). O dicionario de colocacións. Criterios de organización. Dans X. ferro Ruibal (coord.), *Actas do I Coloquio Galego de Fraseoloxía* (P. 63-82). Centro Ramón Piñeiro, Xunta de Galicia.
- Hausmann, F. J. (2007). Lexicographie française et phraséologie. Dans E. Haag (éd.), *Collocations, phraséologie, lexicographie. Études 1977-2007* (p. 121-153). Shaker Verlag.
- Hausmann, F. J. et Blumenthal, P. (2006). Présentation : collocations, corpus, dictionnaires. *Langue française*, 150, 3-13.
- Heid, U. (1994). On ways words work together – Topics in lexical combinatorics. Dans W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg et P. Vossen. (éds.), *EURALEX'94 Proceedings* (p. 226-257). Vrije Universiteit Amsterdam.
- Heitz, T. (2006). Modélisation du prétraitement des textes. Dans J.-M. Viprey (éd), *Actes JADT'06 : 8^e Journées internationales d'analyse statistique des données textuelles* (vol. 1, p. 499-506). Presses universitaires de Franche-Comté.
- Henrot Sostero, G. et González Rey, M. I. (coords.). (2019). Phraséodidactique: de la conscience à la compétence. *Repères-Dorif*, 18. DoRiF Università. http://www.dorif.it/ezine/show_issue.php?iss_id=30.

- Henry, G. (1992). *Dictionnaire des expressions nées de l'histoire*. Tallandier.
- Hernán Ramírez, L. (1979). *Estructura y funcionamiento del lenguaje*. Edición Ultra.
- Hernández Hernández, H. (1989). *Los diccionarios de orientación escolar. Contribución al estudio de la lexicografía monolingüe española*. Max Niemeyer.
- Herrero Ingelmo, J. L. (2002). Los verbos soportes: el verbo *dar* en español. *Léxico y Gramática*. Repéré à <http://diarium.usal.es/joluin/files/2013/12/darsoportepdf>
- Herrero Ingelmo, J. L. (2004). ¿Puede un sustantivo predicar? (de los sustantivos que se pueden conjugar). *Actas del V Congreso de Lingüística General* (p. 453–468). Arco Libros.
- Heyd, S. et Knittel, M.-L. (2009). Les noms d'activité parmi les noms abstraits : propriétés aspectuelles, distributionnelles et interprétatives. *Linguisticae investigationes*, 32(1), 124-148.
- Hoey, M. (1991). *Patterns of lexis in text*. Oxford University Press.
- Hoey, M. (2005). *Lexical priming: a new theory of words and language*. Routledge.
- Hoffmann, L. (1979). Towards a theory of LSP. Elements of a methodology of LSP analysis. *Fachsprache*, 1(1/2), 12-17.
- Hoffmann, L. (1998). *Llenguatges d'especialitat. Selecció de textos*. Institut Universitari de Lingüística aplicada, Universitat Pompeu Fabra.
- Hornby, A. S. (1948). *Learner's dictionary of current english*. Oxford University Press.
- Hougaard, C. (1986). Danish and Slavic phraseology contributions to an analysis of idiomatics on a contrastive basis, *Papers and Studies in Contrastive Linguistics*, 21, 145-196.
- Howarth, P. (1998a). Phraseology and second language proficiency. *Applied Linguistics*, 19/1, 24-44.
- Howarth, P. (1998b). The phraseology of learners. Dans A. P. Cowie (éd.), *Phraseology. Theory, analysis and applications* (p. 161-186). Oxford University Press.
- Hudson, J. (1998). *Perspectives on fixedness: applied and theoretical*. Lund University Press.

- Hundt, C. (1994a). Construções de verbo + substantivo: estrutura semântica e posição dentro da fraseologia. Dans A. Endruschat, M. Vilela et G. Wotjak (éds.), *Verbo e estruturas fráscas do verbo. Colóquio Internacional de Hispanística* (p. 267-275). Universidade do Porto.
- Hundt, C. (1994b). *Untersuchungen zur portugiesischen Phraseologie*. Gottfried Egert.
- Hunston, S. et Francis, G. (2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. John Benjamins Publishing Company.
- Huyghe, R. (2014). Noms syncatégorématiques et degrés de dépendance syntactico-sémantique. Dans E. Higert et al. (éds), *Res Per Nomen 4. Les théories du sens et de la référence* (p. 155-172). Epure.
- Huyghe, R. (2015). Les typologies nominales : présentation. *Langue française*, 185, 5-27.
- Ibrahim, A. H. (dir.). (1996). *Les supports. Langages*, 121.
- Ingria, R. J. P. et Leland, G. (1993). Adjectives, nominals, and the status of arguments. Dans J. Pustejovsky (éd.), *Semantics and lexicon* (p. 107-127). Kluwer Academic Publishers.
- Iñesta Mena, E. M. et Pamies Beltrán, A. (2002). *Fraseología y metáfora: aspectos tipológicos y cognitivos*. Granada Lingvística y Método Ediciones.
- Iordanskaja, L et Mel'čuk, I. (2017). *Le mot français dans le lexique et dans la phrase*. Hermann Éditeurs.
- Isačenco, A. V. (1948). Morphologie, syntaxe et phraséologie. *Cahiers Ferdinand de Saussure*, 7, 17-32.
- ISO 1087 (2019). *Travail terminologique et science de la terminologie — Vocabulaire*. <https://www.iso.org/obp/ui/#iso:std:iso:1087:ed-2:v1:fr:term:3.1.7>
- ISO 1087 (1990). *Terminologie — Vocabulaire*.
- ISO 2008. *ISO DIS 24617-1: 2008 Language resource management - Semantic annotation Framework - Part 1: time and events*.
- Issac, F. (2011). Figement et informatique. Dans J.-C. Anscombe et S. Mejri (dirs.), *Le figement linguistique : la parole entravée* (p. 419-431). Honoré Champion.

- Issac, F. (2016). Outils et instruments pour la dictionnaire. *Le français moderne, 2016-1*, 96-111.
- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. MIT Press.
- Jackendoff, R. (1993). *Semantic Structures*. MIT Press.
- Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford University Press.
- Jacquet, G., Venant, F. et Victorri B. (2005). Polysémie lexicale. Dans P. Enjalbert (dir.), *Sémantique et traitement automatique du langage naturel* (p. 99-132). Hermès.
- Jakobson, R. (1963). *Essais de linguistique générale* (vol. 1). Minuit.
- Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., et Suchomel, V. (2013). The TenTen corpus family. Dans *Proceedings of the 7th International Corpus Linguistics Conference (CL 2013)* (p. 125-127). https://www.sketchengine.eu/wp-content/uploads/The_TenTen_Corpus_2013.pdf
- Jespersen, O. (1917). *Negation in English and other languages*. Publisher Kgl. Danske videnskabernes selskab.
- Jespersen, O. (1922). *Language, its nature, development, and origin*. H. Holt.
- Johns, T. (1991). Should you be persuaded: two examples of data-driven learning. Dans T. Johns et P. King (dirs.), *Classroom Concordancing, English Language Research Journal*, 4 (p. 1-16). ELR.
- Jousse, A. L. et Polguère, A. (2005). *Le DiCo et sa version DiCouèbe. Document descriptif et manuel d'utilisation*. OLST, Université de Montréal. <http://idefix.ling.umontreal.ca/dicouebe/DiCoDOC.pdf>
- Kahane, S. (2003a). Sur le lien entre la définition lexicographique et les fonctions lexicales : une blessure profonde dans le DEC. Dans F. Grossman et A. Tutin (éds.), *Les collocations : analyse et traitement* (p. 61-73) . De Werelt.
- Kahane, S. (2003b). The meaning-Ttext theory. Dans V. Agel, L. Eichinger, H.-W. Eroms, P. Hellwig, H. J. Heringer et H. Lobin (éds.), *Dependency and valency*.

An international handbook of contemporary research (vol. 1, p. 546-570).
Walter de Gruyter.

- Kahane, S. et Polguère, A. (2001). Formal foundations of Lexical Functions. Dans B. Daille et G. Williams (éds), *Proceedings of "COLLOCATION: Computational Extraction, Analysis and Exploitation", 39th Annual Meeting and 10th Conference of the European Chapter of the Association for Computational Linguistics. Toulouse* (p. 8-15).
- Kalonji, M. Z. (1993). *La lexicographie bilingue en Afrique francophone*. L'Harmattan.
- Katz, J. J. et Postal, P. M. (1963). Semantic Interpretation of Idioms and Sentences Containing them. *M.I.T. Research Laboratory of Electronics. Quaterly Progress Report, 70*, 275-282.
- Kauffer, M. (2011). Actes de langage stéréotypés en allemand et en français. Pour une redéfinition du stéréotype grâce à la phraséologie. *Nouveaux Cahiers d'allemand 1*, 35-53.
- Kauffer, M. (2013). Le figement des « actes de langage stéréotypés » en français et en allemand. *Pratiques. Linguistique, littérature, didactique, 159-160*, 42-54.
- Kauffer, M. (2018). Les "actes de langage stéréotypés" : essai de synthèse critique. *Cahiers de lexicologie, 114*, 149-71.
- Kauffer, M. et Keromnes, Y. (2022). Aspects de la recherche actuelle en phraséologie. Présentation. *Langages, 225*, 9-18.
- Kay, P. et Fillmore Ch. (1999). Grammatical constructions and linguistic generalizations: the what's X doing Y? construction. *Language, 75*, 1-33.
- Kayser, D. (1987). Une sémantique qui n'a pas de sens. *Langages, 87*, 33-45.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. et Suchomel, V (2014). The Sketch Engine: ten years on. *Lexicography, 1*, 7-36.
- Kilgarriff, A., Rychlý, P., Smrž, P. et Tugwell, D. (2004). The Sketch Engine. Dans G. Williams et S. Vessier (éds.), *Proceedings of th Eleventh EURALEX International Congress. EURALEX 2004* (p. 105-116). Université de Bretagne Sud.
https://www.sketchengine.eu/wp-content/uploads/The_Sketch_Engine_2004.pdf

- Kingsbury, P. et Palmer, M. (2002). From Treebank to PropBank. Dans *Proceedings of the third international conference on Language Resources and Evaluation Conference (LREC 2002)* (p. 1989–1993). European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2002/>
- Kipper-Schuler, K., Dang, H. T. et Palmer, M. (2000). Class-based construction of a verb lexicon. *Proceedings of the seventeenth national conference on Artificial Intelligence (AAAI-00)*, Austin, USA. AAAI Press. <https://cdn.aaai.org/AAAI/2000/AAAI00-106.pdf>
- Kittredge, R. et Lehrberger, J. (éds.). (1982). *Sublanguage: studies of language in restricted semantic domains*. Walter de Gruyter.
- Kleiber, G. (1990). *La sémantique du prototype*. Presses Universitaires de France.
- Kleiber, G. (1999). *Problèmes de sémantique. La polysémie en question*. Presses Universitaires du Septentrion.
- Kleiber, G. et Vuillaume, M. (2011). Sémantique des odeurs. *Langages*, 181, 17-36.
- Klein, J. R. (2007). Le figement dans les proverbes et les expressions verbales figées : un débat qui n'est pas encore... figé. Dans G. Conde tarrío (éd.), *Nouveaux apports à l'étude des expressions figées. Nuevas aportaciones al estudio de las expresiones fijas* (p. 129-156). E.M.E. et Intercommunications.
- Klein, J. R. et Lamiroy, B. (1995). Lexique-grammaire du français de Belgique : les expressions figées. *Lingvisticae Investigationes*, 18(2), 285–320.
- Kocourek, R. (1982/1991). *La langue française de la technique et de la science*. Oscar Brandstetter Verlag.
- Koike, K. (1996). Verbos colocacionales en español. *Hispánica*, 40, 14-31.
- Koike, K. (2001). *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*. Universidad de Alcalá.
- Koike, K. (2002). Comportamientos semánticos en las colocaciones léxicas. *Lingüística española actual*, 24(1), 5-23.
- Kótova, M. (1998). As achegas rusas ós estudos fraseolóxicos. Dans X. Ferro Ruibal (coord.), *Actas do I Coloquio galego de Fraseoloxía* (p. 247-256). Xunta de Galicia.

- Kövesces, Z. (2000). *Metaphor and emotion*. Cambridge University Press.
- Kraif, O. et Tutin, A. (2020). Collocations et linguistique de corpus : l'intuition des linguistes et les critères quantitatifs convergent-ils ? *Le français moderne*, 2020-1, 84-101.
- Kubeková, J. (2018). Confronto interlinguistico della fraseologia tra l'italiano, il francese, lo spagnolo e lo slovaco nel contesto fraseodidattico. *PHRASIS - Rivista di studi fraseologici e paremiologici*, 2, 77-92.
- Lafleur, B. (1991). *Dictionnaire des locutions idiomatiques françaises*. Duculot.
- Lafourcade, M. et Joubert, A. (2008). JeuxDeMots : un prototype ludique pour l'émergence de relations entre termes. Dans S. Heden et B. Pincemin (dirs.), *JADT 2008. 9^{es} Journées internationales d'Analyse statistique des Données Textuelles* (p. 657-666). Presses Universitaires de Lyon.
- Lafourcade, M. et Joubert, A. (2013). Bénéfices et limites de l'acquisition lexicale dans l'expérience jeuxdemots. Dans N. Gala et M. Zock (dirs.), *Ressources lexicales : contenu, évaluation, utilisation, évaluation. Lingvisticae Investigationes Supplementa*, 30 (p. 187–216). John Benjamins Publishing Company.
- Lafourcade, M. et Le Brun, N. (2020). Jeuxdemots : un réseau lexico-sémantique pour le français, issu de jeux et d'inférences. *Revue Lexique*, 27, 47–86.
- Lakoff, G. (1982). *Categories and cognitive models*. Cognitive Science Program, Institute of Cognitive Studies, University of California at Berkeley.
- Lakoff, G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago University Press.
- Lamiroy, B. (2003). Les notions linguistiques de figement et de contrainte. *Lingvisticae Investigationes*, 26(1), 1-14.
- Lamiroy, B. (2008). Les expressions figées : à la recherche d'une définition. *ZFSL, Zeitschrift für französische Sprache und Literatur*, 36, 85 - 99.
- Lamiroy, B. (coord.). (2010). *Les expressions verbales figées de la francophonie : Belgique, France, Québec et Suisse*. Ophrys.
- Lamiroy, B. et Klein, J. R. (2005). Le problème central du figement est le semi-figement. *Linx*, 53, 135–154.

- Langacker, R. W. (1987). *Foundations of cognitive grammar. Volume I: theoretical prerequisites*. Stanford University Press.
- Langacker, R. W. (2000). A dynamic usage-based model. Dans M. Barlow et S. Kemmer (éds.), *Usage-based models of language* (1-60). Stanford University Press.
- Langacker, R. W. (2013). *Essentials of cognitive grammar*. Oxford University Press.
- Le Fur, D. (dir.). (2007). *Dictionnaire des combinaisons de mots. Les synonymes en contexte*. Dictionnaires Le Robert.
- Le Pesant, D. (1994). Les compléments nominaux du verbe *lire*, une illustration de la notion de "classes d'objets". *Langages*, 115, 31-46.
- Le Pesant, D. et Mathieu-Collas, M. (1998). Introduction aux classes d'objets. *Langages*, 131, 6-33.
- Le Petit Robert de la langue française 2024 (version numérique)* (2023). Éditions Le Robert.
- Leacock, C., Towell, G. et Voorhees E. M. (1996). Towards building contextual representations of word senses using statistical models. Dans B. Boguraev et J. Pustejovsky (éds.), *Corpus processing for lexical acquisition* (p. 97-113). MIT Press.
- Leclère C. (1989). Les mots ont-ils une syntaxe ? *Le français dans le monde*, n° spécial *Et la grammaire ?*, 40-49.
- Leclère C. (1998). Travaux récents en linguistique-grammaire. *Travaux de linguistique*, 37, 155-186.
- Legallois, D. (2006). La phraséologie dans la linguistique contextualiste. *Cahiers du CRISCO*, 21, 28-48.
- Legallois, D. (2012). La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? *Corpus*, 11, 31-54.
- Legallois, D. et Gréa, Ph. (2006). La grammaire de construction. *Cahiers du CRISCO*, 21, 5-27.

- Legallois, D. et Tutin, A. (2013). Vers une extension du domaine de la phraséologie. *Langages*, 189, 3-25.
- Léon J. (2007). Meaning by collocation. The Firthian filiation of Corpus Linguistics. Dans D. Kibbee (éd.), *Proceedings of ICHoLS X, 10th International Conference on the History of Language Sciences* (p. 404-415). John Benjamins Publishing Company.
- Lerat, P. (1995). *Les langues spécialisées*. Presses Universitaires de France.
- Lerat, P. (1997). Approches linguistiques des langues spécialisées. *ASp*, 15-18, 1-10. <https://journals.openedition.org/asp/2926>
- Lerat, P. (2002). Qu'est-ce qu'un verbe spécialisé ? Le cas du droit. *Cahiers de lexicologie*, 80, 201-211.
- Lerat, P. (2008). Restrictions paradigmatiques et traduction des schémas d'arguments. *Meta : Journal des traducteurs/Meta : Translators' Journal*, 53(2), 434-442.
- Lerat, P. (2012). Cohérence conceptuelle et cohésion lexicale dans le discours spécialisé. Dans P. Mogorrón Huerta et S. Mejri (dirs.), *Lenguas de especialidad, traducción, fijación / Langues spécialisées, figement, traduction* (p. 33-48). Publicaciones de la Universidad de Alicante.
- Leska, O. (1966). "Le centre" et "la périphérie" des différents niveaux de la structure linguistique. *Travaux Linguistiques de Prague*, 2, 52-57.
- Levin, B. (1993). *English verb classes and alternations*. University of Chicago Press.
- Levrat, B. (1993). *Le problème du sens dans les systèmes de traitement automatique du langage naturel : une approche alternative au travers de la paraphrase* [thèse de doctorat, Université Sorbonne Paris Nord].
- Lewis, M. (2000). *Teaching collocation: further development in the lexical approach*. Language Teaching Publications.
- L'Homme, M.-C. (1998). Le statut du verbe en langue de spécialité et sa description lexicographique. *Cahiers de lexicologie*, 73, 61-84.
- L'Homme, M.-C. (2004). *La terminologie : principes et techniques*. Presses de l'Université de Montréal.

- L'Homme, M.-C. (2005). Sur la notion de "terme". *Méta : Journal des traducteurs/Meta : Translators' Journal*, 50(4), 1112-1132.
- L'Homme, M.-C. et Bertrand, C. (2000). Specialized lexical combinations: should they be described as collocations or in terms of selectional restrictions? Dans U. Heid, S. Evert, E. Lehmann et C. Rohrer (éds.), *Proceedings of the 9th EURALEX International Congress* (p. 497-506). Institut für Maschinelle Sprachverarbeitung.
- Lim, J.-H. (1998). *La fréquence et son expression en français* [thèse de doctorat, Université Sorbonne Paris Nord].
- Liu, D. (2010). Going beyond Patterns: Involving Cognitive Analysis in the Learning of Collocations. *TESOL Quarterly*, 44, 4-30.
- Longrée, D. et Mellet, S. (2013). Les motifs : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours. *Langages*, 189, 65-79.
- Longrée, D., Loung, X. et Mellet, S. (2008). Les motifs : un outil pour la caractérisation topologique des textes. Dans S. Heiden et B. Pincemin (éds.), *JADT 2008, Actes des 9èmes journées internationales d'analyse statistique des données textuelles* (vol. 2, p.733-744). Presses Universitaires de Lyon.
- Lorenz-Bourjot, M. et Lüger, H.-H. (éds.) (2001). *Phraseologie und phraseodidaktik*. Praesens.
- Loupy, C. de, Combet, V. et Crestan E. (2003). Linguistic resources for information retrieval. Dans *Proceedings ENABLER/ELSNET International Roadmap for Language Resources, Paris 2003*.
https://www.researchgate.net/profile/Eric_Crestan/publication/228436072_Linguistic_resources_for_information_retrieval/links/02e7e537d8984256d6000000/Linguistic-resources-for-information-retrieval.pdf
- Luque Durán, J. de D. et Bertrán A. P. (éds.). (1998). *Léxico y fraseología*. Método Ediciones.
- Lux-Pogodalla, V. et Polguère, A. (2011). Construction of a French lexical network: methodological issues. Dans *First international workshop on lexical resources*

- Lyons, J. (1977). *Semantics*. Cambridge University Press.
- Makkai, A. (1972). *Idiom structure in English*. Walter de Gruyter.
- Makkai, A. (1975). The cognitive organization of idiomaticity: rhyme or reason? *Georgetown University Working Papers on Languages and Linguistics*, 11, 10-29.
- Makkai, A. (1978). Idiomaticity as a language universal. *Universal of Human Language*, 3, 401-448.
- Marcus, M. P., Marcinkiewicz, M. A. et Santorini, B. (1993). Building a large annotated corpus of English: the Penn Treebank. *Computational linguistics*, 19(2), 313–330.
- Marengo, S. (2011). Unités linguistiques et parties du discours : pour une grammaire à l'écoute de la lexicographie. Dans S. Mejri, I. Sfar et M. van Campenhoudt (dirs.), *L'unité en sciences du langage. Actes des neuvièmes journées scientifiques du réseau thématique «Lexicologie, Terminologie, Traduction»* (p. 49-59). Éditions des archives contemporaines.
- Mariani, J., Francopoulo, G., et Paroubek, P. (2019a). The NLP4NLP corpus (I): 50 years of publication, collaboration and citation in speech and language processing. *Frontiers in Research Metrics and Analytics*, 3(36).
- Mariani, J., Francopoulo, G., Paroubek, P. et Vernier, F. (2019b). The NLP4NLP corpus (ii): 50 years of research in speech and language processing. *Frontiers in Research Metrics and Analytics*, 3(37).
- Martelli, A. (2007). *Lexical collocations in learner English: a corpus-based approach*. Edizioni dell'Orso.
- Martin, R. (1983). *Pour une logique du sens*. Presses Universitaires de France.
- Martin, R. (1994). Dictionnaire informatisé et traitement automatique de la polysémie. Dans E. Martin (éd.), *Textes et informatique* (p. 77-113). Didier Erudition.

- Martin, R. (1997). Sur les facteurs de figement lexical. Dans M. Martins-Baltar (éd.), *La locution entre langage et usages* (p. 291-305). ENS Éditions.
- Martinet, A. (1966). Le mot. Dans Fonagy, Schaff et Paude (coords.), *Problèmes de Langage* (p. 39-53). Gallimard.
- Martinet, A. (1967). Syntagme et syntème. *La linguistique*, 2, 1-14.
- Martínez López, J. A. (1996). *La fraseología del español. Acercamiento morfosintáctico, semántico y pragmático* [thèse de doctorat, Universidad de Granada]. <https://digibug.ugr.es/handle/10481/14435>
- Martínez Marín, J. (1991). Las expresiones fijas verbales en el habla culta de Caracas. Dans C. Hernández et al. (éds.), *El español de América* (p. 807-813). Junta de Castilla y León.
- Martínez Marín, J. (1996). *Estudios de fraseología española*. Armand Colin.
- Martínez Marín, J. (1998) Cómo introducir los valores pragmáticos de las expresiones fraseológicas en el diccionario. Dans J. Ferro Ruibal (coord.), *Actas do I Coloquio galego de Fraseoloxía* (p. 127-137). Xunta de Galicia.
- Martins-Baltar, M. (1997). Repères dans les recherches actuelles sur la notion de locution. Dans M. Martins-Baltar (éd.), *La locution entre langage et usages* (p. 19-52). ENS Éditions.
- Massoussi, T. et Mejri, S. (2011). Traitement automatique des discours financiers : application à la bourse. Dans M. van Campenhout, T. Lino et R. Costa, *Passeurs de mots, passeurs d'espoirs. Lexicologie, terminologie et traduction face au défi de la diversité. Actes des Huitièmes Journées Scientifiques du Réseau de chercheurs Lexicologie, terminologie, traduction. Lisbonne 15-17 octobre 2009* (p. 245-258). Éditions des archives contemporaines.
- Massoussi, T. et Mejri, S. (2015). Procédure d'étiquetage des textes stratégiques : examen de deux concepts (délocalisation et internationalisation). *Synergies Tunisie*, 4, 83-99.
- Massoussi, T. et Sfar, I. (2009). Description des prédicats nominaux : de la langue générale aux langues spécialisées. *Neophilologica*, 21, 62-81.

- Matešić, J. (1983). Zum terminus und zur definition der 'phraseologischen einheit'. Dans J. Matešić (éd.), *Phraseologie und ihre Aufgaben. Beiträge zum 1. Internationalem Phraseologie-Symposium vom 12. bis 14. Oktober 1981 in Mannheim* (p. 110-116). Julius Groos Verlag.
- Mathieu-Colas, M (1996). Représentation de la polysémie dans un dictionnaire électronique. Dans A. Clas, Ph. Thoiron et H. Béjoint (dirs.), *Lexicomatique et dictionnaires. Actes du colloque de Lyon 1995* (317-325). AUPELF-UREF.
- Mathieu-Colas, M. (1994). *Les mots à trait d'union. Problèmes de lexicographie informatique*. Didier Erudition.
- Mathieu-Colas, M. (1996). Typologie de la composition nominale. *Cahiers de lexicologie*, 69, 65-118.
- Mathieu-Colas, M. (1996-2006). *Dictionnaire morphologique du français. I. Formes simples* [rapport technique du LLI, Université Paris 13].
- Mathieu-Colas, M. (2007). Domaines et classes sémantiques. *Verbum*, 29(1-2), 11-24.
- Mathieu-Colas, M. (2009). Morfetik : une ressource lexicale pour le TAL. *Cahiers de lexicologie*, 94, 137-146.
- McEnery, T. et Hardie, A. (2011). *Corpus linguistics: method, theory and practice*. Cambridge University Press.
- McIntoch, A. (1966). Patterns and Ranges. Dans A. McIntoch et M. A. K. Halliday (dir.), *Patterns of language* (p. 183-199). Longmans.
- Mejri, S. (1997). *Le figement lexical. Descriptions linguistiques et structuration sémantique* [thèse doctorat, Université de la Manouba].
- Mejri, S. (1999). Unité lexicale et polylexicalité. *Linx*, 40, 79-93.
- Mejri, S. (2002). Le figement lexical : nouvelles tendances. *Cahiers de lexicologie*, 80, 213-225.
- Mejri, S. (2003a). Le figement lexical. *Cahiers de lexicologie*, 82, 23-39.
- Mejri, S. (2004). Introduction : Polysémie et polylexicalité. *Syntaxe et Sémantique*, 5, 13-30.

- Mejri, S. (2005). Figement absolu ou relatif : la notion de degré de figement. *Linx*, 53, 183-196.
- Mejri, S. (2006). Polylexicalité, monolexicalité et double articulation : la problématique du mot. *Cahiers de lexicologie*, 89(2), 209-221.
- Mejri, S. (2007). French phraseology. Dans H. Burger, D. Dobrovolskij, P. Kühn et N. R. Norrick (éds.), *Phraseologie / Phraseology. Ein internationales handbuch zeitgenössischer forschung / An international handbook of contemporary research* (p. 682-691). Walter de Gruyter.
- Mejri, S. (2008a). Constructions à verbes supports, collocations et locutions verbales. Dans P. Mogorrón Huerta et S. Mejri (éds.), *Las construcciones verbo-nominales libres y fijas. Aproximación contrastiva y traductológica* (p. 191-202). Publicaciones de la Universidad de Alicante.
- Mejri, S. (2008b). La place du figement dans la description des langues. Dans P. Blumenthal et S. Mejri (éds.), *Les séquences figées : entre langue et discours* (P. 117-129). Franz Steiner Verlag.
- Mejri, S. (2011a). Collocations et emplois appropriés : des unités lexicales hybrides ? *Cahiers de lexicologie*, 2011(1), 83-94.
- Mejri, S. (2011b). Figement, collocation et combinatoire libre. Dans J.-C. Anscombe et S. Mejri (dirs.), *Le figement linguistique : la parole entravée* (p. 63-77). Honoré Champion.
- Mejri, S. (2011c). Phraséologie et traduction des textes spécialisés. Dans P. Mogorrón Huerta et C. Gonzalez-Roy (éds.), *Estudios y análisis de fraseología contrastiva: lexicografía, traducción y análisis de corpus* (p. 125-138). Publicaciones de la Universidad de Alicante.
- Mejri, S. (2012). Phraséologie et traduction. Pour une typologie des phraséogismes dans les discours spécialisés. Dans P. Mogorrón Huerta et S. Mejri (dirs.), *Lenguas especializadas, fijación y traducción / Langues spécialisées, figement et traduction* (p. 19-31). Publicaciones de la Universidad de Alicante.
- Mejri, S. (2018a). La phraséologie : cotexte, contexte et contenus culturels. *Lublin studies in modern languages and literature*, 42(4), 11-38.

- Mejri, S. (2018b). La phraséologie française : synthèse, acquis théoriques et descriptifs. *Le français moderne*, 2018-1, 5-32.
- Mejri, S. (2019). Figement et relation concessive : une prédication complexe. *Thélème. Revista Complutense de Estudios Franceses*, 34(1), 109-124.
- Mejri, S. (dir.). (2003b). *Le figement lexical. Cahiers de lexicologie*, 82-1.
- Mejri, S. (dir.). (2018). *La phraséologie française. Le français moderne*, 2018-1.
- Mel'čuk, I. (1973). Linguistic theory and "meaning-text" type models. Dans R. Bogdan et I. Niiniluoto (éds.), *Logic, language and probability* (p. 223-235). Reidel.
- Mel'čuk, I. (1981). Un nouveau type de dictionnaire : le *Dictionnaire explicatif et combinatoire du français contemporain* (six articles de dictionnaire). *Cahiers de lexicologie*, 38(1), 3-34.
- Mel'čuk, I. (1988). *Dependency syntax: theory and practice*. State University of New York Press.
- Mel'čuk, I. (1993). La phraséologie et son rôle dans l'enseignement-apprentissage d'une langue étrangère. *Études de Linguistique Appliquée*, 92, 82-113.
- Mel'čuk, I. (1994). Typologie des phrasèmes et leur présentation dans un dictionnaire. *Conférence au Colloque International "La locution"*. Paris, ENS Saint-Cloud.
- Mel'čuk, I. (1995). Phrasemes in language and phraseology in linguistics. Dans M. Everaert, E.-J. van der Linden, A. Schenk and R. Schreuder (éds), *Idioms. Structural and psychological perspectives* (p. 167-232). Lawrence Erlbaum Associates.
- Mel'čuk, I. (1996). Lexical Functions: a tool for the description of lexical relations in a lexicon. Dans L. Wanner (dir.), *Lexical Functions in Lexicography and Natural Language Processing* (p. 37-102). John Benjamins Publishing Company.
- Mel'čuk, I. (1997). *Vers une linguistique Sens-Texte. Leçon inaugurale*. Collège de France.
- Mel'čuk, I. (1998). Collocations and Lexical Functions. Dans P. A. Cowie (éd.), *Phraseology. Theory, Analysis, and Applications* (23-53). Clarendon Press.

- Mel'čuk, I. (2003a). Collocations dans le dictionnaire. Dans T. Szende (éd.), *Les écarts culturels dans les dictionnaires bilingues* (p. 19-64). Honoré Champion.
- Mel'čuk, I. (2003b). Collocations : définition, rôle et utilité. Dans F. Grossmann et A. Tutin (éds.), *Les collocations : analyse et traitement* (p. 23-32). De Werelt.
- Mel'čuk, I. (2004a). Actants in Semantics and Syntax I. Actants in Semantics. *Linguistics*, 42(1), 1-66.
- Mel'čuk, I. (2004b). Actants in Semantics and Syntax II. Actants in Syntax. *Linguistics*, 42(2), 247-291.
- Mel'čuk, I. (2004c). La non-compositionnalité en morphologie linguistique. *Verbum*, 26 (4), 439-458.
- Mel'čuk, I. (2004d). Verbes support sans peine. *Lingvisticae Investigationes*, 27(2), 203-217.
- Mel'čuk, I. (2006a). Explanatory Combinatorial Dictionary. Dans G. Sica (éd.), *Open Problems in Linguistics and Lexicography* (p. 225-355). Polimetrica Publisher.
- Mel'čuk, I. (2006b). Parties du discours et locutions. *Bulletin de la Société de linguistique de Paris*, 101, 29-65.
- Mel'čuk, I. (2011). Phrasèmes dans le dictionnaire. Dans J.-C. Anscombe et S. Mejri (dirs.), *Le figement linguistique : la parole entravée* (p. 41-61). Honoré Champion.
- Mel'čuk, I. (2012). Phraseology in the language, in the dictionary and in the computer. Dans K. Kuiper (éd.), *The Yearbook of Phraseology*, 3 (p. 31-56).
- Mel'čuk, I. (2013). Tout ce que nous voulions savoir sur les phrasèmes, mais... *Cahiers de lexicologie*, 102(1), 129-149.
- Mel'čuk, I. (2015). Clichés, an understudied subclass of phrasemes. *Yearbook of Phraseology*, 6, 55-86.
- Mel'čuk, I., Clas, A. et Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Duculot.

- Mel'čuk, I. et al. (1984). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques* (vol. 1). Presses de l'Université de Montréal.
- Mel'čuk, I. et al. (1988). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques* (vol. 2). Presses de l'Université de Montréal.
- Mel'čuk, I. et al. (1992). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques* (vol. 3). Presses de l'Université de Montréal.
- Mel'čuk, I. et al. (1999). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques* (vol. 4). Presses de l'Université de Montréal.
- Mel'čuk, I. et Iordanskaja, L. (2017). *Le mot français dans le lexique et dans la phrase*. Hermann Éditeurs.
- Mel'čuk, I. et Milićević, J. (2014). *Introduction à la linguistique* (vol. 1). Hermann Éditeurs.
- Mel'čuk, I. et Polguère A. (1987). A formal lexicon in the Meaning-Texte Theory (or how to do lexica with words). *Computational Linguistics*, 13(3-4), 261-275.
- Mel'čuk, I. et Polguère A. (2007). *Le lexique actif du français*. De Boeck & Larcier.
- Mel'čuk, I. et Polguère, A. (2008). Prédicats et quasi-prédicats sémantiques dans une perspective lexicographique. *Revue de linguistique et de didactique des langues (Lidil)*, 37, (99-114).
- Mel'čuk, I. et Polguère, A. (2021). Les fonctions lexicales dernier cri. Dans S. Marengo (dir.), *La théorie Sens-Texte. Concepts et applications* (p. 75-155). L'Harmattan.
- Mel'čuk, I. et Wanner, L. (1996). Lexical Functions and lexical inheritance for emotion lexemes in German. Dans L. Wanner (éd.), *Lexical Functions in Lexicography and Natural Language Processing* (p. 209-279). John Benjamins Publishing Company.
- Mellado Blanco, C. (1997). Fraseologismos alemanes y españoles del campo de las emociones. *Paremia*, 6, 383-388.

- Mellado Blanco, C. (1999). *Los somatismos del alemán: semántica y estructura* [thèse de doctorat, Universidad de Salamanca].
- Mendívil Giró, J. L. (1998). Aspectos teóricos del estudios de las unidades fraseológicas : gramática, pragmática y fraseología. Dans G. Wotjak (éd.), *Estudios de fraseología y fraseografía del español actual* (p. 39-55). Vervuert/Iberoamericana.
- Mendívil Giró, J. L. (1999). *Las palabras disgregadas. Sintaxis de las expresiones idiomáticas y los predicados complejos*. Prensas Universitarias de Zaragoza.
- Merle, J.-M. (2009). La prédication : présentation générale. *Faits de langue*, 31-32, 5-12.
- Mertens, P. (2010) Restrictions de sélection et réalisations syntagmatiques dans DICOVALENCE. Conversion vers un format utilisable en TAL. *Actes TALN 2010 (Montréal, 19-23 juillet 2010)*. <https://aclanthology.org/2010.jeptalnrecital-long.32/>
- Mešková, L. et Kubeková, J. (2019). Les enjeux de l'intercompréhension pour la phraséologie en langues romanes en comparaison avec les langues slaves. *Repères DoRiF*, 18. <https://www.dorif.it/reperes/category/18-phraseodidactique-de-la-conscience-a-la-competence/>
- Meunier, A. (1981). *Nominalisations d'adjectifs par verbes supports* [thèse de 3^e cycle, Université Paris 8].
- Meyer, I. et Mackintosh, K. (1994). Phraseological analysis and conceptual analysis: exploring a symbiotic relationship in the specialized lexicon. Dans W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg et P. Vossen. (éds.), *EURALEX'94 Proceedings*. Vrije Universiteit Amsterdam.
- Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B. et Grishman, R. (2004). The NomBank project : an interim report. Dans *Frontiers in corpus annotation. Proceedings of the workshop. HLT/NAACL 2004* (p. 24-31). <https://aclanthology.org/W04-2705.pdf>
- Mikolov, T., Yih, W.-T. et Zweig, G. (2013). Linguistic regularities in continuous space word representations. Dans *Proceedings of the 2013 Conference of the North*

- American Chapter of the Association for Computational Linguistics (NAACL)* (p. 746–751). Association for Computational Linguistics. <https://aclanthology.org/N13-1.pdf>
- Milicévic J. (1997). *Étiquettes sémantiques dans un dictionnaire formalisé du type Dictionnaire Explicatif et Combinatoire* [mémoire Maitre ès arts, Université de Montréal]. http://olst.ling.umontreal.ca/pdf/Milicevic_1997_memoire.pdf
- Milicévic J., Polguère A. (2010). Ambivalence sémantique des noms de communication langagière en français. Dans F. Neveu, V. Muni Toke, J. Durand, T. Klingler, L. Mondada et S. Prévost (éds.), *Congrès mondial de linguistique française 2010* (p. 1029-1050). https://www.linguistiquefrancaise.org/articles/cmlf/abs/2010/01/cmlf2010_000102/cmlf2010_000102.html
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39-41.
- Mitkov, R. (éd.) (2017). *Computational and corpus-based phraseology: second international conference, Europhras 2017, London, UK, November 13-14, 2017, Proceedings*. Springer.
- Mitkov, R., Monti, J., Corpas Pastor, G. et Seretan V. (éds.) (2018). *Multiword units in machine translation and translation technology*. John Benjamins Publishing Company.
- Mogorrón Huerta, P. (2002). *La expresividad en las locuciones verbales en francés y en español*. Publicaciones de la Universidad de Alicante.
- Mogorrón Huerta, P. (2010). Analyse du figement et de ses possibles variations dans les constructions verbales espagnoles. *Lingvisticae Investigationes*, 33(1), 86-151.
- Mokienko, V. (2000a). *As imaxes da lingua rusa. Ensayos históricos, etimológicos e etnolingüísticos sobre fraseoloxía* (traduit par E. Lossik). Centro Ramón Piñeiro.
- Mokienko, V. (2000b). *Fraseoloxía eslava. Manual universitario para a especialidade de lingua e literatura rusas* (traduit par K. Guerbek). Centro Ramón Piñeiro.
- Moliner, M. (2007). *Diccionario de uso del español* (3^e éd.). Gredos.

- Molinero, M., Sagot, B. et Nicolas, L. (2009a). A morphological and syntactic wide-coverage lexicon for Spanish: the Leffe. Dans G. Angelova, K. Bontcheva, R. Mitkov, N. Nicolov et N. Nicolov, *Proceedings of the international conference "Recent Advances in Natural Language Processing 2009"* (p. 264-269). <https://aclanthology.org/R09-1049/>
- Molinero, M., Sagot, B. et Nicolas, L. (2009b). Construcción y extensión de un léxico morfológico y sintáctico para el español: el Leffe. *Procesamiento del lenguaje natural*, 43, 335-343.
- Montague, R. (1970). English as a formal language. Dans Bruno Visentini (éd.), *Linguaggi nella società e nella tecnica* (p. 188-221). Edizioni di Comunità.
- Montoro del Arco, E. T. (2006). *Teoría fraseológica de las locuciones particulares. Las locuciones prepositivas, conjuntivas y marcadoras en español*. Peter Lang.
- Moon, R. (1992). There is reason in the roasting of eggs': a consideration of fixed expressions in native-speaker dictionaries. Dans H. Thommola, K. Varantola, T. Salmi-Tolonene et J. Schopp, *EURALEX'92 proceedings. Papers submitted to the 5th EURALEX international congress on lexicography in Tampere, Finland* (p. 493-502). University of Tampere.
- Moon, R. (1998). *Fixed expressions and idioms in english: a corpus-based approach*. Oxford University Press.
- Mortureux, M.-F. (2003). Figement lexical et lexicalisation. *Cahiers de lexicologie*, 82(1), 11-22.
- Mortureux, M.-F. et Petit, G. (1989). Fonctionnement du vocabulaire dans la vulgarisation et problèmes de lexicologie. *DRLAV*, 40, 41-62.
- Mounin, G. (1979). La linguistique comme science auxiliaire dans les disciplines juridiques. *Meta : Journal des traducteurs/Meta : Translators' Journal*, 24(1), 9-17.
- Muller, C. (2013). Le prédicat, entre (méta)catégorie et fonction. *Cahiers de lexicologie*, 102(1), 51-65.
- Namer, F. (2000). FLEMM : un analyseur flexionnel du français à base de règles. *Traitement automatique des langues*, 41(2), 523-547.

- Náray-Szabó, M. (2002). Quelques remarques sur la définition de phrasème. *Revue d'études françaises*, 7, 71-81.
- Navarro-Brotons, L. (2013). *Las paremias y sus variantes: análisis sintáctico, semántico y traductológico español/francés* [thèse de doctorat, Universidad de Alicante].
- Navarro-Brotons, L. (2023). *Paremias: estudio lingüístico contrastivo francés-español*. John Benjamins Publishing Company.
- Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. John Benjamins Publishing Company.
- Neveu, F. (2011). *Dictionnaire des sciences du langage* (2^e éd.). Armand-Colin.
- New, B. et Pallier, C. (2005). *Manuel de Lexique 3. Document version 3.11*. Laboratoire de psychologie expérimentale, Université René-Descartes. http://lexique.org/_documentation/Manuel_Lexique.3.pdf
- Nuccorini, S. (1992). Holding a post and filling a job. English collocations in English-Italian, Italian-English dictionaries. Dans M. Alvar Ezquerra (éd.), *EURALEX'90. Proceedings* (p. 369-380). Bibliograf.
- Nunberg, G. et Zaenen, A. (1997). La polysémie systématique dans la description lexicale. *Langue française*, 113, 12-23.
- Nunberg, G., Sag, I. A. et Wasow, T. (1994). Idioms. *Language*, 70, 491-538.
- Olza Moreno, I. (2009). *Aspectos de la semántica de las unidades fraseológicas. La fraseología somática metalingüística del español* [thèse de doctorat, Universidad de Navarra].
- Orol, A. et Alonso Ramos, M. (2013). A comparative study of collocations in a native corpus and a learner corpus of Spanish. *Procedia-Social and Behavioural Sciences*, 96, 563-570.
- Palmer, H. E. (1933). *Second interim report on English collocations*. Kaitakusha.
- Palmer, M., Gildea, D. et Kingsbury, P. (2005). The Proposition Bank: a corpus annotated with semantic roles. *Computational Linguistics Journal*, 31(1), 71-106.

- Palmer, M., Gildea, D. et Xue, N. (2010). Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3(1), 1-103.
- Paul, H. (1880/1966). *Prinzipien der sprachgeschichte*. Max Niemeyer.
- Paumier, S. (2016). *Manuel d'utilisation du logiciel Unitex*. Université Paris-Est Marne-la-Vallée. <http://unitexgramlab.org/releases/3.1/man/Unitex-GramLab-3.1-usermanual-fr.pdf>
- Pawley, A. et Sider, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. Dans J. C. Richards et R. W. Schmidt (éds.), *Language and communication* (p. 191-227). Addison Wesley Longman Limited.
- Pecina, P. et Schlesinger, P. (2006). Combining association measures for collocation extraction. Dans *Proceedings of the 21st international conference on computational linguistics and 44th annual meeting of the association for computational linguistics (COLING/ACL 2006), Poster Sessions*, (p. 651–658). ACL.
- Pecman, M. (2004). L'enjeu de la classification en phraséologie. *Actes du congrès Europhras 2004, 26-29 août 2004, Université de Bâle* (p. 127-146). Schneider Hohengehren Verlag.
- Pecman, M. (2007). Approche onomasiologique de la langue scientifique générale. *Revue française de linguistique appliquée*, 12(2), 76-96.
- Pecman, M. et Klüber, N. (2011). ARTES: An online lexical database for research and teaching in specialized translation and communication. Dans *Proceedings from international workshop on lexical resources (WoLeR). Ljubljana, Slovénie. August 1–5, 2011* (p. 87-93). http://alpage.inria.fr/~sagot/pub/WoLeR_2011_proceedings.pdf#page=87
- Penadés Martínez, I. (1999). *La enseñanza de las unidades fraseológicas*. Arco Libros.
- Penadés Martínez, I. (2002). *Diccionario de locuciones verbales para la enseñanza del español*. Arco Libros.
- Penadés Martínez, I. (2005a). *Diccionario de locuciones adverbiales para la enseñanza del español*. Arco Libros.

- Penadés Martínez, I. (2005b). Resultados y perspectivas de estudio en fraseología española. *LynX. Panorámica de Estudios Lingüísticos*, 4, 5-58.
- Penadés Martínez, I. (2008). *Diccionario de locuciones nominales, adjetivas y pronominales para la enseñanza del español*. Arco Libros.
- Pennington, J., Socher, R. et Manning, C. (2014). GloVe : Global vectors for word representation. Dans A. Moschitti, B. Pang et W. Daelemans (éds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 1532–1543. <https://aclanthology.org/D14-1162/>
- Peris Morant, A. et Taulé Delor, M. (2011). AnCora-Nom: A Spanish lexicon of deverbal nominalizations. *Procesamiento del Lenguaje Natural*, 46, 11-18
- Peytard, J., Jacobi, D. et Petroff, A. (éds.). (1984). Français technique et scientifique : reformulation, enseignement, *Langue française*, 64.
- Picht, H. et Draskau, J. (1985). *Terminology. An introduction*. University of Surrey, Department of Linguistics and International Studies.
- Picoche, J. (1992). *Précis de lexicologie française - L'étude et l'enseignement du vocabulaire*. Nathan Université.
- Pierrel, J.-M. (2006). Le Trésor de la Langue Française Informatisé : un dictionnaire de référence accessible à tous. *AMOPA*, 174, 25-28.
- Pigeon, L. P. (1982). La traduction juridique. L'équivalence fonctionnelle. Dans J. C. Gémar (éd.), *Langage du droit et traduction* (p. 271-281). Linguattech.
- Piirainen, E. (2008). Phraseology in a European framework: a cross-linguistic and cross-cultural research project on widespread idioms. Dans S. Granger et F. Meunier (éds.), *Phraseology: an interdisciplinary perspective* (p. 243-258). John Benjamins Publishing Company.
- Piirainen, E. (2012). *Widespread idioms in Europe and beyond : toward a lexicon of common figurative units* (vol. 1). Peter Lang.
- Piirainen, E. (2016). *Lexicon of common figurative units. Widespread idioms in Europe and beyond*. (vol. 2). Peter Lang.
- Pineaux, J. (1967). *Proverbes et dictons français*. PUF.

- Polguère, A. (1990). *Structuration et mise en jeu procédurale d'un modèle linguistique dans un cadre de génération de texte* [thèse de doctorat, Université de Montréal]. <http://olst.ling.umontreal.ca/pdf/PolguerePhD1990.pdf>
- Polguère, A. (1998a). La théorie sens-texte. *Dialangue*, 8-9, 71-83.
- Polguère, A. (1998b). Pour un modèle stratifié de la lexicalisation en génération de texte. *T.A.L.*, 39(2), 57-76.
- Polguère, A. (2000a). Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. Dans U. Heid, S. Evert, E. Lehmann et C. Rohrer (éds.), *Proceedings of the 9th EURALEX International Congress* (p. 517–527). Institut für Maschinelle Sprachverarbeitung.
- Polguère, A. (2000b). Une base de données lexicales du français et ses applications possibles en didactique. *Revue de linguistique et de didactique des langues (Lidil)*, 21, 75-97
- Polguère, A. (2003). Étiquetage sémantique des lexies dans la base de données DiCo. *Traitement Automatique des Langues n°44(2)*, 39-68.
- Polguère, A. (2007). Soleil insoutenable et chaleur de plomb : le statut linguistique des greffes collocationnelles. Version officielle non publiée, en ligne, <http://olst.ling.umontreal.ca/pdf/Greffescolloc2007.pdf> (consulté le 25/04/2023)
- Polguère, A. (2008a). *Lexicologie et sémantique lexicale. Notions fondamentales*. Les Presses de l'Université de Montréal.
- Polguère, A. (2008b). Pour un transfert des savoirs lexicographiques. Dans J. Durand, B. Habert et B. Laks (éds.), *Congrès Mondial de Linguistique Française - CMLF'08* (p. 1277-1285). <https://www.linguistiquefrancaise.org/articles/cmlf/pdf/2008/01/cmlf08331.pdf>
- Polguère, A. (2009). Lexical systems: graph models of natural language lexicons. *Language Resources and Evaluation*, 43(1), 41–55.
- Polguère, A. (2011a). Classification sémantique des lexies fondée sur le paraphrasage. *Cahiers de lexicologie*, 9, 197-211.

- Polguère, A. (2011b). Figement et ellipse dans une perspective lexicographique : le cas de *dé à jouer* et *dé à coudre*. Dans J.C. Anscombe et S. Mejri (dirs.), *Le figement linguistique : la parole entravée* (p. 365-375). Honoré Champion.
- Polguère, A. (2014). From writing dictionaries to weaving lexical networks. *International Journal of Lexicography*, 27(4), 396-418.
- Polguère, A. (2015a). Lexicon Embedded Syntax. Dans J. Nivre et E. Hajičová (éds.), *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)* (p. 2–9). Upsala University - Department of Linguistics and Philology. <https://aclanthology.org/W15-21.pdf>
- Polguère, A. (2015b). Non-compositionnalité : ce sont toujours les locutions faibles qui trinquent. *Verbum*, 37(2), 257-280.
- Polguère, A. (2016). La question de la géométrie du lexique. Dans F. Neveu, G. Bergounioux, M.-H. Côté, J.-M. Fournier, L. Hriba et S. Prévost (éds.), *SHS Web Conferences (vol. 27). 5e Congrès Mondial de Linguistique Française (CMLF 2016)*. https://www.shs-conferences.org/articles/shsconf/abs/2016/05/shsconf_cmlf2016_01002/shsconf_cmlf2016_01002.html
- Pottier, B. (1974). *Linguistique générale, théorie et description*. Klincksieck.
- Pottier, B. (1987). *Théorie et analyse en linguistique*. Hachette.
- Pradet, Q. (2015). *Annotation en rôles sémantiques du français en domaine spécifique* [thèse de doctorat, Université Paris-Diderot]. https://inria.hal.science/tel-01182711/file/These_QuentinPradet_2015-08-15.pdf
- Pradet, Q., Danlos, L. et de Chalendar G. (2014). Adapting VerbNet to French using existing resources. Dans *Proceedings of the Ninth international conference on Language Resources and Evaluation Conference (LREC'14)* (p. 1122-1126). European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2014/index.html>
- Pradhan, S., Bonn, J., Myers, S., Conger, K., O'Gorman, T., Gung, J. et Palmer, M. (2022). Prop-Bank comes of age. Larger, smarter, and more diverse. Dans *Proceedings of the 11th joint conference on lexical and computational semantics*

- (p. 278-288). Association for Computational Linguistics.
<https://aclanthology.org/2022.starsem-1.24.pdf>
- Prandi, M. (2004). *The building blocks of meaning. Ideas for a philosophical grammar*. John Benjamins Publishing Company.
- Prandi, M. (2016). Les métaphores conflictuelles dans la création de concepts et de termes. *Langue française*, 189, 35-48.
- Prandi, M., Giaufret, A. et Rossi, M. (éds.). (2012). *Il ruolo della metafora nella creazione di terminologie*. GUP.
- Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press.
- Pustejovsky, J., Castaño, J. M., Ingria, R., Sauri, R., Gaizauskas, R., Setzer, A., Katz, G. et Radev, D. (2003). TimeML: robust specification of event and temporal expressions in text. Dans *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)* (p. 28-34).
<https://cdn.aaai.org/Symposia/Spring/2003/SS-03-07/SS03-07-005.pdf>
- Pustejovsky, J., Hanks, P., Saur, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L. et Lazo, M. (2003). The TIMEBANK Corpus. Dans D. Archer, P. Rayson, A. Wilson et T. McEnery, *Proceedings of the Corpus Linguistics conference 2003* (p. 647-656).
<https://ucrel.lancs.ac.uk/publications/cl2003/contents.htm>
- Pustejovsky, J., Verhagen, M., Sauri, R., Littman, J., Gaizauskas, R., Katz, G., Mani, I., Knippen, R. et Setzer, A. (2006). *TimeBank 1.2 LDC2006T08*. Linguistic Data Consortium. <https://catalog.ldc.upenn.edu/LDC2006T08>
- Quemada, B. (1978). Technique et langage. La formation des vocabulaires français de technique. Dans B. Gille (dir.), *Histoire des techniques* (p. 1146-1240). Gallimard.
- Quiniou, S., Cellier, P., Charnois, T., et Legallois, D. (2012). What about sequential data mining techniques to identify linguistic patterns for stylistics? Dans *Proceedings of the 13th international conference on intelligent text processing and computational linguistics* (p. 166-177). Springer.

- Ramajo Caño, A. (1987). *Las gramáticas de la lengua castellana desde Nebrija a Correas*. Universidad de Salamanca.
- Rappaport, M. (1983). On the nature of derive nominals. Dans L. Levin, M Rappaport et A. Zaenen (éds.), *Papers in Lexical-Functional Grammar* (p. 113-142). Indiana University Linguistics Club.
- Real Academia Española. *Diccionario de la lengua española* (23^e éd., version 23.6 en ligne). <https://dle.rae.es>
- Rey, A. (1976). Présentation. Dans A. Rey (éd.), *Néologie en marche, série b : langues de spécialité 2*. Éditeur officiel du Québec.
- Rey, A. (1985). Les implications théoriques d'un dictionnaire phraaséologique. *Le moyen français*, 14, 119-133.
- Rey, A. (1995). *Essays in terminology* (traduit par J. C. Sager). John Benjamins Publishing Company.
- Rey, A. et Chantreau S. (1990). *Dictionnaire des expressions et locutions*. Les usuels du Robert.
- Rey-Debove, J. (1972). La métalangue comme système de référence au signe. *Le français moderne*, 40, 232-241.
- Riegel, M., Pellat, J.-C. et Rioul, R. (1994). *Grammaire méthodique du français*. Paris : Presses Universitaires de France.
- Rodríguez Díez, B. (1981). *Las lenguas especiales. El léxico del ciclismo*. Colegio Universitario de León.
- Rodríguez, H., Farwell, D., Farreres, J., Bertran, M., Alkhalifa, M., Martí, M. A., Black, W., Elkateb, S., Kirk, J., Pease, A., Vossen, P. et Fellbaum, C. (2008). Arabic WordNet: current state and future extensions. Dans A. Tanács, D. Csendes, V. Vincze, C. Fellbaum et P. Vossen (éds.), *Proceedings of the Fourth Global WordNet Conference (GWC 2008)* (p. 387-405). <https://www.cs.upc.edu/~nlp/papers/rodriguez08.pdf>
- Rodríguez-Piñero Alacalá, A. I. et García Antuña, M. (2009). Lenguas de especialidad y lenguas para fines específicos: precisiones conceptuales y terminológicas e implicaciones didácticas. Dans A. Vera et I. Martínez (éds.), *El español en*

- contextos específicos: enseñanza e investigación* (vol. 2, p. 907-932). Fundación Comillas et ASELE.
- Romary, L., Salmon-Alt, S. et Francopoulo, G. (2004). Standards going concrete: from LMF to Morphalou. Dans *Proceedings of the workshop on enhancing and using electronic dictionaries, COLING-2004, Geneva, Switzerland*, (p. 22–28). <https://aclanthology.org/W04-2104/>
- Rondeau, G. (1984). *Introduction à la terminologie*. Gaëtan Morin.
- Ruiz Gurillo, I. (1997). Aspectos de fraseología teórica española. Anejo XXIV de *Cuadernos de Filología*. Universitat de València.
- Ruiz Gurillo, I. (1998). *La fraseología del español coloquial*. Ariel.
- Ruiz Gurillo, I. (2001). *Las locuciones en español actual*. Arco/Libros.
- Ruiz Gurillo, I. (2002). Cuadernos de bitácora para navegantes fraseólogos. *Verba*, 29, 403-419.
- Ruppenhofer, J., Ellsworth, M., Petruck, M., Johnson, C. R., Baker, C. et Scheffczyk, J. (2016). *FrameNet II: Extended theory and practice*. <https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf>.
- Russel, S. J. et Norvig, P. (1995). *Artificial Intelligence: a modern approach*. Prentice Hall.
- Ruwet, N. (1983). Du bon usage des expressions idiomatiques dans l'argumentation en syntaxe générative. *Recherches linguistiques*, 11, 5-84.
- Sager, J. C. (1990). *A practical course in terminology processing*. John Benjamins Publishing Company.
- Sager, J. C., Dungworth, D. et McDonald, P. (1980). *English special languages. Principles and practice in science and technology*. Brandstetter.
- Sagot, B. (2010). The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. Dans N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner et D. Tapias, *Proceedings of the seventh international conference on Language Resources and Evaluation (LREC 2010)* (p. 2744-2751). European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2010/>

- Sagot, B. (2018). Informatiser le lexique. Modélisation, développement et exploitation de lexiques morphologiques, syntaxiques et sémantiques [Habilitation à diriger les recherches, Sorbonne Université]. <https://inria.hal.science/tel-01895229/document>
- Sagot, B. et Danlos, L. (2008). Méthodologie lexicographique de constitution d'un lexique syntaxique de référence pour le français. *Actes du colloque "Lexicographie et informatique : bilan et perspectives"*, Nancy. <https://hal.inria.fr/inria-00524742/document>
- Sagot, B. et Fišer, D. (2008). Building a free French WordNet from multilingual resources. Dans *OntoLex 2008*, Marrakech, Maroc. <https://inria.hal.science/inria-00614708/document>
- Sagot, B. et Fišer, D. (2011). Extending wordnets by learning from multiple resources. Dans *Proceedings of the 5th Language and Technology Conference (LTC 2011)*, Poznań, Pologne. <https://inria.hal.science/hal-00655785/document>
- Sagot, B. et Fišer, D. (2014). Classification-based extension of wordnets from heterogeneous resources. Dans Z. Vetulani et J. Mariani (éds.), *Human Language Technology Challenges for Computer Science and Linguistics. 5th Language and Technology Conference, LTC 2011. Poznań, Poland, November 25-27, 2011. Revised Selected Papers* (p. 396–407). Springer.
- Samardžić, T., van der Plas, L., Kashaeva, G., et Merlo P. (2010). The scope and the sources of variation in verbal predicates in English and French. Dans M. Dichinson, K. Müürisep et M. Passarotti, *Proceedings of the ninth international workshop on Treebanks and Linguistic Theories* (p. 199–211). Northern European Association for Language Technology (NEALT).
- Saracho Arnáiz, M. (2015). *La fraseología del español: una propuesta de didactización para la clase de ELE basada en los somatismos* [thèse de doctorat, Universidade de Santiago de Compostela].
- Sauri, R. et Badia, T. (2012). *Spanish TimeBank 1.0 LDC2012T12*. Linguistic Data Consortium. <https://catalog.ldc.upenn.edu/LDC2012T12>

- Saussure, F. de (1916/1978). *Cours de linguistique générale*. Payot.
- Schafroth, E. (2019). FRAME : Fraseologia multilingue elettronica : i fondamenti teorici. *Repères DoRiF*, 18. <https://www.dorif.it/reperes/elmar-schafroth-frame-fraseologia-multilingue-elettronica-i-fondamenti-teorici/>
- Schapira, C. (1999). *Les stéréotypes en français : proverbes et autres formules*. Ophrys.
- Schifko, P. (2001). ¿Existen las lenguas de especialidad? Dans M. Bargalló, E. Forgas, C. Garriga, A. Rubio et J. Schnitzer (éds.), *Las lenguas de especialidad y su didáctica: Actas del Simposio Hispano-Austríaco* (p. 21-30). Servicio de Publicaciones Universitat Rovira i Virgili.
- Schmale, G. (2013). Qu'est-ce qui est préfabriqué dans la langue ? - Réflexions au sujet d'une définition élargie de la préformation langagière. *Langages*, 189, 27-45.
- Schuler, K. K. (2005). *VerbNet: a broad-coverage, comprehensive verb lexicon* [thèse de doctorat, University of Pennsylvania].
- Scott, M. et Tribble, C. (2006). *Textual patterns: key words and corpus analysis in language education*. John Benjamins Publishing Company.
- Sechehaye, A. (1921). Locutions et composés. *Journal de psychologie normale et pathologique*, 18, 654-675.
- Seco, M., Andrés, O. et Ramos, G. (2004). *Diccionario fraseológico documentado del español actual: locuciones y modismos españoles*. Aguilar.
- Selva T., Verlinde S. et Binon J. (2003). Vers une deuxième génération de dictionnaires électroniques. *Traitement Automatique des Langues* 44(2), 177-197.
- Sevilla Muñoz, J. (1988). *Hacia una aproximación conceptual de las paremias francesas y españolas*. Editorial de la Universidad Complutense.
- Sevilla Muñoz, J. (1992a). La terminologie parémiologique française et sa correspondance espagnole. *Terminologie et traduction*, 2/3, 331-343.
- Sevilla Muñoz, J. (1992b). Los "dictons": propuesta de sistematización. *Revista dde filología francesa*, 1, 175-187.
- Sevilla Muñoz, J. (1993). Las paremias españolas: clasificación, definición y correspondencia francesa. *Paremia*, 2, 15-20.

- Sevilla Muñoz, J. et Arroyo Ortega, A. (1993). La noción de "expresión idiomática" en francés y en español. *Thélème. Revista complutense de estudios franceses*, 4, 247-261.
- Sevilla Muñoz, J. et Cantera Ortiz de Urbina, J. (2004). *Diccionario temático de locuciones francesas con su correspondencia española*. Gredos.
- Sevilla Muñoz, J. et González Rodríguez, A. (1994). La traducción y la didáctica de las expresiones idiomáticas (francés-español). *Équivalences*, 24(2), 171-182.
- Sfar, I. (2007). *Les racines prédicatives : entre équivalence et rupture d'emplois* [thèse de doctorat, Université Sorbonne Paris Nord].
- Sfar, I. (2010). L'analyse prédicative : un outil pour l'évaluation de la traduction des textes bilingues alignés. *Synergies Tunisie*, 2, 99-113.
- Sfar, I. (2012). Les collocations dans les discours spécialisés : le cas de la terminologie ornithologique. *Language Design*, 14, 19-38.
- Siepmann, D. (2006). Collocations et dictionnaires d'apprentissage onomasiologiques bilingues : questions aux théoriciens et pistes pour l'avenir. *Langue française*, 107, 99-118.
- Silberztein, M. (1993a). *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX*. Masson.
- Silberztein, M. (2005). NooJ: a linguistic annotation system for corpus processing. Dans D. Byron, A. Venkataraman et D. Zhang (éds.), *Proceedings of HLT/EMNLP on interactive demonstrations* (p. 10-11). Association for Computational Linguistics. <https://aclanthology.org/H05-2006/>
- Sinclair, J. (1966). Beginning the study of lexis. Dans C. E. Bazell et al. (dir.), *In Memory of John Firth* (p. 410-430). Longmans.
- Sinclair, J. (2003). *Reading concordances. An introduction*. Longman.
- Sinclair, J. (2004). *Trust the text*. Routledge.
- Sinclair, J. (éd.) (1987). *Looking up. An account of the COBUILD project in lexical computing*. Collins Cobuild.
- Sinclair, J. (éd.) (1991). *Corpus, concordance, collocation*. Oxford University Press.

- Soutet, O., Sfar, I. et Mejri S. (éds.) (2018). *La phraséologie contrastive*. Honoré Champion.
- Stefanowitsch, A. (2014). Collostructional analysis. A case study of the English into-causative. Dans T. Herbst, H.-J. Schmid, et S. Faulhaber (éds.), *Constructions – Collocations – Patterns* (p. 217–238). Mouton de Gruyter.
- Stefanowitsch, A. et Gries, S. (2003). Collostructions: investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8, 209–243.
- Steinlin, J., Kahane, S. et Polguère, A. (2005). Compiling a “classical” explanatory combinatorial lexicographic description into a relational database. Dans D. Apresjan et L. L. Iomdin, *East West encounter. Proceedings of the Second International Conference on the Meaning-Text Theory (MTT’2005)* (p. 477–485). Slavic Culture Languages Publishing House.
- Steinlin, J., Kahane, S., Polguère, A. et El Ghali, A. (2004). De l’article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux. Dans G. Williams et S. Vessier (éds.), *Proceedings of the Eleventh EURALEX International Congress. EURALEX 2004* (p. 177-186). Université de Bretagne Sud.
- Strauss, A. et Corbin, J. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Sage.
- Subirats Rüggeberg and Petruck, M. (2003). Surprise: Spanish FrameNet! Dans E. Hajičová, A. Kotěšovcová et J. Mirovský (éds.), *Proceedings of the workshop on frame semantics, XVII International Congress of Linguists (CIL)*. http://elies.rediris.es/miscelanea/misc_6/Subirats-Petruck.pdf
- Subirats Rüggeberg, C. (2013a). Frames, constructions, and metaphors in Spanish FrameNet. Dans I. Verdager, N. J., Laso et D. Salazar (éds.), *Biomedical English. A corpus-based approach* (p. 185-210). John Benjamins Publishing Company.
- Subirats Rüggeberg, C. (2013b). La integración de la semántica de marcos y la semántica de simulación: aplicaciones al procesamiento semántico automático del español. Dans M. L. Calero et M. A. Hermosilla (éds.), *Lingüística, poética y*

- cognición* (p. 307-337.). Servicio de Publicaciones de la Universidad de Córdoba.
- Subirats Rüggeberg, C. (2001). *Introducción a la sintaxis léxica del español*. Iberoamericana.
- Sułkowska, M. (2013). *De la phraséologie à la phraséodidactique. Études théoriques et pratique*. Wydawnictwo UŚ.
- Sułkowska, M. (2016). Phraséodidactique et phraséotraduction : quelques remarques sur les nouvelles disciplines de la phraséologie appliquée. *Yearbook of Phraseology*, 7, 35-54.
- Svensson, M. H. (2004). *Critères de figement. L'identification des expressions figées en français contemporain* [thèse de doctorat, Umeå Universitet]. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A143138&dswid=9637>
- Svensson, M. H. (2008). A very complex criterion of fixedness: Non-compositionality. Dans S. Granger et F. Meunier (éds.), *Phraseology: an interdisciplinary perspective* (p. 81-93). John Benjamins Publishing Company.
- Sweet, H. (1891/1960). *A new English grammar, logical and historical*. 2 vols. The Oxford University Press.
- Sypnicki, J. (1991). Quelques remarques sur les modifications comiques des structures lexicales et phraséologiques. *Studi italiani di lingüística teorica ed applicata*, 1, 151-163.
- Talmy L. (1983). How language structures space. Dans H. Pick et L. Acredelo (éds.), *Spatial orientation: theory, research, and application* (p. 225–282). Plenum.
- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive science*, 12, 49-100.
- Tamba, I. (2014). Sémantique lexicale et phrastique : le sens compositionnel revisité. *Cahiers de lexicologie*, 105(2), 99-106.
- Telija, V. (1996). *Russkaja fraseologuija. Semantičeskij, pragmatičeskij i lingvokul'turulogičeskij aspektj*. Shkola "Yazyki russkoj kultury".
- Temmermann, R. (2000). *Towards new ways of terminological description. The sociocognitive approach*. John Benjamins Publishing Company.

- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Klincksieck.
- Thun, H. (1975). Quelques relations systématiques entre groupements de mots figés. *Cahiers de lexicologie*, 27, 52-71.
- Thun, H. (1978). *Probleme der phraseologie. Untersuchungen zur wiederholten Rede mit Beispielen aus den Französischen, Italienischen, Spanischen und Rumänischen*. Max Niemeyer Verlag.
- Timofeeva, L. (2008). *Acerca de los aspectos traductológicos de la fraseología española* [thèse de doctorat, Universidad de Alicante]. https://rua.ua.es/dspace/bitstream/10045/7707/1/tesis_doctoral_larissa_timofeeva.pdf
- Titone, D. A. et Connine, C. M. (1994). Descriptive norms for 171 idiomatic expressions: familiarity, compositionality, predictability, and literality, *Metaphor & Symbolic Activity*, 9, 247-270.
- TLFi : Trésor de la Langue Française informatisé*. ATILF - CNRS & Université de Lorraine. <http://www.atilf.fr/tlfi>
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins Publishing Company.
- Tomasello, M. (2002). *New psychology of language*. Lawrence Erlbaum Associates Inc. Publishers.
- Tono, Y. (2003). Learner corpora: design, development and applications. Dans D. Archer, P. Rayson, A. Wilson et T. McEnery (éds.) *Proceedings of the corpus linguistics 2003 conference* (p. 800-809). Lancaster University, University Centre for Computer Corpus Research on Language.
- Torrent, T., Borin, L. et Baker, C. (éds.). (2018). *LREC 2018 Workshop. International FrameNet workshop 2018. Multilingual Framenets and Constructions. Proceedings*. http://lrec-conf.org/workshops/lrec2018/W5/pdf/book_of_proceedings.pdf
- Tristá Pérez, M. A. (1980). Estructura interna de las unidades fraseológicas. *Anuario L/L*, 10, 93-104.
- Tristá Pérez, M. A. (1988). *Fraseología y contexto*. Editorial de Ciencias Sociales.

- Tufis, D., Cristea, D. et Stamou, S. (2004). BalkaNet: aims, methods, results and perspectives. A general overview. *Romanian Journal of Information science and technology*, 7(1-2), 9-43.
- Tutin, A. (2007). Collocations du lexique transdisciplinaire des écrits scientifiques : annotation et extraction des propriétés linguistiques dans la perspective d'une application didactique. *Cahiers de l'Institut de linguistique de Louvain*, 31(2-4), 247-262.
- Tutin, A. (2005). Le dictionnaire de collocations est-il indispensable ? *Revue française de linguistique appliquée*, 10(2), 31-48.
- Tutin, A. (2008a). For an extended definition of lexical collocations. Dans E. Bernal et J. DeCesaris (éds), *Proceedings of the 12th EURALEX International Congress* (p. 1453-1460). IULA.
- Tutin, A. (2008b). L'apport des corpus annotés pour l'élaboration semi-automatique d'une base de collocations de la langue scientifique générale. Dans F. Maniez, P. Dury, N. Arlin et C. Rougemont (dirs.), *Corpus et dictionnaires de langues de spécialité* (p. 45-65). Presses universitaires de Grenoble.
- Tutin, A. (2013). Les collocations lexicales : une relation essentiellement binaire définie par la relation prédicat-argument. *Langages*, 189, 47-63.
- Tutin, A. (2020). *Tu parles ! Et puis quoi encore !* Phrases préfabriquées à fonction expressive dans les dictionnaires français. Dans F. Neveu, B. Harmegnies, L. Hriba, S. Prévost et A. Steuckardt (éds.), *SHS Web Conferences (vol. 78). 7^e Congrès Mondial de Linguistique Française (CMLF 2020)*. https://www.shs-conferences.org/articles/shsconf/pdf/2020/06/shsconf_cmlf2020_05013.pdf
<https://doi.org/10.1051/shsconf/20207805013>
- Tutin, A. et Grossmann, F. (2002). Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue française de linguistique appliquée*, 7(1), 7-25.
- Uzcanga Vivar, I. (2017). La néologie dans le domaine de la phraséologie en espagnol et en français. Dans, M. Sánchez Ibáñez, N. Maroto, J. Torres del Rey, G. De Sterck, D. Linder et J. Gacía Palacios (éds.), *La renovación léxica en las lenguas*

románicas: proyectos y perspectivas (p. 209-228). Ediciones de la Universidad de Murcia.

Uzcanga Vivar, I. (2020a). La phraséologie, un domaine favorable à l'innovation lexicale. Dans, V. Arigne, S. Pech-Pelletier, Ch. Rocq-Migette et J. F. Sablayrolles (dirs.), *Études lexicales. Mélanges offerts à Ariane Desporte* (p. 129-146). Université Sorbonne Paris Nord. <https://hal-univ-paris13.archives-ouvertes.fr/hal-02971758/document>.

Uzcanga Vivar, I. (2020b). Metáfora y factores socioculturales en el ámbito de la fraseología en francés y español. Dans P. Cotta Ramusino et F. Mollica (éds.), *Phraseology : languages and cultures in comparison* (p. 271-281). Cambridge Scholars Publishing.

Uzcanga Vivar, I. et Gómez Fernández, A. (2011). Intercompréhension et inférences lexicales : le cas des langues romanes. *L'intercomprensione in ricordo di Claire Blanche-Benveniste. REDINTER-Intercomprensão* (vol. 3, p. 95-112). Edições Cosmos.

van den Eynde, K. et Blanche-Benveniste, C. (1978). Syntaxe et mécanismes descriptifs : présentation de l'approche pronominale. *Cahiers de lexicologie*, 32(1), 3-27.

van den Eynde, K. et Mertens, P. (2003). La valence : l'approche pronominale et son application au lexique verbal. *Journal of French Language Studies* 13(1), 63-104.

van den Eynde, K. et Mertens, P. (2006). *Le dictionnaire de valence Dicovalence: manuel d'utilisation*. <http://bach.arts.kuleuven.be/dicovalence/manuel061117.pdf>.

van den Eynde, K., Mertens, P. et Eggermont, C. (2017). *Dicovalence [Lexique]*. <https://hdl.handle.net/11403/dicovalence/v1>.

van der Plas, L. et Apidianaki, M. (2014). Cross-lingual word sense disambiguation for predicate labelling of french. *Proceedings of the 21st TALN Conference* (p. 46-55). Association pour le Traitement Automatique des Langues. <http://www.aclweb.org/anthology/F14-1005>

- Van der Plas, L., Samardžić, T. et Merlo, P. (2010). Cross-lingual validity of PropBank in the manual annotation of French. Dans *Fourth Linguistic Annotation Workshop (The LAW IV). Proceedings of the workshop* (p.113-117). The Association for Computational Linguistics. <https://aclanthology.org/W10-18.pdf>
- Varantola, K. (1986). Translators, dictionaries and text corpora. Dans S. Bernardini et F. Zanettin (éds.), *I corpora nella didattica della traduzione* (p. 117-133). CLUEB.
- Varela, F. et Kubarth, H. (1994). *Diccionario fraseológico del español moderno*. Gredos
- Velasco Menéndez, J. (2010). La aportación de V.V. Vinogradov al desarrollo de la fraseología rusa. *Eslavística Complutense*, 10, 125-134.
- Vendler, Z. (1967). *Linguistics in Philosophy*. Cornell University Press.
- Veniard, M. (2012). Questionnements croisés sur le sens et la phraséologie : lexique et axe syntagmatique dans la perspective de l'analyse du discours. *Pratiques*, 155-156, 66-82. DOI : 10.4000/pratiques.3465
- Verlinde, S., Selva, T. et Binon, J. (2005). L'informatique au service de la lexicographie. Dans H. Béjoint et F. Maniez (éds.), *De la mesure dans les termes* (p. 109-123). Presses Universitaires de Lyon.
- Véronis, J. (2004). Quels dictionnaires pour l'étiquetage sémantique ? *Le français moderne*, 72(1), 27-38.
- Verstraten, L. (1992). Fixed phrases in monolingual learners' dictionaries. Dans P. Arnaud et H. Béjoint (éds.), *Vocabulary and applied linguistics* (p. 28-40). MacMillan.
- Victorri, B. et Fuchs, C. (1996). *La polysémie. Construction dynamique du sens*. Hermès.
- Vinogradov, V. (1946/1977). Osnovnije ponjatija ruskij fraseologii, kak lingvističeskoj distsipliny. Dans V. Vinogradov (1977), *Izbrannyje trudy. Leksikologuija y leksikografija*. Nauka.
- Vinogradov, V. (1947/1977). Ob osnovnykh tipakh fraseologičeskikh edinit v russkom jazyke. Dans V. Vinogradov (1977), *Izbrannyje trudy. Leksikologuija y leksikografija*. Nauka.

- Vinogradov, V. (1953/1977). Osnovnyje tipy leksicheskikh znachenij slov. Dans V. Vinogradov (1977), *Izbrannyje trudy. Leksikologuija y leksikografia*. Nauka.
- Vinogradov, V. (1977). *Izbrannyje trudy. Leksikologuija y leksikografia*. Nauka.
- Vivès, R. (1983). *Avoir, prendre, perdre : constructions à verbe support et extensions aspectuelles* [thèse de 3^e cycle, Université Paris 8].
- Vivès, R. (1993). La prédication nominale et l'analyse par verbes supports. *L'Information Grammaticale*, 59, 8-15.
- Vossen, P. (éd.). (1998). *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer Academic Publishers.
- Wanner, L. (dir.). (1996). *Lexical Functions in Lexicography and Natural Language Processing*. John Benjamins Publishing Company.
- Wanner, L., Alonso Ramos, M., Vincze, O., Nazar, R. Ferraro, G., Mosqueira, E. et Prieto, S. (2013). Annotation of collocations in a learner corpus for building a learning environment. Dans S. Granger (éd.), *Twenty years of learner corpus research. Looking back, moving ahead* (p. 493-503). Presses universitaires de Louvain.
- Watts, D. J. et Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440–442.
- Weinreich, U. (1966/1969). Problems in the analysis of idioms. Dans J. Puhvel (éd.), *Substance and structure of language* (p. 23-8). University of California Press.
- Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Taylor, A., Kaufman, J., Franchini, M., El-Bachouti, M., Belvin, R. et Houston, A. (2013). *OntoNotes Release 5.0*. <https://catalog.ldc.upenn.edu/LDC2013T19>
- Wierzbicka, A. (1982). Why can you *have a drink* when you can't **have an eat*? *Laguage*, 58, 753-799.
- Wierzbicka, A. (2009). The Theory of the Mental Lexicon. Dans S. Kempgen, P. Kosta, T. Berger et K. Gutschmidt (éds), *Die slavischen sprachen/The Slavic Languages. Ein internationales handbuch zu ihrer struktur, ihrer geschichte und ihrer erforschung / An international handbook of their structure, their history and their investigation* (p. 848–863). Mouton de Gruyter.

- Williams, E. (1981). Argument structure and morphology. *The Linguistic Review*, 1(1), 81-114.
- Williams, E. (1987). English as an ergative language: the theta-structure of derived nouns. *Proceedings of the Chicago Linguistic Society*, 23, 366-375.
- Williams, G. (2001). Sur les caractéristiques de la collocation. Dans d. Maurel, *Actes de la 8^e conférence annuelle sur le Traitement Automatique des Langues Naturelles. Tours, 2-5 juillet, 2001* (p. 9-16). ATALA.
- Williams, G. (2003). Les collocations et l'école contextualiste britannique. Dans F. Grossmann et A. Tutin (éds.), *Les collocations : analyse et traitement* (p. 33-44). Éditions De Werelt.
- Winter, C. (1992). Verbal stereotypes. *Meta : Journal des traducteurs/Meta: Translators' Journal*, 35(2), 378-389.
- Wotjak, B. (1989). Ansatz eines modular-intergrativen Beschreibungs-modells für verbale Phraseolexeme (PL). Dans G. Gréciano (éd.), *Europhras 88 : phraséologie contrastive. Actes du Colloque International Klingenthal-Strasbourg, 12-16 mai 1988* (p. 459-467). Université des Sciences Humaines.
- Wotjak, B. (1992). *Verbale phraseolexeme in system und text*. Max Niemeyer.
- Wotjak, G. (1983). En torno a la traducción de unidades fraseológicas (con ejemplos tomados del español y el alemán). *Linguistische Arbeitsberichte*, 40, 56-80.
- Wotjak, G. (1984). No hay que estarse con los brazos cruzados. Algunas observaciones acerca del significado de expresiones idiomáticas verbales del español actual. *Linguistische Arbeitsberichte*, 45, 75-84.
- Wotjak, G. (1988). Uso y abuso de las unidades fraseológicas. Dans *Homenaje a Alonso Zamora Vicente* (vol. 1, p. 535-548). Castalia.
- Wotjak, G. (1992). Estructuras en el léxico. Dans G. Wotjak, *Estudios de lexicología y metalexigrafía del español* (p. 108-124). Max Niemeyer Verlag.
- Wotjak, G. (éd.) (1998). *Estudios de fraseología y fraseografía del español actual*. Vervuert/Iberoamericana.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

- Wray, A. (2008). *Formulaic language: pushing the boundaries*. Cambridge University Press.
- Wray, A. et Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, 20, 1-28.
- Wright, S. E. et Budin, G. (éds.) (1997). *Handbook of terminology management*. John Benjamins Publishing Company.
- Wüster, E. (1996). *Terminología: selecció de textos d'E. Wüster*. Universitat de Barcelona, Servei de Llengua Catalana.
- Wüster, E. (1998). *Introducción a la teoría general de la terminología y a la lexicografía terminológica*. Institut Universitari de Lingüística aplicada, Universitat Pompeu Fabra.
- Yan, R., Tutin, A. et Tran, T. T. H. (2018). Routines verbales pour les français langue étrangère : des corpus d'experts aux corpus d'apprenants. *Lidil. Revue de linguistique et de didactique des langues*, 58. <https://journals.openedition.org/lidil/5411>
- Yllera, A. (2001). Linguistique contrastive, linguistique comparée ou linguistique tout court ? Dans I. Uzcanga Vivar, E. Llamas Pombo et J. M. Pérez Velasco (éds.), *Presencia y renovación de la lingüística francesa* (p. 435-446). Ediciones Universidad de Salamanca.
- Zinglé, H. et Brobeck-Zinglé, M.-L. (2003). *Dictionnaire combinatoire du français. Expressions, locutions et constructions*. La maison du dictionnaire.
- Žolkovskij, A. et Mel'čuk, I. (1970). Sur la synthèse sémantique. *T.A. Informations*, 2, 1-85. (Traduit du russe : 1967, O semantičeskom sinteze. *Problemy kibernetiki*, 19, 177-238)
- Zuluaga, A. (1975). La fijación fraseológica. *Thesaurus XXX* (2), 225-248.
- Zuluaga, A. (1980). *Introducción al estudio de las expresiones fijas*. Peter Lang.
- Zuluaga, A. (2002). Los "enlaces frecuentes" de María Moliner. Observaciones sobre las llamadas colocaciones. *LEA*. 24(1), 97-114.

OUTILS LEXICOGRAPHIQUES CONSULTÉS

- aa. ii, (2001). *Lexique des termes juridiques*, sous la direction de S. Guinchard et G. Montagnier, Éditions Dalloz.
- Barilari, A. et Drapé, R.. (1987). *Lexique fiscal*. Éditions Dalloz.
- Beau, P. et Beau, R. (1986). *Lexique de la protection sociale*, Éditions Dalloz.
- Bernard, Y., Colli, J.-C. et Lewandowski, D. (1975). *Dictionnaire économique et financier*, Éditions du Seuil.
- Cornu, G. (1987). *Vocabulaire juridique*, Presses Universitaires de France.
- Commission ministérielle de terminologie auprès du Secrétaire d'État chargé des personnes âgées, (1984). *Dictionnaire des personnes âgées, de la retraite et du vieillissement*, sous la direction de J.-C. Sournia, Franterm diffusé par Nathan.
- Frangi, M. et Schulz, P. (1995). *Droit des relations internationales*, Éditions Dalloz
- Lemeunier, (1988). *Dictionnaire juridique*, avec la collaboration du Centre International du Droit des Affaires, La maison du dictionnaire.
- Magsof, G. (1994). *Le langage des impôts*, Éditions Méthodes et Stratégies.
- Percebois, J., Aben, J. et Euzéby, A. (1995). *Dictionnaire de finances publiques*, Armand Colin Éditeur.
- Sournia, J.-C. (1973). *Dictionnaire des assurances sociales*, Maison et Cie Éditeurs.
- Sainrapt, C. (1996). *Dictionnaire général de l'assurance*, Éditions Arcature.
- Salomon, J. (2001). *Dictionnaire de droit international public*, Etablissements Émile Bruylant.
- Sousi-Roubi, B. (2001). *Lexique de la banque et des marchés financiers*, Éditions Dalloz.

ANNEXE I

Nom	Catégorie grammaticale	Classe d'objet	N0	N1	N2	N3	Domaine	Acronyme	espagnol
abattement	nm	aide financière : dégrèvement	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhum, Ncoll, #pour Nhum, Ncoll, #en faveur de Nhum, Ncoll	droit fiscal		deducción
abattement	nm	aide financière : dégrèvement	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhum, Ncoll, #pour Nhum,	commerce		deducción

						Ncol l, #en fave ur de Nhu m, Ncol l			
abattement fiscal	nm	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l, #po ur Nhu m, Ncol l, #en fave ur de Nhu m, Ncol l	droit fiscal		deducción fiscal

aide à l'embauche dans les TPE	nf	aide financière : allocation	coll_hum <Etat>	à <coll_hum:entreprise:TP E>	de <montant>	pour Vinf	droit du travail		ayuda al empleo para PYMES
aide au logement	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		ayuda al alquiler
aide au mérite	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca al mérito académico
aide financière	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	général		ayuda económica
aide judiciaire	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	justice		ayuda jurídica
aide médicale à domicile	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale / santé		ayuda médica a domicilio
aide personnalisée au logement	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	logement / protection sociale	APL	ayuda al alquiler
aide sociale	nf	aide financière	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		ayuda social

		: allocation							
aide sociale à domicile	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	protection sociale		ayuda social a domicilio
allégement fiscal	nm	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l, #po ur Nhu m, Ncol l, #en fave ur de Nhu m, Ncol l	droit fiscal		desgravación fiscal
allocation chômeurs âgés	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	protection sociale	ACA	subsidio por desempleo para trabajadores mayres de 52 años / subsidio por desempleo para trabajadores mayres de 55 años

allocation compensatrice pour tierce personne	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	ACTP	subsidio de ayuda a tercera persona
allocation compensatrice tierce personne	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	ACTP	subsidio de ayuda a tercera persona
allocation compensatrice tierce personne	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	ACTP	subsidio de ayuda a tercera persona
allocation d'adoption	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	AAD	Prestación económica por nacimiento o adopción
allocation de chômage	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit du travail		paro / prestación contributiva / prestación contributiva por desempleo / prestación por desempleo / renta activa de inserción / subsidio de desempleo / subsidio de desempleo
allocation de logement	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	logement / protection sociale		ayuda al alquiler
allocation de	nf	aide financière	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	logement /	ALF	ayuda al alquiler

logement familial		: allocation					protection sociale		
allocation de maternité	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit du travail		prestación económica por maternidad
allocation de présence parentale	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		prestación económica por cuidado de menores afectados por cáncer u otra enfermedad grave
allocation de recherche	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca de investigación
allocation de revenu minimum d'insertion	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		paro / prestación contributiva / prestación contributiva por desempleo / prestación por desempleo / renta activa de inserción / subsidio de desempleo / subsidio de desempleo
allocation de veuvage	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit du travail		pensión de viudedad
allocation d'éducation de l'enfant handicapé	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	AEEH	ayuda complementaria para alumnado con necesidades específicas de apoyo educativo

allocation d'éducation spéciale	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	AES	ayuda complementaria para alumnado con necesidades específicas de apoyo educativo
allocation d'éducation spécialisée	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	AES	ayuda complementaria para alumnado con necesidades específicas de apoyo educativo
allocation d'études	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca
allocation familiale	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		prestación familiar
allocation logement à caractère familial	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	logement / protection sociale	ALF	ayuda al alquiler
allocation pour assistance d'une tierce personne	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	handicapé / santé		subsídio de ayuda a tercera persona
allocation supplémentaire d'invalidité	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	ASI	pensión no contributiva de invalidez

allocation veuvage	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión de viudedad
allocations de solidarité	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión de viudedad
allocs	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		subsidio
appui	nm	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	général		apoyo
bourse	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca
bourse de mobilité	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca de movilidad
bourse de thèse	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca de doctorado

bourse d'enseignement supérieur	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		becas para cursar estudios universitarios de Grado
bourse d'études	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca
bourse du mérite	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca al mérito académico
bourse Erasmus	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca Erasmus
bourse Erasmus Mundus	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca Erasmus Mundus
bourse postdoctorale	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca postdoctoral
bourse postdoc	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		beca postdoctoral

décompte	nm	aide financière : dégrèvement	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhum, Ncoll, #pour Nhum, Ncoll, #en faveur de Nhum, Ncoll	général		descuento / rebaja
déduction	nf	aide financière : dégrèvement	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhum, Ncoll, #pour Nhum, Ncoll, #en	général		deducción

						fave ur de Nhu m, Ncol l			
déduction fiscale	nf	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l, #po ur Nhu m, Ncol l, #en fave ur de Nhu m, Ncol l	droit fiscal		deducción fiscal
dégrève ment	nm	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l,	droit fiscal		desgravación

						#pour Nhum, Ncoll, #en faveur de Nhum, Ncoll			
dégrèvement fiscal	nm	aide financière : dégrèvement	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhum, Ncoll, #pour Nhum, Ncoll, #en faveur de Nhum,	droit fiscal		desgravación fiscal

						Ncol l			
escompte	nm	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l, #po ur Nhu m, Ncol l, #en fave ur de Nhu m, Ncol l	commerce		descuento / rebaja
indemnisation	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	droit du travail / protection sociale		indemnización

indemnisation	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit du travail / justice / protection sociale		indemnización
indemnisation chômage	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit du travail / protection sociale		subsídio de desempleo / prestación de desempleo
indemnité	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit du travail / protection sociale		indemnización
indemnité	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit du travail / justice / protection sociale		indemnización
indemnité contractuelle de licenciement	nf	aide financière : allocation	coll_hum <entreprise>	à <hum:travailleur:salarié: licencié>	de <montant>	pour Vinf	droit du travail		indemnización por despido
indemnité de guerre	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	politique internationale		indemnización de guerra
indemnité de licenciement	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit du travail		indemnización por despido

indemnité de licenciement	nf	aide financière : allocation	coll_hum <Etat>	à <hum:travailleur:salarié: agent technique du Ministère de la Défense>	de <montant>	pour Vinf	droit du travail		indemnización por despido
indemnité journalière	nf	aide financière : allocation	coll_hum <Etat, Pôle emploi>	à <hum:travailleur:salarié: malade, parent>	de <montant>	pour Vinf	droit du travail	IJ	subsidio / prestación
indemnité légale de licenciement	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit du travail		indemnización por despido procedente
indemnité parlementaire	nf	aide financière : allocation	coll_hum <Etat>	à <hum:travailleur:parlementaire>	de <montant>	pour Vinf	droit du travail		retribución ed los miembros de la Diputación Permanente
indemnités Assedic	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		subsidio de desempleo / prestación de desempleo
pension	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión
pension	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	enseignement		pensión

pension alimentaire	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	droit civil		pensión alimenticia
pension de guerre	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	armée / protection sociale		pensión derivada de la guerra civil
pension de retraite	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión de jubilación
pension de reversion	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión de viudedad
pension de reversion d'orphélin	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión de orfandad
pension de reversion du conjoint survivant	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión de viudedad
pension de veuf	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión de viudedad
pension de veuve	nf	aide financière	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión de viudedad

		: allocation							
pension de veuve de guerre	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión familiar derivada de la guerra civil
pension d'invalidité	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	armée / protection sociale		prestación económica por incapacidad permanente / prestación económica por incapacidad temporal / prestación económica por lesiones permanentes no invalidantes
pension militaire de retraite	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	armée / protection sociale		pensión de retiro
prestation	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		prestación
prestation compensatoire	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		indemnización
prestation d'accueil du jeune enfant	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	PAJE	prestación por nacimiento y cuidado de menor

prestation d'accueil du jeune enfant	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	PAJE	prestación por nacimiento y cuidado de menor
prestation de maladie	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		prestación por incapacidad temporal
prestation d'invalidité	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		prestación económica por incapacidad permanente / prestación económica por incapacidad temporal / prestación económica por lesiones permanentes no invalidantes
prestation sociale	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		prestación social
prestation spécifique de dépendance	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale	PSD	ayuda para personas dependientes
prestation vieillesse	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		pensión de jubilación
prime	nf	aide financière	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		prima

		: allocation							
prime	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	général		prima
prime	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	économie		prima
prime à la naissance ou à l'adoption	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	protection sociale		prestación económica por nacimiento o adopción de hijo, en supuestos de familias numerosas, monoparentales y en los casos de madres discapacitadas
prime de naissance ou d'adoption	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	protection sociale		Prestación económica por nacimiento o adopción
escompte de caisse	nf	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l, #po ur Nhu m, Ncol	commerce		descuento por pronto pago

						l, #en fave ur de Nhu m, Ncol l			
exonératio n	nf	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l, #po ur Nhu m, Ncol l, #en fave ur de Nhu m, Ncol l	droit fiscal		exoneración

exonération d'impôt	nf	aide financière : dégrèvement	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhum, Ncoll, #pour Nhum, Ncoll, #en faveur de Nhum, Ncoll	droit fiscal		exención tributaria
exonération fiscale	nf	aide financière : dégrèvement	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhum, Ncoll, #pour Nhum, Ncoll, #en	droit fiscal		exención fiscal

						fave ur de Nhu m, Ncol l			
rabais	nm	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l, #po ur Nhu m, Ncol l, #en fave ur de Nhu m, Ncol l	général		descuento / rebaja
réduction	nf	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l,	général		reducción

						#po ur Nhu m, Ncol l, #en fave ur de Nhu m, Ncol l			
remise	nf	aide financière : dégrève ment	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhu m, Ncol l, #po ur Nhu m, Ncol l, #en fave ur de Nhu m,	commerce		descuento / rebaja

						Ncol l			
rente	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	assurance s / protection sociale		renta
rente	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	assurance s / protection sociale		renta
rente d'accident du travail	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	protection sociale		pensión por incapacidad
rente de conjoint	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	assurance s		pensión de viudedad
rente de conjoint survivant	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	assurance s		pensión de viudedad
rente d'incapacit é permanent e	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <monta nt>	pour Vinf	protection sociale		pensión de incapacidad permanente

ristourne	nf	aide financière : dégrèvement	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhum, Ncoll, #pour Nhum, Ncoll, #en faveur de Nhum, Ncoll	commerce		descuento / rebaja
subside	nm	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	protection sociale		subsídio
subvention	nf	aide financière : allocation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	culture / politique économique		subvención
don	nm	aide financière : donation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	général		donativo

donation	nf	aide financière : donation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	général		donación
aide financière	nf	aide financière : dégrèvement	Ncoll	de <montant>, #de N	sur N, #pour N	à Nhum, Ncoll, #pour Nhum, Ncoll, #en faveur de Nhum, Ncoll	général		ayuda económica
aide financière	nf	aide financière : donation	Nhum, Ncoll	à Nhum, Ncoll	de <montant>	pour Vinf	général		ayuda económica

Nom	Catégorie grammaticale	Classe d'objet	N0	N1	N2	N3	Domaine	Acronyme	français
ayuda económica	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	général		aide financière
deducción	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	droit fiscal		abattement
deducción	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	commerce		abattement
apoyo	nm	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant >	para Vinf	général		appui
ayuda al alquiler	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant >	para Vinf	logement / protection sociale		aide au logement / allocation de logement / aide personnalisée au logement / allocation de logement

									familial / allocation logement à caractère familial
ayuda al empleo para PYMES	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	droit du travail		aide à l'embauche dans les TPE
ayuda complementaria para alumnado con necesidades específicas de apoyo educativo	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	protection sociale		allocation d'éducation de l'enfant handicapé / allocation d'éducation spéciale / allocation d'éducation spécialisée
deducción	nf	aide financière: dégrève ment	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	général		déduction
ayuda económica	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	général		aide financière

ayuda jurídica	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	justice		aide judiciaire
ayuda médica a domicilio	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale / santé		aide médicale à domicile
ayuda para personas dependientes	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prestation spécifique de dépendance
ayuda social	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		aide sociale
ayuda social a domicilio	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		aide sociale à domicile
beca	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseignement		bourse / bourse d'études / allocation d'études
beca al mérito académico	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseignement		aide au mérite / bourse du mérite

beca de doctorado	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseñemen t		bourse de thèse
beca de investigación	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseñemen t		allocation de recherche
beca de movilidad	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseñemen t		bourse de mobilité
beca Erasmus	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseñemen t		bourse Erasmus
beca Erasmus Mundus	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseñemen t		bourse Erasmus Mundus
beca postdoctoral	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseñemen t		bourse postdoctorale / bourse postdoc
becas para cursar estudios universitarios de Grado	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseñemen t		bourse d'enseñeme nt supérieur

indemnización	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	droit du travail / justice / protection sociale		indemnisation / indemnité / prestation compensatoire
indemnización de guerra	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	politique internationa le		indemnité de guerre
indemnización por despido	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	droit du travail		indemnité contractuelle de licenciement / indemnité de licenciement / indemnité de licenciement
indemnización por despido procedente	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	droit du travail		indemnité légale de licenciement
pensión	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	protection sociale		pension

pensión	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	enseignement		pension
pensión alimenticia	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit civil		pension alimentaire
pensión de incapacidad permanente	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		rente d'incapacité permanente
pensión de jubilación	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		pension de retraite
pensión de jubilación	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prestation vieillesse
pensión de orfandad	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		pension de reversion d'orphélin
pensión de retiro	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	armée / protection sociale		pension militaire de retraite

pensión de viudedad	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit du travail		allocation de veuvage / rente de conjoint / rente de conjoint survivant / allocation veuvage / allocations de solidarité / pension de reversion / pension de veuf / pension de veuve / pension de reversion du conjoint survivant
deducción fiscal	nf	aide financière: dégrèvement	Ncoll	de <montant>, #de N	en N, #por N, #para N	a Nhum, Ncoll	droit fiscal		abattement fiscal
deducción fiscal	nf	aide financière: dégrèvement	Ncoll	de <montant>, #de N	en N, #por N, #para N	a Nhum, Ncoll	droit fiscal		déduction fiscale

descuento	nm	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	général		décompte / rabais
descuento	nm	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	commerce		escompte / remise
rebaja	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	commerce		escompte / remise
rebaja	nm	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	général		rabais / décompte
descuento / rebaja	nm	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	commerce		ristourne
pensión derivada de la guerra civil	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant >	para Vinf	armée / protection sociale		pension de guerre
pensión familiar derivada de la guerra civil	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant >	para Vinf	protection sociale		pension de veuve de guerre

pensión no contributiva de invalidez	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		allocation supplémentaire d'invalidité
pensión por incapacidad	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		rente d'accident du travail
prestación	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit du travail		indemnité journalière
prestación	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prestation
prestación contributiva	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit du travail		allocation de chômage / allocation de revenu minimum d'insertion
prestación contributiva por desempleo	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit du travail		allocation de chômage / allocation de revenu minimum d'insertion

prestación de desempleo	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	droit du travail / protection sociale		indemnisation chômage / indemnités Assedic
prestación económica por cuidado de menores afectados por cáncer u otra enfermedad grave	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	protection sociale		allocation de présence parentale
prestación económica por incapacidad permanente	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	armée / protection sociale		pension d'invalidité
prestación económica por incapacidad permanente	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	protection sociale		prestation d'invalidité
prestación económica por incapacidad temporal	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	armée / protection sociale		pension d'invalidité
prestación económica por lesiones permanentes no invalidantes	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	protection sociale		prestation d'invalidité
prestación económica por lesiones permanentes no invalidantes	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	armée / protection sociale		pension d'invalidité

prestación económica por maternidad	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit du travail		allocation de maternité
Prestación económica por nacimiento o adopción	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		allocation d'adoption
prestación económica por nacimiento o adopción de hijo, en supuestos de familias numerosas, monoparentales y en los casos de madres discapacitadas	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prime à la naissance ou à l'adoption
prestación económica por nacimiento o adopción	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prime de naissance ou d'adoption
prestación familiar	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		allocation familiale
prestación por desempleo	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit du travail		allocation de chômage / allocation de revenu minimum d'insertion

prestación por incapacidad temporal	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prestation de maladie
prestación por nacimiento y cuidado de menor	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prestation d'accueil du jeune enfant
prestación por nacimiento y cuidado de menor	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prestation d'accueil du jeune enfant
prestación social	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prestation sociale
prima	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		prime
prima	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	général		prime
prima	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	économie		prime

renta	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	assurances / protection sociale		rente
renta activa de inserción	nm	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit du travail		allocation de chômage / allocation de revenu minimum d'insertion
retribución de los miembros de la Diputación Permanente	nf	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit du travail		indemnité parlementaire
subsidio	nm	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		allocs / subside
subsidio	nm	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	droit du travail		indemnité journalière
subsidio de ayuda a tercera persona	nm	aide financière: allocation	Nhum, Ncoll	a Nhum, Ncoll	de <montant>	para Vinf	protection sociale		allocation compensatrice pour tierce personne / allocation compensatrice tierce personne /

									allocation pour assistance d'une tierce personne
descuento por pronto pago	nm	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	commerce		escompte de caisse
desgravación	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	droit fiscal		dégrèvement
desgravación fiscal	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	droit fiscal		allègement fiscal / dégrèvement fiscal
exención fiscal	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	droit fiscal		exonération fiscale
exención tributaria	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	droit fiscal		exonération d'impôt
exoneración	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	droit fiscal		exonération

subsidio de desempleo	nm	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	droit du travail		allocation de chômage / allocation de revenu minimum d'insertion / indemnisation chômage / indemnités Assedic
subsidio por desempleo para trabajadores mayores de 52 años	nm	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	protection sociale		allocation chômeurs âgés
subsidio por desempleo para trabajadores mayores de 55 años	nm	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	protection sociale		allocation chômeurs âgés
subvención	nf	aide financière: allocation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	culture / politique économique		subvention
donativo	nm	aide financière: donation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	général		don
donación	nf	aide financière: donation	Nhu m, Ncoll	a Nhum, Ncoll	de <montan t>	para Vinf	général		donation

reducción	nf	aide financière: dégrèvement	Ncoll	de <montant >, #de N	en N, #por N, #para N	a Nhum, Ncoll	général		réduction
ayuda económica	nf	aide financière: donation	Nhum, Ncoll	a Nhum, Ncoll	de <montant >	para Vinf	général		aide financière

UNIVERSIDAD DE SALAMANCA
FACULTAD DE FILOLOGÍA
DEPARTAMENTO DE FILOLOGÍA FRANCESA



VNiVERSiDAD
DSALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

DOCTORADO EN LENGUAS MODERNAS

RESUMEN EN CASTELLANO DE TESIS DOCTORAL

**« Dictionnaire bilingue français-espagnol des noms
prédicatifs des <Aides financières>/<Ayudas
financieras>. Étude syntactico-sémantique »**

A handwritten signature in blue ink, likely belonging to the author or a member of the thesis committee.

**ELENA DIEGO
HERNÁNDEZ**
(autora)

**VºBº
ISABEL UZCANGA
VIVAR**
(directora)

**VºBº
ARACELI GÓMEZ
FERNÁNDEZ**
(codirectora)

RESUMEN DE TESIS DE DOCTORADO

Estudiante de doctorado: Elena Diego Hernández

Universidad: Universidad de Salamanca

Título de la tesis: *Diccionario bilingüe francés-español de sustantivos predicativos de <aides financières>/<ayudas financieras>. Estudio sintáctico semántico*

Directora: Isabel Uzcanga Vivar. Universidad de Salamanca

Codirectora: Araceli Gómez Fernández. Universidad Nacional de Educación a Distancia

INDICE

LISTA DE FIGURAS Y CUADROS	10
LISTA DE ABREVIATURAS, SÍMBOLOS Y CONVENCIONES GRÁFICAS.	13
CAPÍTULO I: INTRODUCCIÓN GENERAL.....	14
1. Objeto de estudio	14
1.1. Breve presentación del concepto de diccionario electrónico.....	16
2. Objetivos.....	19
3. Metodología.....	19
4. Esquema de la tesis	20
CAPÍTULO II: PREDICADOS NOMINALES DE <AYUDAS FINANCIERAS>.....	23
1. La noción de predicado nominal.....	23
1.1. Cuestionamiento de la noción de <i>partie du discours</i>	24
1.2. El predicado nominal: diferentes posiciones teóricas	27
1.2.1. Sustantivo predicativo: sustantivo morfológicamente derivado	28
1.2.2. Sustantivo predicativo: un sustantivo abstracto	30
1.2.3. Sustantivo predicativo: sustantivo con estructura argumental	33
1.2.3.1. El predicado nominal en la teoría Sens-Texte.....	34
1.2.3.2. El predicado nominal en el modelo de clases de objetos	36
2. El discurso especializado	42
2.1. ¿Lengua de especialidad o discurso de especialidad?.....	43

2.2.	Características del discurso especializado en nuestro corpus.....	47
2.2.1	Características léxicas del discurso de especialidad.....	48
2.2.1.1.	Presencia de sustantivos compuestos	49
2.2.1.2.	Presencia de neologismos.....	51
2.2.1.3.	Presencia de siglas.....	51
2.2.1.4.	Presencia del lenguaje figurado.....	52
2.2.2.	Características sintácticas del discurso especializado.....	52
3.	Terminología y diccionarios	53
3.1.	Terminología	54
3.1.1.	Óptica conceptual.....	54
3.1.1.1.	Cuestionamiento del enfoque conceptual.....	55
3.1.2.	El enfoque léxico-semántico	57
3.2.	Terminografía	59
3.2.1.	Los diccionarios especializados	63
3.2.1.1.	Diccionarios monolingües, diccionarios bilingües y diccionarios multilingües.....	63
3.2.1.2.	La nomenclatura de los diccionarios especializados.....	66
3.2.1.3.	La sintaxis en los diccionarios especializados.....	66
4.	Conclusiones.....	67
CAPÍTULO III: LA FRASEOLOGÍA		69
1.	Breve historia de la fraseología.....	70
1.1.	Fraseología francesa	89
1.2.	Fraseología española	91
2.	El estatuto de la fraseología	94
2.1.	Disciplina autónoma o independiente.....	94
2.2.	Objeto de estudio y límites de la fraseología.....	98
2.3.	Fenómeno marginal o central de la lengua.....	101
3.	¿Qué es una unidad fraseológica?.....	103
3.1.	Terminología fraseológica.....	103

3.2.	Características de las unidades fraseológicas	105
3.2.1.	Polilexicalidad	108
3.2.2.	La frecuencia	108
3.2.3.	Institucionalización.....	109
3.2.4.	El producto	109
3.2.5.	Idiomaticidad o no composicionalidad.....	116
3.3.	Clasificación de las unidades fraseológicas.....	120
3.3.1.	Tipología Gläser.....	121
3.3.2.	Tipología de Mel'čuk	122
3.3.3.	Tipología de G. Gross	124
3.3.5	Tipología Corpas Pastor	125
4.	Colocaciones	127
4.1.	¿Qué es una colocación?	128
4.2.	Propiedades de las colocaciones	132
4.2.1.	La arbitrariedad de la colocación	133
4.2.2.	El sentido composicional o la transparencia de la colocación	135
4.2.3.	El carácter binario de la colocación	138
4.2.4.	Asimetría de los componentes de colocación	139
4.2.5.	La coocurrencia léxica restringida de la colocación	140
4.3.	El estatuto fraseológico de las colocaciones.....	142
5.	Conclusiones	144
CAPÍTULO IV: EL MODELO DE CLASES DE OBJETOS		146
1.	La noción de predicado en la teoría de las clases de objetos	148
1.1.	Los orígenes harrisianos de la noción de predicado	148
1.2.	El predicado como relación	151
1.3.	Formas morfológicas del predicado	152
1.4.	Predicados polimórficos	153
1.5.	Raíces predicativas	154
1.6.	El esquema de argumentos vs la polisemia	155

2.	La frase elemental	156
2.1.	El concepto de frase elemental	156
2.2.	La organización del léxico en tres funciones primarias	158
2.3.	La frase elemental vs la polisemia	159
2.4.	La noción de contexto	161
2.4.1.	Contextos aleatorios	161
2.4.2.	Contextos restringidos	162
2.4.3.	Contextos significativos	162
3.	Descripción de predicados: la noción de empleo	163
3.1.	Un empleo está asociado a un esquema de argumentos	164
3.2.	Un empleo está asociado a un significado	165
3.3.	Un empleo está asociado a una forma morfológica	166
3.4.	Un empleo está asociado a una actualización	167
3.5.	Un empleo está asociado a un sistema aspectual	167
3.6.	Un empleo está asociado a unas modificaciones	169
3.6.1.	Las transformaciones	169
3.6.2.	Las tematizaciones	170
3.6.3.	Las supresiones	170
3.6.4.	Las reestructuraciones	171
3.7.	Un empleo está asociado a un campo específico	172
3.8.	Un empleo está asociado a un registro de lengua	172
4.	Descripción de argumentos: clases de objetos	173
4.1.	La noción de clase de objeto	173
4.2.	Elaboración de clases de argumentos: los operadores apropiados	176
4.3.	Jerarquización de las clases de objetos	178
4.4.	La noción de herencia	180
5.	Descripción de los actualizadores	181
5.1.	Actualización predicativa vs actualización argumentativa	182

5.2.	Actualización de predicados nominales	183
5.2.1.	Actualización mediante el verbo soporte	183
5.2.1.1.	Propiedades sintácticas del verbo soporte	186
5.2.1.1.1.	Supresión del verbo soporte	186
5.2.1.1.2.	Análisis del complemento doble <i>N Prep N</i>	187
5.2.1.1.3.	Restricciones sobre la determinación del nombre predicativo	188
5.2.1.1.4.	Imposibilidad de nominalización del verbo soporte	189
5.2.1.2.	Tipos y funciones de los verbos soporte	191
5.2.1.2.1.	Verbos soporte generales	193
5.2.1.2.2.	Verbos soporte apropiados	193
5.2.1.2.3.	Verbos soporte aspectuales	195
5.2.1.2.4.	Variantes diastráticas	198
5.2.1.2.5.	Verbos soporte y metáfora	199
5.2.1.2.6.	Verbos soporte pasivos	199
5.2.2.	La actualización mediante la determinación	203
5.2.2.1.	La heterogeneidad de la determinación	203
5.2.2.2.	Determinación y clases de objetos	204
5.2.2.2.	La determinación de los predicados	206
6.	La frase compuesta	208
6.1.	Los completivas son argumentos proposicionales	208
6.2.	La subordinada circunstancial	210
6.2.1.	Descripción de los conectores	212
7.	Conclusiones	214
CAPÍTULO V: TRATAMIENTO DE LENGUAS NATURALES (TLN)		217
1.	Las dificultades del TLN	219
1.1.	Polisemia	219
1.1.1.	La noción de polisemia en el TLN	220
1.1.2.	Análisis automático de la polisemia	221
1.1.3.	Tratamiento lexicográfico de la polisemia	222
1.2.	El tratamiento de la fijación	227

1.2.1.	La noción de fijación en el modelo de clases de objetos	228
1.2.2.	Reconocimiento y generación automática de unidades poliléxicas	231
2.	Métodos de análisis automático: método estadístico vs método mixto	232
2.1.	El método estadístico	232
2.2.	El método mixto	233
3.	Recursos léxicos para el etiquetado	235
3.1.	Recursos para el etiquetado morfosintáctico	237
3.1.1.	DicoValence	237
3.1.2.	Lefff (Lexique des Formes Fléchies du Français) - Leffe (Léxico de Formas Flexionadas del Español)	238
3.2.	Recursos para el etiquetado semántico	239
3.2.1.	WordNet	239
3.2.2.	DiCo (Diccionario de Combinatoria)	244
3.2.2.1.	DiCouèbe	247
3.2.2.2.	DiCoPop	249
3.2.2.3.	Lexique Actif du Français (LAF)	251
3.2.3.	Réseau Lexical du français (RL-fr)	253
3.2.4.	Definiens	257
3.3.	Recursos para el etiquetado semántico de funciones	259
3.3.1.	FrameNet	259
3.3.2.	VerbNet	264
3.3.3.	PropBank (Propositional Bank)	266
3.4.	Recursos para etiquetar el aspecto y la temporalidad	269
3.4.1.	TimeBank	269
3.4.2.	Nomage	273
3.5.	Informatización de los diccionarios en papel	277
4.	El tratamiento automático y el método de la clase de objeto	279

4.1.	Recursos lingüísticos para el sistema de TLN del modelo de clases de objetos	281
4.1.1.	Diccionarios electrónicos	281
4.1.1.1.	Diccionarios de morfosintaxis	282
4.1.1.2.	Diccionarios sintáctico-semánticos	285
4.1.2.	Gramáticas locales	288
4.1.3.	Corpus	289
4.2.	Pasos de procesamiento automático	290
4.2.1.	Pretratamiento: normalización de documentos	291
4.2.2.	Segmentación en frases y palabras	291
4.2.3.	Anotación morfosintáctica	292
4.2.4.	Anotación semántica sintáctica	293
5.	Conclusiones	295
CAPÍTULO VI: METODOLOGÍA		297
1.	Principios de análisis	297
1.1.	La lengua como metalenguaje	298
1.2.	La frase elemental como marco de análisis	298
1.3.	El enfoque sintáctico-semántico del léxico	298
1.4.	Preocupación por la exhaustividad y la sistematicidad	299
1.5.	La necesidad de coherencia	299
1.6.	Análisis ascendente	299
2.	Etapas de trabajo y herramientas de análisis	300
2.1.	Establecimiento de la nomenclatura	300
2.2.	Constitución del corpus	302
2.3.	Elaboración de una tabla de análisis	308
3.	Conclusiones	310
CAPÍTULO VII: DESCRIPCIÓN DE PREDICADOS NOMINALES DE <AYUDA FINANCIERA>		312

1. <Ayuda financiera: subsidio>	314
2. <Ayuda financiera: donación>	339
3. <Ayuda financiera: desgravación fiscal>	349
CAPÍTULO VIII: CONCLUSIONES GENERALES	369
BIBLIOGRAFÍA	372
HERRAMIENTAS LEXICOGRÁFICAS CONSULTADAS	373

1. Introducción general

En los albores del tratamiento automático del lenguaje, el optimismo de los especialistas en lingüística computacional les hizo creer que el análisis lingüístico podía basarse en la extracción de reglas gramaticales generales y su aplicación a palabras sencillas. Sin embargo, el discurso no se construye únicamente a partir de la aplicación de un cierto número de reglas gramaticales. Estos análisis pronto se toparon con la esencia misma de las lenguas naturales: la polisemia y la fijación.

La supuesta libertad combinatoria, es decir, la capacidad del hablante para construir un discurso a partir de la libre selección de unidades léxicas y la aplicación de reglas gramaticales, no es más que una utopía lingüística. La lingüística computacional y la lingüística de corpus han demostrado rápidamente que nuestro discurso está muy limitado. Está sujeto a numerosas restricciones de selección.

Por otra parte, la polisemia es una de las características fundamentales de las lenguas naturales. El ser humano es capaz de resolver los problemas de interpretación relacionados con la polisemia basándose en el contexto lingüístico en el que aparece una palabra determinada. En otras palabras, una unidad léxica adquiere un significado determinado en un contexto lingüístico determinado. Del mismo modo, la herramienta informática podrá interpretar correctamente el discurso si puede hacer uso de este contexto lingüístico.

2. Objeto de estudio

El objeto de nuestro estudio es el sustantivo predicativo que designa una ayuda económica. Estos sustantivos pertenecen a veces al lenguaje general (*pensión, reducción, ayuda*), a veces al discurso especializado (*subsidio por desempleo, desgravación*). Como la clase semántica de <ayuda financiera> es muy amplia, la nomenclatura de nuestro diccionario incluye términos procedentes de diferentes ámbitos (derecho, finanzas, educación, protección social, etc.). La originalidad de este estudio se basa, por tanto, en dos aspectos. Por un lado, reunimos en una misma obra sustantivos de la lengua general y sustantivos de los llamados *lenguajes especializados*, y los describimos utilizando las mismas herramientas lingüísticas. De hecho, consideramos que la separación entre lenguas especializadas y lenguas generales carece de fundamento lingüístico. Por otra parte, en lo que se refiere a los términos, no limitamos nuestra investigación a un único

campo especializado, como es práctica común en la investigación terminológica o lingüística especializada.

Nos centramos en la predicación porque es el elemento central del análisis semántico sintáctico. Describir la predicación nos permite describir el contenido proposicional. En nuestro caso, describir la predicación nominal nos permite describir la frase en la que estos sustantivos se incrustan y adquieren su significado.

La gramática tradicional, fuertemente influida por la concepción binaria de la oración procedente de la filosofía antigua (la oración se divide en *tema* y *rema*) y asumida por la gramática generativa (la oración se divide en *grupo nominal* y *grupo verbal*), siempre se ha centrado en el verbo. Sin embargo, la predicación no es una función exclusiva del verbo. Otras partes de la oración, como los sustantivos, los adjetivos e incluso las preposiciones, pueden ser predicativas, pero a menudo se han descuidado en el análisis sintáctico-semántico de la lengua.

Además, como afirma Béjoint (2007), " les lexicographes ont toujours eu pour objet de description l'élément lexical isolé. Dans les dictionnaires, cet élément est débarrassé de son contexte pour que les informations données à son sujet puissent mieux s'adapter à tous les contextes possibles " (p. 17). Las únicas descripciones del entorno sintagmático de la palabra principal contenidas en los diccionarios son mucho más numerosas para los verbos (y para ciertos adjetivos) que para los nombres. Este es claramente el caso de las descripciones de lenguajes especializados, que suelen limitarse a la terminología. El estudio de los sustantivos ha quedado a menudo relegado a simples listas con definiciones o traducciones a otras lenguas. Según Català (2003)

Le terminologue-terminographe s'attache principalement à délimiter et à définir des concepts représentés par des termes. Ses fiches terminologiques donnent priorité aux données de type encyclopédique comme le domaine et la définition, de type signalétique (les sources documentaires) et de type traductologique (équivalents en une ou plusieurs langues cibles). Mais le terme, pour sa part, demeure une réalité linguistique dont le fonctionnement syntacticosémantique est rarement abordé ou même simplement ignoré. (p. 324)

Se han realizado algunos estudios sobre la clase <ayudas> (cf. G. Gross, 2012, 2017; Massoussi y Sfar, 2009; Mel'čuk et al., 1984). Massoussi y Sfar (2009) y G. Gross (2017) realizaron estudios preliminares sobre esta clase semántica en el marco del modelo de clases de objetos. Estos estudios permitieron realizar una primera clasificación de los diferentes tipos de verbos de apoyo seleccionados por estos predicados nominales. Sin

embargo, sólo se trata de estudios embrionarios, ya que no pretenden ser ni exhaustivos ni sistemáticos.

Nos hemos basado en los léxicos francés y español para crear la nomenclatura de nuestro diccionario electrónico. Es importante señalar que el resultado no es un diccionario bilingüe, sino dos diccionarios monolingües coordinados.

2.1. Breve introducción al concepto de diccionario electrónico

El término *diccionario electrónico* se refiere a dos objetos diferentes: los diccionarios en formato electrónico destinados al uso humano y los diccionarios diseñados para ser utilizados por programas de tratamiento automático. Entre los diccionarios electrónicos para uso humano, cabe distinguir entre los que son una versión electrónica de un diccionario en papel (a menudo denominados *diccionarios informatizados*) y los que se elaboran desde el principio en un soporte electrónico.

Veronis (2004) señala que, tradicionalmente, los diccionarios de uso humano han tratado de describir el significado de las palabras más que sus usos (con la excepción de algunas indicaciones sobre el dominio o el registro). Hacia mediados del siglo XX^e, siguiendo los trabajos lexicográficos de Hornby (1942, 1954), algunos diccionarios empezaron a incorporar sistemáticamente información sintáctica, distributiva o paradigmática. El *Dictionnaire du français contemporain* de Dubois (1966) incorpora patrones de valencia (número y naturaleza de los complementos) al agrupar las entradas. Sin embargo, este diccionario carece de sistematicidad. Posteriormente, el *Trésor de la Langue Française* (1971-1994), basado en el estudio de un amplio corpus, también contiene algunas indicaciones de combinatoria, pero distan mucho de ser sistemáticas. En español, el *Diccionario de uso del español* (1966-1967) de María Moliner¹ es muy interesante desde el punto de vista combinatorio y ha dejado una fuerte huella en la lexicología y la lexicografía españolas, pero sigue careciendo de sistematicidad.

La llegada de la informática ha cambiado la forma de representar los diccionarios y el propio trabajo del lexicógrafo. El desarrollo de grandes corpus y de las herramientas informáticas que permiten utilizarlos ha revolucionado la lexicografía. Por un lado, la lexicografía se ha orientado hacia la descripción de los usos de las palabras, es decir, hacia la combinatoria. Por otro, los lexicógrafos comprendieron la necesidad de descripciones

¹ Cuando María Moliner hojeó el *Learner's Dictionary of Current English* de Hornby (1948), sembró la semilla de su *Diccionario de uso del español*.

exhaustivas y sistemáticas. Desde los años 80, los lexicógrafos centran su atención en la combinatoria de palabras² (cf. Verlinde et al., 2005). En este contexto se publicó el *Dictionnaire explicatif et combinatoire du français contemporain* de Mel'čuk et al (1984, 1988, 1992, 1999)³. En español, el diccionario de combinaciones más importante no se publicó hasta 2004, editado por Bosque: *REDES, Diccionario combinatorio del español contemporáneo*⁴.

Estas obras lexicográficas más recientes son muy interesantes desde el punto de vista lexicológico y lexicográfico, pero siguen siendo insuficientes para el tratamiento automático. El diccionario electrónico destinado a los programas de tratamiento automático no es un diccionario de consulta. Se trata más bien de una herramienta que contiene información lingüística formalizada que será utilizada por los sistemas informáticos para reconocer o generar textos. Este tipo de diccionario es, por tanto, el núcleo de los sistemas automáticos de tratamiento del lenguaje natural. Como el autómata no tiene conocimiento previo de la lengua ni del mundo, para que el diccionario electrónico sea eficaz debe contener información mucho más explícita y precisa que la presentada en un diccionario informatizado o en un diccionario de papel (cf. Pruvost, 2000).

A modo de resumen, Courtois y Silberztein (1990) repasan las diferencias más importantes entre los diccionarios electrónicos destinados a los autómatas y lo que ellos denominan "diccionarios comerciales", es decir, diccionarios para uso humano:

- a) Los diccionarios electrónicos deben ser lo más completos posible; sin embargo, un diccionario comercial nunca está completo;
- b) La información de un diccionario electrónico debe ser explícita, mientras que los diccionarios comerciales dejan implícita información que se da por sabida; así, los diccionarios comerciales suelen recurrir a conocimientos previos del mundo ;

² A menudo se han utilizado herramientas estadísticas para identificar las co-ocurrencias relevantes: *información mutua, puntuación Z o puntuación T*.

³ Hay que señalar que las definiciones lexicográficas de este diccionario se basan en la introspección, y que el corpus sólo sirve de comprobación a posteriori. Se trata de un diccionario en papel, pero se han desarrollado interfaces informáticas que contienen gran parte del diccionario.

⁴ Se trata de un diccionario en papel. Posteriormente, siempre bajo la dirección de Bosque, se publicó el *Diccionario combinatorio práctico del español contemporáneo* (2006) (también en papel).

c) Los diccionarios electrónicos están intrínsecamente ligados a los programas de tratamiento automático. Toda la información que contienen estos diccionarios es precisa y coherente, ya que debe formalizarse para poder ser utilizada por los programas informáticos.

Sin embargo, la oposición entre diccionarios en papel y diccionarios electrónicos no está justificada, ya que los segundos se nutren de los datos de los primeros. Así pues, los diccionarios en papel contienen información semántica que puede utilizarse en la elaboración de diccionarios electrónicos, como definiciones, marcas vinculadas al ámbito técnico de uso de una determinada acepción, determinadas construcciones combinatorias o restricciones, vínculos semánticos con otras unidades léxicas (sinónimos, antónimos, hiperónimos o hipónimos) e incluso listas de palabras compuestas construidas a partir de la palabra de entrada.

3. Objetivos

El objetivo de este estudio es describir y clasificar la semántica sintáctica de los sustantivos predicativos <ayuda financiera> en francés y español con vistas a elaborar un módulo para un diccionario electrónico. Este módulo no adopta la forma de un diccionario al uso, sino que se construye en forma de base de datos. Nos interesa especialmente la integración de estos sustantivos predicativos en el marco oracional. Por ello, nos proponemos describir la estructura argumental de estos sustantivos predicativos y analizar su combinatoria.

Nuestra hipótesis de trabajo es que la descripción exhaustiva y sistemática del funcionamiento sintáctico-semántico de los sustantivos predicativos estudiados debería permitirnos agruparlos en clases semánticas con un mismo comportamiento en la oración. De este modo, la herramienta de procesamiento automático que utilizara este recurso podría predecir automáticamente la combinatoria verbal de estos sustantivos.

Nos gustaría señalar que nuestro enfoque es estrictamente lingüístico. El objetivo de este estudio es investigar cómo funciona un léxico específico en la frase y no hacer un inventario exhaustivo de los distintos tipos de ayudas financieras disponibles en Francia y España, ni clasificarlas según el tipo de ayuda o el beneficiario u otro tipo de criterios extralingüísticos.

4. Metodología

El presente trabajo se inscribe en el marco de una teoría léxica, la teoría de las clases de objetos. Esta teoría es la continuación de la gramática transformacional de Harris (1976) y de los trabajos sobre léxico-gramática (M. Gross, 1975; Leclère, 1998). Se basa en la clasificación semántica, completada con reglas sintácticas.

Las unidades léxicas se conciben como partes primordiales de la lengua. Esta metodología permite una descripción multilateral del léxico, es decir, las unidades léxicas descritas se consideran simultáneamente desde su vertiente semántica (establecimiento de clases semánticas), desde su vertiente sintáctica (descripción de su patrón argumental) y desde su vertiente léxico-combinatoria (colocaciones).

En nuestro estudio proponemos una descripción de los sustantivos predicativos de <ayuda financiera> que abarca varios niveles de análisis. Desde un punto de vista semántico, los sustantivos predicativos se agrupan en clases semánticas. La pertenencia a una clase se basa principalmente en criterios de sinonimia o cuasi-sinonimia. Desde el punto de vista morfológico, en nuestro diccionario electrónico, las unidades léxicas estudiadas se caracterizan en género y número. Además, presentamos todas las formas que puede tener un determinado sustantivo predicativo; por ejemplo, las lexías *alloc* o *allocs* se utilizan a menudo en un registro lingüístico coloquial (como pudimos comprobar consultando foros de Internet) en lugar de los términos *allocation* o *allocations*; también hemos enumerado las siglas que corresponden a determinados nombres de ayudas financieras, ya que se utilizan a menudo en el discurso (por ejemplo, *APL*). Por último, desde un punto de vista sintáctico, todas las unidades léxicas estudiadas se definen por su patrón de argumentos y por su combinatoria verbal.

5. Esquema de la tesis

En el primer capítulo, discutimos las nociones fundamentales para identificar el objeto de nuestro estudio: el sustantivo predicativo de <ayuda financiera>. De este modo, examinamos diferentes concepciones de la noción de sustantivo predicativo y explicamos cómo se concibe y describe en el marco de la teoría de las clases de objetos. En este capítulo, que nos permite identificar y comprender mejor el objeto de nuestra investigación, abordamos también las características del discurso especializado para observar en qué se diferencia del discurso general, y hacemos una breve aproximación a los estudios de terminología y terminografía.

El segundo capítulo está dedicado a la fraseología. Partiendo de una breve reseña histórica, seguimos el desarrollo de esta disciplina y la ampliación de la noción de unidad fraseológica. Los límites de la noción de unidad fraseológica siguen siendo imprecisos, dependiendo del enfoque lingüístico adoptado. Por lo tanto, es necesario definir lo que entendemos por unidad fraseológica. Nos centraremos en la noción de colocación, ya que es uno de los elementos clave de nuestra investigación.

En el capítulo tres, hacemos una presentación detallada de la teoría de clases de objetos. Explicamos sus postulados fundamentales y cómo puede describirse el lenguaje dentro de este marco teórico. Esta presentación nos permite comprender por qué este enfoque teórico responde a las necesidades del procesamiento automático.

En el capítulo cuarto, hacemos una breve incursión en el tratamiento automático, ya que determina el contenido y la forma de nuestra investigación. En primer lugar, analizamos las dos principales dificultades a las que se enfrenta el tratamiento automático de las lenguas naturales: la polisemia y la fijación. A continuación, presentamos los dos principales métodos de análisis lingüístico que existen en la actualidad y especificamos el que adoptamos, al tiempo que justificamos nuestra elección. Para hacer un breve repaso de los trabajos de lingüística computacional en relación con la lexicología y la lexicografía, presentamos algunos recursos léxicos para el etiquetado morfosintáctico y semántico, el etiquetado de aspecto y temporalidad y la informatización de diccionarios en papel. Por último, nos centramos en el tratamiento automático dentro del marco teórico de las clases de objetos. Así, presentamos algunos recursos lingüísticos y las etapas del sistema de tratamiento automático específico de esta teoría.

En el quinto capítulo explicamos la metodología que hemos seguido para elaborar nuestro diccionario electrónico. Comenzamos presentando los principios de análisis que rigieron nuestras elecciones metodológicas. A continuación, detallamos las fases de trabajo que hemos seguido y las herramientas de análisis que hemos desarrollado para llevar a cabo nuestra investigación.

Por último, en el capítulo seis, presentamos el diccionario electrónico que hemos desarrollado. Describimos las propiedades sintáctico-semánticas de los predicados nominales de <ayuda financiera> y comparamos los resultados obtenidos en las dos lenguas.

Por último, en la conclusión extraemos conclusiones de los resultados de nuestra investigación y revisamos la eficacia de la metodología de trabajo y el marco teórico que hemos adoptado.

6. Noción de predicado nominal

Como muchos términos del metalenguaje lingüístico, el término *predicado* designa distintas entidades lingüísticas. A primera vista, el término tiene dos significados principales. Por un lado, en la tradición lógica clásica, el *predicado* designa lo que completa al sujeto para formar una proposición. Por otro lado, la lingüística moderna, heredera de la lógica renovada de Frege (1867/1979), entiende por *predicado* el centro funcional (u organizador) de la oración, que rige variables que se denominan *argumentos* (tanto el sujeto como los complementos son argumentos)⁵. En otras palabras, "le prédicat est une catégorie logico-sémantique de nature relationnelle, une sorte de fonction⁶ sur la base de laquelle se calcule le sens de tout énoncé" (Mejri, 2019, p. 112). En lo sucesivo, es sólo este segundo significado el que conservaremos en nuestro trabajo.

El sustantivo también ocupa un lugar importante en lingüística, pero su definición y clasificación también están sujetas a interpretaciones divergentes.

En esta sección, repasamos los problemas de definición de las nociones *parte de oración* y *sustantivo*. Seguidamente, resentamos las principales interpretaciones del término *sustantivo predicativo*: para algunos, el sustantivo predicativo es un sustantivo deverbal; otros lo entienden como un sustantivo abstracto; y otros consideran que el sustantivo predicativo es un sustantivo con estructura argumental. Por último, definimos brevemente qué se entiende por predicado nominal y presentamos las características de los predicados de <ayudas>.

⁵ La antigua dicotomía aristotélica *sujeto/predicado* o *tema/rhema* se sustituye por *predicado/argumento* (o *función/argumento*, según la terminología utilizada por Frege). Para un resumen de los usos del término *predicado*, véase Merle (2009, pp. 6-8); Muller (2013, pp. 51-53).

⁶ F(x), où F est la fonction et x, la variable.

7. Discurso especializado

La nomenclatura del diccionario que hemos elaborado incluye sustantivos de la lengua general y sustantivos de lo que suele llamarse la lengua especializada. Consideramos que nada impide estudiar conjuntamente todas estas unidades léxicas. Además, la particularidad de nuestra investigación reside en el hecho de que estos sustantivos predicativos pertenecen a varios ámbitos especializados, como el derecho civil, el derecho laboral, la legislación social, las finanzas, la educación, etc.

El término *lengua especializada* y la definición del concepto al que se refiere se han debatido ampliamente en la literatura lingüística. Existe una falta de consenso que deriva, en parte, de la dificultad de establecer la frontera entre lo que es general y lo que es especializado en la lengua. Así, podemos distinguir tres concepciones principales de la relación entre lenguas generales y especializadas:

- i. Las lenguas especializadas son códigos lingüísticos completos, diferentes de la lengua general. Son, por tanto, sub-lenguas al mismo nivel que la lengua general.
- ii. Las lenguas especializadas son variantes léxicas de la lengua general. La única especificidad de las lenguas especializadas es el léxico. En cuanto a las características sintácticas, semánticas y enunciativas, no hay nada que las diferencie de la lengua general.
- iii. Otros lingüistas han adoptado la postura intermedia de que las lenguas especializadas son subconjuntos dependientes de la lengua general, pero tienen sus propios elementos específicos, esencialmente léxico-sintácticos. Estas especificidades dependen de la pragmática y la extralingüística. Por tanto, ya no se trataría de una yuxtaposición entre la lengua general y la lengua especializada (primera concepción), ni de una inclusión (segunda concepción), sino más bien de un continuo entre la lengua general y la lengua especializada.

Es bajo esta última concepción de las lenguas especializadas como hemos emprendido nuestra investigación. Coincidimos con Lerat (1997) cuando afirma que "les langues de spécialité ne sont rien d'autre que des usages spécialisés des langues naturelles" (p. 15). Una lengua especializada no es más que la utilización de las reglas de la lengua general para dar cuenta técnicamente de conocimientos especializados (Lerat, 1995, p. 21).

8. Breve repaso a la fraseología

La fraseología ha recorrido un largo camino desde que se la consideraba una disciplina lingüística. Ignorada durante mucho tiempo por los lingüistas y asociada después a la lexicología, la gramática y la estilística, la fraseología tuvo dificultades para encontrar su lugar como disciplina independiente. Antes de convertirse en una disciplina científica reconocida por los lingüistas, la fraseología se manifestó en la lexicología y la paremiología. En la mayoría de las lenguas, los primeros indicios de interés por las unidades fraseológicas se encuentran en las colecciones de máximas, proverbios, etcétera. A lo largo de los años, estas obras han ido incorporando todo tipo de expresiones más o menos fijas. A partir del siglo XIX, con el auge de la lexicografía, estas secuencias fijas se introdujeron en los diccionarios. Las primeras referencias a la existencia de elementos fraseológicos datan de finales del siglo XIX y aparecen en obras de lingüística general de diversas lenguas (Hermann Paul, 1880, para el alemán; Henry Sweet, 1891, para el inglés). Por lo que respecta a las lenguas que estudiamos, en 1897 Michel Bréal, en su *Essai de sémantique*, señaló la existencia de grupos de palabras fijas, que denominó *locuciones, fórmulas o grupos articulados*, y que destacó como preconstruidos e idiomáticos. En español, hubo que esperar hasta 1950 para que Julio Casares publicara *Introducción a la lexicografía moderna*. En esta obra, el autor intenta identificar y clasificar las unidades que constituyen el objeto de estudio de la fraseología.

En las últimas décadas del siglo XX, la investigación sobre fraseología se desarrolló en torno a tres grandes bloques: el bloque oriental soviético, el estructuralismo europeo occidental y la lingüística norteamericana. Los dos primeros bloques mantuvieron relaciones muy estrechas, como lo demuestran los trabajos sobre fraseología alemana, fuertemente influidos por la investigación soviética (cf. Häusermann, 1977; Dobrovolskij, 1988, entre otros). En Estados Unidos, la gramática generativa ha excluido durante mucho tiempo la descripción de construcciones fraseológicas. De hecho, encontró serios problemas a la hora de aplicar transformaciones sintácticas a este tipo de unidades (cf. Chafe, 1968; Fraser, 1970; Ruwet, 1983; Weinreich, 1966/1969). Fue la gramática transformacional de M. Gross (1982, 1988, 1993) la que consiguió integrar la descripción de las unidades fraseológicas.

9. El modelo de clases de objetos

Entre las teorías que se centran en el léxico, hay algunas que conceden un lugar central a la semántica; otras que abordan la descripción de las unidades léxicas desde el punto de vista de la sintaxis (en particular, léxico-gramática); y otras que adoptan una posición intermedia de equilibrio entre semántica y sintaxis, como es el caso del modelo de clases de objetos (cf. G. Gross, 1992, 1994a, 1995, 2012; Le Pesant y Mathieu-Colas, 1998). Se trata de un método de descripción del lenguaje destinado al tratamiento automático: proporciona una relación exhaustiva y sistemática de los hechos del lenguaje, adaptándose así a las características específicas de la herramienta informática. Es sobre los fundamentos teóricos de este modelo que hemos emprendido la descripción de los sustantivos predicativos del <ayudas>⁷.

Este método de descripción de la lengua se basa en la observación de las posibles construcciones sintácticas y combinaciones léxicas. El principio básico de este enfoque es la gramática transformacional de Harris (1971, 1976), que postula que la unidad mínima de análisis es la oración y no el signo lingüístico (en el sentido saussureano)⁸. El modelo de clases de objetos se inspira en la gramática transformacional de Harris (1971, 1976) y en los trabajos sobre léxico-gramática de M. Gross (1975, 1981)⁹ desarrollados en el laboratorio LADL (Laboratoire d'Automatique Documentaire et Linguistique), y fue desarrollado por G. Gross en el laboratorio LLI (Laboratoire de Linguistique Informatique), posteriormente denominado LDI (Lexiques, Dictionnaires et Informatique)¹⁰. Las clases de objetos permiten que la semántica desempeñe un papel más importante que el que tuvo para Harris o M. Gross. De hecho, el modelo de clases de objetos es una respuesta a la falta de precisión de las características sintáctico-semánticas

⁷ Por convención, en el modelo de clases de objetos, el nombre representativo de una clase de objetos se encierra entre comillas <>.

⁸ Según Harris, las propiedades sintácticas de una unidad léxica deben analizarse en el marco de la oración simple, que está formada por un operador (predicado) y sus objetos (argumentos: sujeto gramatical y varios complementos). Es el operador el que selecciona el número y la distribución de sus objetos. Para afinar la clasificación de los operadores y distinguir los distintos usos de una misma forma, Harris los tamiza a través de una serie de transformaciones, siempre observables en el marco de la oración.

⁹ Para comprender los principios generales del léxico-gramatical, véase también Boons et al, 1976a, 1976b; Leclère, 1989; y el número 59 de la revista *L'information grammaticale*, dedicado íntegramente al léxico-gramatical.

¹⁰ En la actualidad, el laboratorio LLI se ha convertido en el TTN (Text, Digital Theories), donde aún podemos encontrar la huella de la investigación sobre el modelo de clases de objetos.

con las que habíamos estado trabajando durante años, y también una respuesta a la naturaleza intuitiva de otras taxonomías existentes. Para que el ordenador pueda reconocer y generar textos, necesita una descripción completa de la lengua. No podemos contentarnos con reglas generales de gramática, morfología y sintaxis. Necesitamos una descripción detallada de todas las propiedades lingüísticas de cada palabra. Estas descripciones se plasmarán en léxicos, de ahí la importancia que se concede al desarrollo de diccionarios electrónicos. Resulta que dos de los principales problemas del tratamiento automático son fenómenos lingüísticos muy extendidos en las lenguas naturales: la polisemia y la fijación. Demostraremos que las clases de objetos son extremadamente útiles para resolver estos problemas.

Una descripción detallada de los conjuntos léxicos requiere a menudo la subdivisión de las clases en subclases, es decir, una clasificación jerárquica en hiperónimos e hipónimos. La noción de clase objeto redefine la relación hiperónimo/hipónimo. Esta jerarquización del léxico no se basa en una concepción ontológica, como es el caso de los árboles semánticos. Nuestra concepción del universo no está necesariamente correlacionada con la expresión lingüística. El enfoque de las clases de objetos es rigurosamente lingüístico: la jerarquización de las clases de objetos se basa en la observación de las construcciones sintácticas y las posibles combinaciones léxicas en los corpus. Las clases de objetos no se basan en relaciones semánticas internas (sinonimia o hiperonimia) ni en clasificaciones ontológicas, sino en la combinatoria de elementos dentro de la frase.

La creación de clases de argumentos se basa exclusivamente en criterios lingüísticos, es decir, todos los sustantivos pertenecientes a una clase deben tener las mismas propiedades semánticas y sintácticas. En cuanto al tratamiento automático, la ventaja de las clases de objetos (y de las subclases e hiperclases) es que todos los sustantivos de una clase determinada pueden factorizarse y describirse de forma compacta. Por último, si asociamos a cada sustantivo el código de la clase a la que pertenece y describimos los patrones argumentales de los predicados utilizando estas clases, el autómata podrá reconocer y generar todas las oraciones posibles para un uso determinado.

10. Tratamiento de la polisemia

La polisemia es uno de los problemas más difíciles en TLN. Desde un punto de vista teórico, muchos enfoques se han centrado en delimitar y relacionar los significados de una palabra polisémica. Su tratamiento depende del grado de abstracción del análisis y de

la precisión de las discriminaciones entre los distintos significados. Por consiguiente, la polisemia plantea grandes problemas para el tratamiento automático del lenguaje. Más aún si se trata de analizar la polisemia del predicado, porque, no lo olvidemos, el predicado desempeña un papel clave en la oración al imponer sus restricciones de selección al resto de los elementos de su entorno lingüístico.

11. Fases de trabajo y construcción del corpus

El modelo de clases de objetos forma parte de la lingüística descriptiva, es decir, pretende explicar todos los usos de las unidades léxicas de una lengua determinada, observados en contexto.

El proceso de trabajo no se realizó simultáneamente en ambas lenguas. Primero analizamos los predicados nominales de <aides> en francés y, después, en español. En cuanto al análisis y la descripción, seguimos los mismos pasos en ambas lenguas. En primer lugar, establecimos la nomenclatura que se iba a describir. En segundo lugar, estudiamos cada unidad léxica de nuestro nomenclátor en su contexto, utilizando varios corpus ya recopilados y, en particular, el motor de búsqueda Google. En tercer lugar, a partir de la observación de las unidades léxicas en contexto, pudimos elaborar una tabla de análisis que tiene en cuenta las propiedades morfológicas, semántico-sintácticas y pragmáticas. Por último, la aplicación de la rejilla de análisis nos permitió identificar y enumerar todos los usos de las palabras conservados en el nomenclátor, y agrupar en clases de objetos los usos que comparten las mismas propiedades sintáctico-semánticas. Obtuvimos así dos diccionarios electrónicos de predicados nominales: uno en francés y otro en español. La última etapa de nuestro trabajo consistió en coordinar estos dos diccionarios. Recordemos que se trata de dos diccionarios de usos de palabras, y que la noción de uso combina propiedades léxicas, sintácticas y semánticas. Así pues, la coordinación de los dos diccionarios no se limitaba a la traducción de las palabras que constituían las entradas de los diccionarios, sino que también era necesario comparar las propiedades combinatorias de cada entrada en las dos lenguas. A continuación presentamos las tres etapas principales de nuestro trabajo.

La especificidad de nuestro objeto de estudio reside en que es muy heterogéneo. Por un lado, identificamos unidades monoléxicas y poliléxicas. Por otra parte, identificamos empleos generales y empleos especializados pertenecientes a distintos ámbitos de

especialización. Por consiguiente, las fuentes de información utilizadas para elaborar el nomenclátor han sido muy variadas y multidisciplinarias.

El inventario de los sustantivos predicativos de <ayuda financiera> se basó en criterios semánticos, a partir de su significado tal y como se especifica en las obras lexicográficas. Comenzamos consultando un gran número de diccionarios de lengua generales y especializados, así como obras especializadas. Como resultado de esta primera fase, recuperamos una base léxica de nombres para <ayudas financieras> que desarrollamos de dos formas distintas. Por lo que respecta a las palabras de la lengua general, buscamos sinónimos en diccionarios generales y diccionarios de sinónimos, con el fin de enriquecer la nomenclatura establecida inicialmente. En cuanto a los términos especializados (como los nombres concretos de determinadas prestaciones), nos centramos sobre todo en la exploración de la web (sitios de instituciones, asociaciones, empresas, foros especializados, etc.). La búsqueda de términos especializados en Internet resultó más fructífera que la consulta de obras impresas especializadas. En primer lugar, porque nuestro campo de estudio (las ayudas) está en constante evolución: se introducen nuevas desgravaciones, nuevos tipos de crédito adaptados a realidades cambiantes, han surgido nuevas formas de colaborar en un proyecto gracias a internet, etc. En segundo lugar, nuestro uso de la red nos ha permitido identificar usos más o menos comunes que se pasaban por alto en diccionarios y obras especializadas.

La tarea de identificar (y también de analizar sintáctica y semánticamente) los términos fue mucho más complicada debido al carácter multidisciplinar de nuestro tema y a los conocimientos extralingüísticos necesarios para describir estas unidades léxicas. En efecto, la descripción de estas unidades léxicas requiere no sólo competencias lingüísticas, sino también competencias extralingüísticas que nos permitan comprender plenamente el significado de estos términos y su funcionamiento en la frase.

La naturaleza cambiante de algunos de sus términos nos planteó un problema metodológico: tuvimos que decidir si incluíamos en nuestra nomenclatura los nombres de subvenciones (en particular, desgravaciones) que ya no existen o que han cambiado de nombre. Como estos nombres aún pueden encontrarse en los archivos y en la web, el autómata podría tener que analizarlos. Además, estos términos siguen formando parte del vocabulario de los hablantes nativos. Por ello, hemos decidido enumerarlos y describirlos. No obstante, queremos subrayar que el inventario de auxiliares sustantivos pertenecientes al discurso especializado es exhaustivo, pero nunca puede ser completo ni definitivo.

Además, hemos observado que existen variaciones diatópicas en el léxico de los ámbitos estudiados. Los nombres de prestaciones, seguros, créditos, etc. no son necesariamente los mismos en todos los países hispanohablantes o francófonos. Es más, no todos los países tienen las mismas rentas sociales. Así que tuvimos que tomar una decisión: en este trabajo de investigación describimos los sustantivos predicativos de <ayudas> en español de España y en francés de Francia.

12. El corpus

Puesto que uno de los principios metodológicos que guían nuestra investigación es considerar la oración elemental como marco para analizar los usos de las unidades léxicas, y puesto que afirmamos que la descripción de la lengua debe basarse en métodos estrictamente lingüísticos, resultó necesario recurrir al corpus. En efecto, la distancia entre la intuición del lingüista y los datos recogidos objetivamente de los textos puede ser enorme y sistemática (Sinclair, 1991).

Lerat (2002) sostiene que "les corpus sont indispensables à la description des emplois des mots spécialisés. Corpus textuels mais aussi corpus lexicographies" (p. 205-204). Así pues, utilizamos estos dos tipos de corpus para identificar las colocaciones verbales. En primer lugar, buscamos en los diccionarios generales¹¹ y en los diccionarios de combinaciones de palabras¹² para identificar una lista inicial de colocaciones. A continuación, a partir de estas combinaciones de palabras iniciales, lanzamos consultas en corpus textuales informatizados con un doble objetivo: en primer lugar, probar otras combinaciones [verbo de apoyo + sustantivo predicativo] que no habíamos encontrado en el corpus lexicográfico; en segundo lugar, observar el comportamiento de estos sustantivos predicativos en contexto.

El acceso a corpus informatizados ha animado a los lexicógrafos a centrarse en las afinidades entre unidades léxicas, considerándolas auténticas unidades funcionales del discurso (Sinclair, 1991). Las posibilidades técnicas que ofrece la informática han permitido poner de relieve las coocurrencias. El uso de corpus informatizados también ha

¹¹ Para el francés, hemos utilizado la versión digital del *Petit Robert* y el *TLFi*. Para el español, hemos consultado el *Diccionario de uso del español* de María Moliner.

¹² En francés, consultamos el *Dictionnaire des combinaisons de mots* du Robert (Le Fur, 2007) y el diccionario de co-ocurrencias contenido en la herramienta Antidote (en varias versiones, de la versión 6.4 a la versión 10). En español, hemos utilizado los diccionarios *Redes* (Bosque, 2004) y *Práctico* (Bosque, 2006).

facilitado el estudio de las relaciones entre los entornos sintagmáticos de las unidades léxicas y su significado. Los lexicógrafos ya no se interesan por la palabra aislada, sino por la unidad léxica insertada en un entorno, es decir, por lo que hemos definido como el uso de las palabras.

El primer paso de nuestra metodología de investigación fue establecer una nomenclatura. En otras palabras, teníamos una lista de sustantivos predicativos y necesitábamos identificar o comprobar su combinatoria verbal en contexto. Como punto de partida para construir el corpus, tuvimos que pensar en el tipo de corpus que necesitaríamos para encontrar ocurrencias de los sustantivos predicativos de nuestra nomenclatura. Nos guiaba una preocupación por la homogeneidad y la heterogeneidad al mismo tiempo. Por un lado, queríamos trabajar con un corpus lo suficientemente homogéneo como para poder obtener datos fiables. Por lo tanto, consideramos ciertas características que garantizarían la homogeneidad del corpus:

- establecer límites cronológicos (siglos XX^e y XXI^e);
- desterrar las variaciones diatópicas;
- Deben evitarse los textos poéticos, ya que las preocupaciones estilísticas de estos últimos pueden dar lugar a un discurso que no se corresponda con el uso corriente de la lengua;
- prohibir las traducciones, es decir, trabajar sólo con textos originales.

Por otra parte, también queríamos que el corpus fuera lo suficientemente heterogéneo como para ser representativo del francés o del español perteneciente tanto al discurso general como al especializado. Así, era necesario tener en cuenta textos literarios, periodísticos, administrativos, etc.

Además, el corpus debe ser pertinente para los ámbitos tratados. Dado que la nomenclatura de nuestro diccionario abarca tanto los usos lingüísticos generales como los especializados, nuestro estudio se basó en dos tipos de recursos diferentes: vastos corpus generales y la web. Para los corpus generales, utilizamos Frantext para el francés, CREA y CORPES XXI para el español, y Sketch Engine (Kilgarriff et al., 2004, 2014) para ambos idiomas. Las interfaces de estos corpus nos permitieron orientar nuestras consultas seleccionando determinados criterios; por ejemplo, delimitación cronológica (sólo textos de los siglos XX y XXI), delimitación geográfica (textos producidos en Francia y España), género literario (prohibimos la poesía), etc. La utilización de estos corpus ha

dado a veces resultados escasos, sobre todo en ámbitos especializados. Por ello, hemos considerado necesario utilizar la web como corpus.

La utilización de la web como *corpus*, tendencia conocida como *web as corpus*, no cuenta con el apoyo unánime de los lingüistas, pero en las dos últimas décadas ha atraído cada vez más partidarios. Los que están en contra citan la naturaleza artificial del discurso en la web, su incorrección (hay muchos errores y algunos son muy frecuentes), su irregularidad (algunos dominios están mucho más representados que otros) y su naturaleza cambiante. Por ejemplo, según Sinclair (2004): "The World Wide Web is not a corpus, because its dimensions are unknown and constantly changing, and because it has not been designed from a linguistic perspective". Sin embargo, consideramos que ni el carácter "ilimitado" de la web ni su naturaleza cambiante son obstáculos para el uso de la web como corpus (cf. Berland y Grabar, 2002; Colson, 2007, 2010; Ferraresi et al., 2010; Fletcher, 2013). La web como corpus ofrece muchas más ventajas que inconvenientes reales. El desarrollo de la web ha dado la vuelta a la noción de corpus y, en nuestra opinión, ya no es tan rígida como antes. La web proporciona una enorme cantidad de recursos que, con ciertas precauciones, son perfectamente válidos y explotables para la investigación lingüística.

Por lo tanto, decidimos utilizar la web como corpus. Si entendemos un corpus tal y como lo define Sinclair (2004): " A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research ", el corpus de la web cumple nuestros criterios de investigación y constituye una verdadera representación de la lengua. En primer lugar, como ya se ha mencionado, muchas de las áreas especializadas en las que hemos estado trabajando están infrarrepresentadas en los corpus generales. Las ocurrencias de determinados usos que ofrecían estos corpus no eran suficientemente representativas y, en ocasiones, no se ofrecía ninguna ocurrencia. Por otro lado, la web constituye una enorme base de datos de documentos pertenecientes a múltiples campos de especialización. Por otro lado, consideramos que la web constituye un corpus muy completo, ya que devuelve una imagen aproximada de todo el discurso. Dado que el objeto de nuestro estudio está en constante innovación (se crean constantemente nuevos tipos de ayudas, becas, créditos, etc.), el corpus más actualizado es sin duda la web.

De hecho, muchos de los sitios tratan de los temas especializados que abarca nuestra nomenclatura, lo que nos permite observar el uso sintagmático de las unidades léxicas

que nos interesan. Los foros han resultado especialmente interesantes para identificar colocaciones verbales. Por supuesto, nos hemos encontrado con algunos usos erróneos. El trabajo de validación de las colocaciones verbales fue, por tanto, lento y arduo: establecimos una base mínima de 100 ocurrencias para aceptar una colocación verbal dada. Hay que decir que el criterio de frecuencia no es suficiente, ya que algunas de las unidades léxicas de nuestro nomenclátor tienen una frecuencia extremadamente baja (por ejemplo, *étrennes*). En los casos en los que obtuvimos pocas ocurrencias (por debajo de la base), nos vimos obligados a observar el contexto en el que aparecían estas ocurrencias para poder comprobar si el emisor de este texto dominaba bien la lengua o si cometía numerosos errores combinatorios (en este último caso, podría tratarse de un hablante que utiliza el francés o el español sin dominarlo). En el primer caso, consideramos que se trataba de coocurrencias infrecuentes pero correctas¹³; en el segundo, sin embargo, las consideramos errores. Por lo tanto, hay que examinar cuidadosamente los corpus de documentos recogidos en la web, ya que su calidad y fiabilidad varían enormemente.

Hemos utilizado el motor de búsqueda comercial Google por el volumen del corpus y su potencial de consulta.

13. Tabla de análisis

La tarea de identificar los sustantivos predicativos de <ayudas económicas> ha sido en gran medida intuitiva, por lo que hemos obtenido una lista muy heterogénea de sustantivos. Con el fin de clasificar estos sustantivos en función de sus propiedades comunes, elaboramos una tabla de análisis que aplicamos sistemáticamente a todos los sustantivos identificados¹⁴. Las propiedades que componen nuestra tabla de análisis se eligieron por su compatibilidad con los principios de análisis del modelo de clases de objetos expuesto anteriormente. Así pues, hemos seleccionado cuatro categorías generales de propiedades lingüísticas.

Las propiedades configuracionales se refieren al patrón de los argumentos en los sustantivos predicativos. Se subdividen en propiedades estructurales (el número de

¹³ Estos casos escapan a menudo al análisis de corpus "tradicionales" como Frantext, CREA, CORDE o incluso Sketch Engine.

¹⁴ Para desarrollar este cuadro de análisis, nos basamos en trabajos realizados en el laboratorio LDI, como Grezka (2006); Grezka y Buvet (2007); G. Gross (2008, 2017), Massoussi y Sfar (2009).

argumentos y el modo en que están estructurados) y propiedades distribucionales (la naturaleza morfosintáctica de los argumentos). Recordemos que para una forma dada existen tantos usos diferentes como patrones argumentales¹⁵. La descripción de los patrones argumentales nos ha permitido identificar todos los usos de una misma forma léxica. Nos gustaría señalar que muy a menudo estos diferentes usos no se describen en los artículos del diccionario (por ejemplo, en forma de diferentes significados).

Estas propiedades configuracionales se complementan con propiedades semánticas, porque al caracterizar el patrón de argumentos también describimos su naturaleza semántica. Las propiedades semánticas son el núcleo del modelo de clases de objetos, y esto es claramente lo que distingue a este modelo de descripción del lenguaje del léxico-gramatical. Así, cada uso se asocia a una clase sintáctico-semántica, a un tipo semántico (o rasgo semántico) y a un dominio de especialización. La especificación del dominio de especialización resultó especialmente interesante, ya que la nomenclatura de nuestro diccionario electrónico incluye un gran número de términos pertenecientes al discurso especializado.

Por último, nos interesaba observar propiedades morfológicas especialmente relevantes en determinados ámbitos especializados. Los nombres de muchas ayudas suelen ir asociados a una forma acrónima o a una forma que ha sufrido un truncamiento y se utiliza con mucha frecuencia (por ejemplo, *RMI* o *allocs*). Por ello, es necesario indicar estas formas asociadas y hacer que constituyan entradas independientes en nuestro diccionario electrónico.

14. Análisis de corpus : Descripción de los predicados nominales de <ayuda financiera>.

La aplicación sistemática de la rejilla de análisis a los sustantivos predicativos de <ayuda financiera> identificados condujo a la constitución de subclases semánticas. Una subclase está formada por diferentes sustantivos predicativos sinónimos o cuasisinónimos con propiedades homogéneas. Obtuvimos así una primera clasificación en once clases semánticas. En segundo lugar, cuando describimos la actualización verbal de estas subclases semánticas y las comparamos, nos dimos cuenta de que algunas subclases tenían el mismo patrón de argumentos y la misma combinatoria verbal. Así pues, no nos

15

pareció pertinente hacer esta distinción en once subclases, sobre todo porque la clase <ayuda financiera> ya constituye una subclase de la hiperclase <ayuda>.

El interés del método de clases objeto reside en el carácter económico de estas clases: las unidades léxicas con el mismo funcionamiento sintáctico-semántico se agrupan en la misma clase para obtener una descripción de la lengua económica y eficaz desde el punto de vista del tratamiento automático.

Finalmente, obtuvimos una clasificación en tres subclases: <ayuda financiera: subsidio>, <ayuda financiera: donación y <ayuda financiera: desgravación>. Las describimos brevemente a continuación.

<ayuda financiera: subsidio

Esta subclase incluye sustantivos simples (*pension, allocation, indemnité, subsidie*, etc.), sustantivos compuestos (*revenu de solidarité active, aide personnalisée au logement, rente d'incapacité permanente*, etc.) y acrónimos (*APL, PPE, RSA*, etc.). Como acabamos de ver en los ejemplos anteriores, incluye sustantivos del lenguaje general, del discurso especializado (sobre todo en el ámbito de la protección social y el derecho laboral) y sustantivos pertenecientes a un registro coloquial del lenguaje (por ejemplo, *allocs*).

Estos nombres se caracterizan en primer lugar por sus propiedades configuracionales, es decir, su patrón de argumentos, tal como se indica a continuación:

N0	N1	N2	N3
<hum, coll_hum>	to <hum, coll_hum >	de <importe>	para Vinf

Tabla 1: Esquema de los argumentos de <aide financière : allocation> en francés

N0	N1	N2	N3
<hum, coll_hum>	a <hum, coll_hum >	de <importe>	para Vinf

Tabla 2: Esquema de argumentos para <ayuda financiera: asignación> en español

La primera posición argumental está ocupada por un sustantivo caracterizado por el rasgo humano (*hum*) o un colectivo humano (*coll_hum*, por ejemplo: un estado, una institución, una organización, etc.). La posición N1 corresponde al primer complemento, es decir, la persona (*hum*) o el colectivo (*coll_hum*) que recibe la ayuda. Al tratarse de una ayuda económica, el argumento N2 difiere del resto de subclases de <ayuda>. Así, esta posición argumental está ocupada por grupos preposicionales que denotan una

cantidad de dinero. Por último, la posición N3, aunque se actualiza con menos frecuencia, está ocupada por un complemento preposicional que contiene un verbo en infinitivo. Este N3 indica el objetivo o la finalidad de la ayuda en cuestión.

Nos gustaría señalar que algunos nombres de <ayuda financiera: asignaciones> incluyen un quinto argumento, porque son pagos de ayuda que se extienden a lo largo de un periodo de tiempo específico. Por tanto, su patrón de argumentos sería el siguiente:

N0	N1	N2	N3	N4
<hum, coll_hum>	to <hum, coll_hum >	de <importe>	para, durante, en <período_de _tiempo> (<período_de _tiempo>)	para Vinf

Tabla 3: Esquema de argumentos para <aide financière : allocation> con 5 argumentos en francés

N0	N1	N2	N3	N4
<hum, coll_hum>	a <hum, coll_hum >	de <importe>	de, durante, por <period_tps>	para Vinf

Tabla 4: Esquema de argumentos para <ayuda financiera: asignación> con 5 argumentos en español

En cuanto a la combinatoria con verbos de apoyo y verbos predicativos apropiados, presentamos en este resumen simplemente una ficha descriptiva:

Clase	FR	ES
Entrées	Aide à la recherche d'emploi, allocation, allocs, appui, APL, bourse Erasmus, pension, prestation d'accueil du jeune enfant, rente d'accident du travail, secours viager, soutien, subvention d'investissement...	Ayuda al alquiler, beca postdoctoral, pensión alimenticia, prestación de orfandad por violencia contra la mujer, renta, subsido agrícola, subsidio extraordinario por desempleo...

Schéma d'arguments	N0:Nhum, Ncoll /N1: à Nhum, Ncoll/N2: de <montant>/N3: pour Vinf	N0:Nhum, Ncoll /N1: a Nhum, Ncoll/N2: de <montant>/N3: para Vinf
Verbes supports appropriés	Accorder, allouer, apporter, attribuer, consentir (FIN., DR.), donner, fournir, octroyer, offrir, procurer.	Aportar, asignar, brindar, conceder, dar, ofrecer, otorgar, prestar, proporcionar, suministrar.
Verbes supports passifs (N1N0)	Avoir, avoir droit à, bénéficier de, disposer de, encaisser, jouir de, palper (fam.), percevoir, profiter de, recevoir, toucher (fam.).	Contar con, disfrutar de, gozar de, percibir, recibir, tener, tener derecho a.
Verbes supports aspectuels inchoatifs	Décrocher, gagner, obtenir, trouver.	Conseguir, ganar, lograr, obtener.
Verbes supports aspectuels progressifs	Maintenir. N1N0 : conserver, garder, maintenir.	Prolongar, prorrogar, renovar. N1N0 : conservar, mantener.
Verbes supports aspectuels terminatifs	Annuler, couper, geler, retirer, supprimer, suspendre. N1N0 : perdre.	Anular, cancelar, congelar, eliminar, quitar, retirar, suprimir, suspender N1N0 : perder
Verbes supports aspectuels itératifs	Débloquer, reconduire, renouveler.	Desbloquear, reconducir, renovar.
Verbes supports aspectuels intensifs à valeur progressive-augmentative	Améliorer, augmenter, majorer.	Ampliar, aumentar, elevar, incrementar, mejorar, subir.

Verbes supports aspectuels intensifs à valeur progressive-atténuative	Diminuer, limiter, réduire.	Bajar, congelar, cortar, disminuir, rebajar, recortar, reducir
Verbes prédicatifs appropriés de <demande>	Demander, réclamer, solliciter.	Demandar, exigir, pedir, reclamar, solicitar.
Verbes prédicatifs appropriés de <mise en place>	Introduire, mettre en place, voter.	Aprobar, establecer, instituir, poner en marcha.
Verbes prédicatifs appropriés de <refus>	Refuser. N1N0: refuser, renoncer à.	Denegar, negar, rechazar. N1N0: rechazar, renunciar a
Verbes prédicatifs appropriés de <versement>	Verser.	Abonar, ingresar, pagar

<Ayuda financiera: donación

La classe <ayuda financiera: donación> contiene un número muy limitado de nombres: *don* y *donation* en francés, y *donación* y *donativo* en español. No hemos considerado nombres del ámbito especializado del derecho como *donation-partage* (fr.) *libéralité* (fr.), *legs* (fr.), *liberalidad* (es.) o *legado* (es.) porque no son nombres de <aide>, sino de <héritage>.

Los nombres pertenecientes a esta clase se caracterizan por el siguiente patrón de argumentos:

N0	N1	N2	N3
<hum, coll_hum>	to <hum, coll_hum >	de <importe>	para Vinf

Tabla 5: Esquema de argumentos para <aide financière : donation> en francés

N0	N1	N2	N3
<hum, coll_hum>	a <hum, coll_hum >	de <importe>	para Vinf

Tabla 6: Esquema de argumentos para <ayuda financiera: donación> en español

La primera posición argumental está ocupada por un sustantivo caracterizado por el rasgo humano (*hum*) o un colectivo humano (*coll_hum*, por ejemplo: una asociación, una institución, una organización, etc.). La posición N1 corresponde al primer complemento, es decir, la persona (*hum*) o el colectivo (*coll_hum*) que se beneficia de esta <donación>. El N1 aparece a menudo en forma de grupo preposicional que comienza por la preposición *à* (fr.) o *a* (es.), pero también hemos encontrado frases preposicionales como *à faveur de*. El argumento N2 corresponde a una frase preposicional que denota una cantidad de dinero. Por último, la posición N3 está ocupada por una frase preposicional que contiene un verbo en infinitivo. En el diagrama argumental, indicamos que la frase preposicional comienza con la preposición *pour* (fr.) o *para* (es.). Como vemos en el ejemplo 119, el N3 también puede empezar por *dans le but de*.

En cuanto a la combinatoria con verbos de apoyo y verbos predicativos apropiados, presentamos en este resumen simplemente una ficha descriptiva:

Clase	FRANÇAIS	ESPAGNOL
Entrées	Don, donation.	Donación, donativo.
Schéma d'arguments	N0:Nhum, Ncoll /N1: à Nhum, Ncoll/N2: de <montant>/N3: pour Vinf	N0:Nhum, Ncoll /N1: a Nhum, Ncoll/N2: de <montant>/N3: para Vinf
Verbes supports appropriés	Accorder, consentir, effectuer, faire, octroyer, offrir, réaliser.	Aportar, brindar, conceder, dar, efectuar, entregar, enviar, hacer, hacer entrega de, ofrecer, otorgar, realizar.
Verbes supports passifs (N1N0)	Bénéficiaire de, percevoir, recevoir.	Recibir, percibir, recaudar, recoger.
Verbes supports aspectuels inchoatifs	Obtenir.	Obtener, conseguir.

Verbes supports aspectuels itératifs	Distribuer, répartir. N1NO : accumuler, collecter, cumuler, rassembler, récolter, recueillir.	Distribuir, repartir. N1NO : acumular, colectar, recolectar, reunir.
Verbes prédicatifs appropriés de <demande>	Demander, solliciter.	Pedir, solicitar.
Verbes prédicatifs appropriés de <refus>	Refuser. N1NO : refuser.	Rechazar.
Verbes prédicatifs appropriés de <versement>	Verser.	Abonar, ingresar, pagar.

<Ayuda financiera: desgravación>

A continuación describiremos los sustantivos predicativos de la clase <aide financière : dégrèvement>. Esta clase contiene unidades monoléxicas (por ejemplo, *ristourne*, *abattement* o *exonération* en francés; *rebaja*, *reducción* o *exoneración* en español) y unidades poliléxicas (por ejemplo, *allègement fiscal*, *escompte de caisse* o *exonération d'impôt* en francés; *reducción fiscal*, *descuento por pronto pago* o *exención tributaria* en español). Como podemos ver en los ejemplos que acabamos de dar, esta clase contiene muchos sustantivos que pertenecen al discurso especializado del derecho fiscal (por ejemplo, *allègement fiscal* / *reducción fiscal*; *exonération d'impôt* / *exención tributaria*) y del comercio (por ejemplo, *escompte de caisse* / *descuento por pronto pago*), y también contiene sustantivos del discurso general como *rabais*, *réduction* o *remise* en francés, y *rebaja*, *reducción* o *descuento* en español.

Estos nombres se caracterizaron en primer lugar por su patrón argumental:

N0	N1	N2	N3
<coll_hum>	de <importe>, #de N	en N, #para N	a <hum, coll_hum >, #para <hum, coll_hum >, #a favor de <hum, coll_hum >

Cuadro 7: Esquema de argumentos para <aide financière : allègement> en français

N0	N1	N2	N3
<coll_hum>	de <importe>, #de N	en N, #por N, #para N	a <hum, coll_hum >

Cuadro 7: Esquema de argumentos para <aide financière : allègement> en español

Copiamos aquí la ficha resumen sobre la combinatoria verbal de estos sustantivos:

	FRANÇAIS	ESPAGNOL
Entrées	Abattement, allègement, escompte de caisse, exonération fiscale, ristourne...	Reducción, reducción fiscal, descuento por pronto pago, exención fiscal, rebaja...
Schéma d'arguments	N0: Ncoll / N1: de <montant>, #de N / N2: sur N, #pour N / N3 : à Nhum, Ncoll, #pour Nhum, Ncoll, #en faveur de Nhum, Ncoll	N0: Ncoll / N1: de <montant>, #de N / N2: en N, #por N, #para N / N3 : a Nhum, Ncoll,
Verbes supports appropriés	Accorder, consentir, faire, octroyer, offrir, pratiquer.	Aplicar, conceder, hacer, ofrecer, otorgar, practicar, proponer, realizar.
Verbes supports passifs (N1N0)	Avoir, avoir droit à, bénéficier de, profiter de.	Beneficiarse de, disfrutar de, gozar de, percibir, recibir, tener.
Verbes supports aspectuels inchoatifs	Obtenir.	Conseguir, lograr, obtener.
Verbes supports aspectuels progressifs	Pérenniser, prolonger.	Mantener, prorrogar.
Verbes supports aspectuels terminatifs	Supprimer. Conversif : perdre	Anular, eliminar, quitar, retirar et suprimir. Conversif : perder
Verbes supports aspectuels itératifs	Cumuler.	Acumular.
Verbes supports aspectuels intensifs à valeur progressive-augmentative	Accroître, augmenter.	Aumentar, incrementar.
Verbes supports aspectuels intensifs à valeur progressive-atténuate	Limiter, plafonner.	Limitar, rebajar, recortar.

Verbes pr�dicatifs appropri�s de <demande>	Demander, exiger, r�clamer, solliciter.	Exigir, pedir, solicitar, reclamar
Verbes pr�dicatifs appropri�s de <mise en place>	Instaurer, instituer.	Establecer, implantar.
Verbes pr�dicatifs appropri�s de <refus>	Refuser. Conversif : refuser, renoncer �.	Denegar, negar, rechazar. Conversif : renunciar a.

15. Conclusiones generales

El objetivo de nuestro trabajo ha sido realizar un estudio ling stico de los predicados nominales de <ayuda financiera> aplicando la teor a de clases de objetos. Nuestro objetivo ha sido obtener una descripci n fina y exhaustiva que pudiera ser utilizada por el aut mata sin generar ruido o silencio. Al final, las clases sint ctico-sem nticas no han resultado ser tan finas como esper bamos y la descripci n del l xico de clases obtenida no puede ser tan exhaustiva como hubi ramos deseado.

La primera clasificaci n en clases y subclases sem nticas se bas  en la combinatoria de sustantivos predicativos con verbos aspectuales de apoyo. De este modo, creamos numerosas subclases. Por ejemplo, en el caso de la clase <ayuda financiera: subsidio>, hab amos distinguido <ayuda financiera: subsidio  nico>, <ayuda financiera: subsidio peri dico>, <ayuda financiera: subsidio vitalicio>. Esto se debe a que los sustantivos de <ayuda financiera: subsidio puntual> no se combinan con verbos incoativos, progresivos o terminativos de apoyo aspectual. Sin embargo, hemos observado que en Espa a una determinada ayuda puede ser puntual o peri dica, dependiendo de la regi n. Este es el caso de la *beca Erasmus*, ya que en algunas regiones se asigna en un  nico pago, en otras regiones el pago se realiza en dos plazos, y en otras el pago se realiza mensualmente. Tambi n hemos observado que la forma de pago puede cambiar a lo largo de los a os. Es el caso, por ejemplo, del t rmino *bonificaci n*. La forma de pago de una determinada bonificaci n puede cambiar a lo largo de los a os, pasando de la clase <ayuda financiera: desgravaci n> (una reducci n de un tipo a un porcentaje determinado sobre una cantidad a pagar) a la clase <ayuda financiera: subsidio> (una cantidad fija de dinero que se devuelve o se asigna al beneficiario). As  pues, si quer amos garantizar cierta estabilidad

y validez a nuestra clasificación, la distinción en distintas subclases no podía ser tan fina. Por lo tanto, tuvimos que eliminar estas clases basándonos en diferencias aspectuales.

En segundo lugar, fue necesario dejar de lado muchos verbos soporte, ya que de otro modo sería imposible agrupar diferentes unidades léxicas en la misma clase. Tomemos como ejemplo los sustantivos *allègement* y *dégrèvement*, pertenecientes a la clase <aide financière : allègement>. El *Petit Robert 2024 (versión digital)* los presenta como sinónimos. De hecho, mostramos a continuación la definición de *allègement* en este diccionario para demostrar que estas dos unidades léxicas están muy próximas desde el punto de vista semántico:

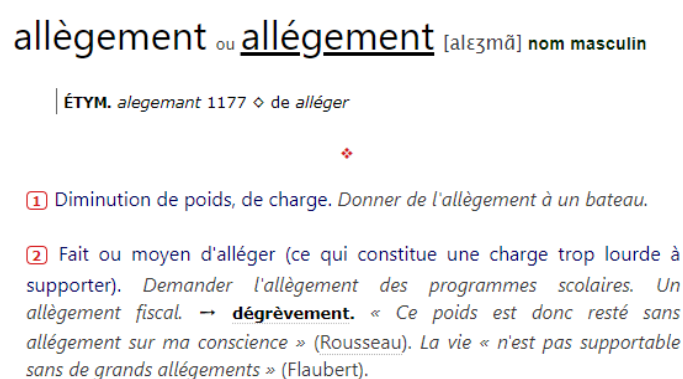


Figura 1 : Definición de aligeramiento (extracto del *Petit Robert, 2024 - versión digital*)

Sin embargo, estas dos unidades léxicas no tienen exactamente las mismas propiedades combinatorias. Por ejemplo, *dégrèvement* selecciona los verbos *imputer* y *accroitre*, pero *allègement* no se combina con estos verbos.

Por poner otro ejemplo, en el caso de la clase <ayuda financiera: subsidio>, identificamos un total de 106 verbos (verbos de apoyo y verbos predicativos apropiados) que se combinan con algunos sustantivos de esta clase. Sólo 54 de estos verbos se combinan con todos los sustantivos de la clase.

En tercer lugar, tras haber realizado este estudio, nos preguntamos qué sentido tendrían las clases que contienen muy pocos sustantivos, como es el caso de la clase <aide financière : donation>. En efecto, esta clase sólo contiene dos sustantivos en español (*donativo* y *donación*) y otros dos en francés (*don* y *donation*). En el proceso de desarrollo de las clases, esta clase contenía inicialmente otros sinónimos como *libéralité*, *aumône*, *charité*, *étrennes*, *générosité* y *largesse* en francés, y *liberalidad*, *limosna*, *caridad*, *aguinaldo* y *dádiva* en español. En francés, partimos de un conjunto de 34 verbos (13

verbos de apoyo apropiados, 3 verbos de apoyo pasivos, 13 verbos de apoyo aspectuales y 5 verbos predicativos apropiados); en español, probamos 44 verbos (11 verbos de apoyo apropiados, 7 verbos de apoyo pasivos, 11 verbos de apoyo aspectuales y 12 verbos predicativos apropiados). Comprobamos la combinatoria de estos verbos con todos los sustantivos susceptibles de pertenecer a esta clase. Este análisis dio como resultado que en francés sólo los verbos de apoyo *accord* y *accepter* son seleccionados por todos los sustantivos de la clase, y en español, los verbos *conceder*, *dar*, *entregar* y *recibir*. Observamos que *don* y *donation* (en francés), y *donativo* y *donación* (en español) tienen un patrón combinatorio muy similar, mientras que los demás sustantivos muestran un patrón combinatorio muy irregular. Por tanto, fue necesario excluir de esta clase todos aquellos otros sustantivos que, siendo sinónimos o cuasi-sinónimos, presentan un patrón combinatorio muy dispar. También ha sido imposible crear otra clase específica para estos sustantivos, que hemos eliminado de la clase <ayuda financiera: donación>.

Sin embargo, hemos obtenido un recurso que enriquece la descripción de la combinatoria de los sustantivos <ayuda financiera>. La ventaja de las clases semánticas sintácticas reside en su eficacia para el tratamiento automático y en el hecho de que proporcionan una descripción económica del lenguaje. Como las clases son generalizaciones, ahorran esfuerzo y permiten obtener más rápidamente una descripción del léxico de una lengua (a pesar de ser una tarea a largo plazo). En otras palabras, la tarea de describir una lengua utilizando el método de las clases de objetos requiere mucho tiempo, pero es factible.

En resumen, el método de las clases de objetos proporciona una descripción rica, eficaz y económica del lenguaje, pero no nos parece exhaustivo, al menos con los parámetros propuestos por este enfoque. Por ello, dejamos este campo abierto para futuras investigaciones. Consideramos que, para ser exhaustivos, sería necesario describir cada unidad léxica individualmente.

Las investigaciones que hemos llevado a cabo nos han hecho reflexionar profundamente sobre las restricciones de selección y sobre la importancia del léxico en la propia configuración de la lengua. El léxico es una parte esencial de la lengua. Hemos observado que las reglas gramaticales y sintácticas deben formularse en términos de lexías. En otras palabras, el léxico de una lengua tiene prioridad sobre su gramática.

