



Université de Montréal

**Du terme prédicatif au cadre sémantique :  
méthodologie de compilation d'une ressource terminologique pour  
les termes arabes de l'informatique**

par Nizar Ghazzawi

Département de linguistique et de traduction  
Faculté des arts et des sciences

Thèse présentée  
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)  
en Traduction  
option Terminologie

mai 2016

© Nizar Ghazzawi, 2016

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée :

**Du terme prédicatif au cadre sémantique :  
méthodologie de compilation d'une ressource  
terminologique pour les termes arabes de l'informatique**

présentée par :

Nizar Ghazzawi

a été évaluée par un jury composé des personnes suivantes :

Patrick Drouin, Président du jury

Marie-Claude L'Homme, Directrice de recherche

Fatiha Sadat, Co-directrice

Jean Quirion, Membre du jury

Mercè Lorente, Examinatrice externe

## Résumé

La description des termes dans les ressources terminologiques traditionnelles se limite à certaines informations, comme le terme (principalement nominal), sa définition et son équivalent dans une langue étrangère. Cette description donne rarement d'autres informations qui peuvent être très utiles pour l'utilisateur, surtout s'il consulte les ressources dans le but d'approfondir ses connaissances dans un domaine de spécialité, maîtriser la rédaction professionnelle ou trouver des contextes où le terme recherché est réalisé. Les informations pouvant être utiles dans ce sens comprennent la description de la structure actancielle des termes, des contextes provenant de sources authentiques et l'inclusion d'autres parties du discours comme les verbes.

Les verbes et les noms déverbaux, ou les unités terminologiques prédicatives (UTP), souvent ignorés par la terminologie classique, revêtent une grande importance lorsqu'il s'agit d'exprimer une action, un processus ou un événement. Or, la description de ces unités nécessite un modèle de description terminologique qui rend compte de leurs particularités. Un certain nombre de terminologues (Condamines 1993, Mathieu-Colas 2002, Gross et Mathieu-Colas 2001 et L'Homme 2012, 2015) ont d'ailleurs proposé des modèles de description basés sur différents cadres théoriques.

Notre recherche consiste à proposer une méthodologie de description terminologique des UTP de la langue arabe, notamment l'*arabe standard moderne* (ASM), selon la théorie de la Sémantique des cadres (*Frame Semantics*) de Fillmore (1976, 1977, 1982, 1985) et son application, le projet FrameNet (Ruppenhofer et al. 2010). Le domaine de spécialité qui nous intéresse est l'informatique. Dans notre recherche, nous nous appuyons sur un corpus recueilli du web et nous nous inspirons d'une ressource terminologique existante, le DiCoInfo (L'Homme 2008), pour compiler notre propre ressource. Nos objectifs se résument comme suit. Premièrement, nous souhaitons jeter les premières bases d'une version en ASM de cette ressource. Cette version a ses propres particularités : 1) nous visons des unités bien spécifiques, à savoir les UTP verbales et déverbaux; 2) la méthodologie développée pour la compilation du DiCoInfo original devra être adaptée pour prendre en compte une langue sémitique. Par la suite, nous souhaitons créer une version en cadres de cette ressource, où nous

regroupons les UTP dans des cadres sémantiques, en nous inspirant du modèle de FrameNet. À cette ressource, nous ajoutons les UTP anglaises et françaises, puisque cette partie du travail a une portée multilingue.

La méthodologie consiste à extraire automatiquement les unités terminologiques verbales et nominales (UTV et UTN), comme *Ham~ala* (حمل) (télécharger) et *taHmiyl* (تحميل) (téléchargement). Pour ce faire, nous avons adapté un extracteur automatique existant, TermoStat (Drouin 2004). Ensuite, à l'aide des critères de validation terminologique (L'Homme 2004), nous validons le statut terminologique d'une partie des candidats. Après la validation, nous procédons à la création de fiches terminologiques, à l'aide d'un éditeur XML, pour chaque UTV et UTN retenue. Ces fiches comprennent certains éléments comme la structure actancielle des UTP et jusqu'à vingt contextes annotés. La dernière étape consiste à créer des cadres sémantiques à partir des UTP de l'ASM. Nous associons également des UTP anglaises et françaises en fonction des cadres créés. Cette association a mené à la création d'une ressource terminologique appelée « DiCoInfo : A Framed Version ». Dans cette ressource, les UTP qui partagent les mêmes propriétés sémantiques et structures actancielles sont regroupées dans des cadres sémantiques. Par exemple, le cadre sémantique *Product\_development* regroupe des UTP comme *Taw~ara* (طور) (développer), *to develop* et *développer*.

À la suite de ces étapes, nous avons obtenu un total de 106 UTP ASM compilées dans la version en ASM du DiCoInfo et 57 cadres sémantiques associés à ces unités dans la version en cadres du DiCoInfo. Notre recherche montre que l'ASM peut être décrite avec la méthodologie que nous avons mise au point.

**Mots-clés** : unité terminologique prédicative, cadre sémantique, FrameNet, informatique, Arabe Standard Moderne, Frame Semantics

## **Abstract**

The description of terms in traditional terminological resources is limited to certain details, such as the term (which is usually a noun), its definition, and its equivalent. This description seldom takes into account other details, which can be of high importance for the users, especially if they consult resources to enhance their knowledge of the domain, to improve professional writing, or to find contexts where the term is realized. The information that might be useful includes the description of the actantial structure of the terms, contexts from authentic resources and the inclusion of other parts of speech such as verbs.

Verbs and deverbal nouns, or predicative terminological units (PTUs), which are often ignored by traditional terminology, are of great importance especially for expressing actions, processes or events. But the description of these units requires a model of terminological description that takes into account their special features. Some terminologists (Condamines 1993, Mathieu-Colas 2002, Gross et Mathieu-Colas 2001 et L'Homme 2012, 2015) proposed description models based on different theoretical frameworks.

Our research consists of proposing a methodology of terminological description of PTUs of the Arabic language, in particular Modern Standard Arabic (MSA), according to the theory of Frame Semantics of Fillmore (1976, 1977, 1982, 1985) and its application, the FrameNet project (Ruppenhofer et al. 2010). The specialized domain in which we are interested is computing. In our research, we compiled a corpus that we collected from online material and we based our method on an existing online terminological resource called the DiCoInfo (L'Homme 2008) in our pursuit to compile our own. Our objectives are the following. First, we will lay the foundations of an MSA version of the aforementioned resource. This version has its own features: 1) we target specific units, namely verbal and deverbal PTUs; 2) the developed methodology for the compilation of the original DiCoInfo should be adapted to take into account a Semitic language. Afterwards, we will create a framed version of this resource. In this version, we organize the PTUs in semantic frames according to the model of FrameNet. Since this frame version has a multilingual dimension, we add English and French PTUs to the resource.

Our methodology consists of automatically extracting the verbal and nominal terminological units (VTUs and NTUs) such as *Ham~ala* (حمل) (download). To do this, we integrated the MSA to an existing automatic extractor, TermoStat (Drouin 2004). Then, with the help of terminological validation criteria, we validate the terminological status of the candidates. After the validation, we create terminological files with an XML editor for each VTU and NTU. These files contain elements, such as the actantial structure of the PTUs and up to 20 annotated contexts. The last step consists of creating semantic frames from the MSA PTUs. We also associate English and French PTUs to the created frames. This association resulted in the creation of a second terminological resource called “DiCoInfo: A Framed Version”. In this resource, the PTUs that share the same semantic features and actantial structures are organized in semantic frames. For example, the semantic frame *Product\_development* groups PTUs such as *Taw~ara* (طور) (develop), *to develop* and *développer*.

As a result of our methodology, we obtained a total of 106 PTUs in MSA compiled in the MSA version of DiCoInfo and 57 semantic frames associated to these units in the framed version. Our research shows that the MSA can be described using the methodology that we set up.

**Keywords:** predicative terminological units, semantic frames, FrameNet, computing, Modern Standard Arabic, Frame semantics

# Table des matières

<b>RÉSUMÉ.....</b>	<b>I</b>
<b>ABSTRACT .....</b>	<b>III</b>
<b>TABLE DES MATIÈRES.....</b>	<b>V</b>
<b>LISTE DES TABLEAUX.....</b>	<b>XII</b>
<b>LISTE DES FIGURES .....</b>	<b>XIV</b>
<b>LISTE DES SIGLES ET DES ABRÉVIATIONS.....</b>	<b>XVIII</b>
<b>CONVENTIONS D'ÉCRITURE .....</b>	<b>XVIII</b>
<b>TRANSLITÉRATION (BUCKWALTER 2016).....</b>	<b>XIX</b>
<b>REMERCIEMENTS.....</b>	<b>XXI</b>
<b>1. INTRODUCTION .....</b>	<b>1</b>
1.1 Problématique de la recherche .....	1
1.2 Objectifs de la recherche.....	2
1.3 Hypothèses de la recherche.....	3
1.4 Structure de la thèse .....	4
<b>2. QUELQUES QUESTIONS RELATIVES À L'ARABE STANDARD MODERNE ET AU TRAVAIL TERMINOLOGIQUE DANS CETTE LANGUE .....</b>	<b>6</b>
2.1 Importance de la langue arabe .....	6
2.2 Pourquoi l'ASM.....	8
2.3 Morphologie de l'ASM.....	10
2.3.1 La question de la translitération.....	13
2.4 Travail terminologique dans l'espace arabophone .....	14
2.4.1 Particularités et problèmes de la terminologie technique en ASM.....	14
2.4.2 Création terminologique des termes techniques : méthodes.....	18
2.5 Défis et problèmes .....	20
<b>3. ÉTAT DE L'ART .....</b>	<b>22</b>
3.1 Partie 1 : Les travaux sur les unités terminologiques prédicatives en terminologie .....	22



3.1.1 Classes d'objets.....	24
3.1.1.1 Les classes d'objets en langue de spécialité .....	28
3.1.1.2 Le domaine médical.....	29
3.1.1.3 Le domaine de l'informatique.....	30
3.1.1.4 Le domaine juridique .....	32
3.1.2 Étude des propriétés syntaxico-sémantiques du verbe par l'établissement de classes .....	33
3.1.2.1 Traitement des verbes .....	34
3.1.2.2 Application au domaine bancaire.....	37
3.1.3 L'optique lexico-sémantique .....	39
3.1.4 Frame-based terminology (FBT) .....	47
3.1.4.1 Les verbes spécialisés dans la FBT.....	51
3.1.5 Autres classifications des verbes spécialisés .....	55
3.1.5.1 Les classifications .....	56
3.1.5.2 Critères de validation .....	58
3.1.6 De multiples typologies de classification des verbes : Gross, Condamines, Mathieu- Colas et Lorente .....	62
3.1.7 Autres travaux sur les unités prédicatives.....	64
3.1.7.1 Les unités phraséologiques en discours spécialisé.....	64
3.1.7.2 Les verbes du domaine juridique .....	67
3.1.7.3 Simplification des textes médicaux au moyen des verbes .....	68
3.1.8 Importance du verbe spécialisé et notre apport personnel .....	70
3.2 Partie 2 : les ressources lexicales de l'arabe standard moderne.....	73
3.2.1 Les ressources générales .....	74
3.2.1.1 Propbank .....	74
3.2.1.2 The Arabic Propbank (APB).....	76
3.2.1.3 WordNet.....	80
3.2.1.4 The Arabic WordNet (AWN) .....	82
3.2.1.5 VerbNet.....	89
3.2.1.6 Arabic VerbNet (AVN).....	90

3.2.1.7 Les contraintes liées à la constitution d'une ressource lexicale de l'arabe standard moderne .....	95
3.2.2 Les ressources spécialisées .....	97
3.2.2.1 WIPO Pearl .....	97
3.2.2.2 ARABTERM (ArT) .....	101
3.2.2.3 Unified Medical Dictionary (UMD) .....	105
3.2.3 Les travaux ontologiques sur la langue arabe .....	107
3.2.3.1 Ontologie basée sur les verbes .....	108
3.2.3.2 Ontologie pour l'extraction du contenu des blogues .....	109
3.2.3.3 Arabic Ontology Project .....	112
3.3 Vers une nouvelle méthodologie .....	115
<b><u>4. CADRE THÉORIQUE .....</u></b>	<b>117</b>
4.1 La révolution cognitive et les travaux sur les verbes .....	117
4.2 De la syntaxe structurale à la Sémantique des cadres .....	118
4.3 Sémantique des cadres .....	123
4.3.1 Transfer_scenario frame .....	126
4.4 Applications de la Sémantique des cadres .....	129
4.4.1 La langue générale : FrameNet .....	129
4.4.1.1 Le cadre .....	130
4.4.1.2 L'unité lexicale .....	132
4.4.1.3 La valence .....	134
4.4.1.4 L'annotation .....	135
4.4.1.5 Relations entre les cadres .....	136
4.4.2 FrameNet dans d'autres langues .....	137
4.4.2.1 Verbes de mouvement du point de vue de la Sémantique des cadres .....	138
4.4.2.2 Arabic FrameNet (AFN) .....	142
4.4.3 Application de la Sémantique des cadres et le FN en langues de spécialité .....	145
4.4.3.1 The Quranic FrameNet Project .....	145
4.4.3.2 BioFrameNet .....	149
4.4.3.3 Framed DiCoEnviro : A framed version of DiCoEnviro .....	151

4.4.4 Choix de cadre théorique et ce qui distingue notre projet .....	157
<b>5. MÉTHODOLOGIE.....</b>	<b>159</b>
5.1 Corpus.....	159
5.1.1 Particularité et importance d'un corpus arabe de spécialité.....	161
5.1.2 Le sujet.....	163
5.1.3 Langue du corpus.....	164
5.1.4 Provenance géographique .....	166
5.1.5 Type de textes .....	166
5.1.6 Taille du corpus.....	169
5.1.7 Les textes du corpus.....	169
5.1.8 Le nombre de textes .....	170
5.1.9 Le format du corpus.....	170
5.1.10 Paternité .....	171
5.1.11 Date de publication .....	172
5.2 Extraction de candidats termes .....	173
5.2.1 TermoStat et la méthode de comparaison.....	174
5.2.2 Intégration de l'ASM à TermoStat .....	175
5.2.3 Le système de MADA de Habash.....	176
5.2.4 Étapes pour bâtir un système d'extraction.....	178
5.2.4.1 Corpus général .....	178
5.2.4.2 Étiquetage des corpus .....	179
5.2.4.3 Analyse morphosyntaxique.....	182
5.2.4.4 AraComLex versus MADA .....	184
5.2.4.5 Gestion de spécificité.....	186
5.2.4.6 Résultats de l'extraction avec TermoStat .....	187
5.2.4.7 Traitement des résultats : filtrage.....	189
5.2.4.8 Résultats de filtrage pour les UTN.....	189
5.2.4.9 Résultats pour les UTV.....	189
5.2.4.10 Analyse des résultats.....	190
5.3 Validation des candidats UTP.....	193

5.3.1 UTP candidates écartées .....	201
5.3.2 Distinction des sens.....	201
5.4 Choix de contextes pour les UTP retenues .....	204
5.5 Annotation des contextes des UTP dans XML .....	209
5.5.1 Vocable .....	210
5.5.2 Lexie .....	211
5.5.3 Information grammaticale.....	211
5.5.4 Structure actancielle.....	212
5.5.5 Réalisations .....	213
5.5.6 Contextes.....	213
5.5.6.1 Participant .....	215
5.5.6.2 Étiquettes des rôles .....	215
5.5.6.3 Lexie .....	216
5.5.7 Équivalent .....	217
5.5.8 Particularités de l'annotation des contextes en langue arabe.....	218
5.6 Identification des cadres .....	220
5.6.1 Différences avec FrameNet.....	221
5.6.2 Association des cadres aux UTP .....	222
5.6.3 Critères de sélection des cadres .....	224
5.6.4 Relations entre cadres .....	227
5.6.5 Encodage des cadres .....	229
5.6.5.1 Frame .....	229
5.6.5.2 Core Frame Element (FE).....	230
5.6.5.3 Actants .....	231
5.6.5.4 Lexies.....	231
5.6.5.5 Définition .....	232
5.6.5.6 Exemple .....	232
5.6.5.7 Relations .....	233
<b>6. RÉSULTATS.....</b>	<b>234</b>
6.1 Les unités terminologiques prédicatives (UTP) de l'ASM.....	234

6.1.1 Les UTV.....	234
6.1.1.1 Les UTV validées .....	235
6.1.2 Les UTN.....	239
6.1.3 Particularités .....	243
6.1.4 Correspondance entre les UTV et les UTN .....	245
6.1.4.1 Extraction .....	245
6.1.4.2 Proportion des UTN par rapport aux UTV .....	247
6.1.4.3 Correspondance avec les UTP françaises et anglaises.....	248
6.2 DiCoInfo : la version ASM.....	251
6.3 Cadres identifiés.....	256
6.3.1 Les UTP et les cadres correspondants .....	257
6.3.2 Relations entre UTP dans les cadres.....	263
6.3.2.1 Relation entre forme verbale et nominale.....	263
6.3.2.2 Relations de synonymie et d'antonymie.....	264
6.3.2.3 Relation de correspondance .....	267
6.3.3 Relations entre cadres : les scénarios conceptuels.....	268
6.3.3.1 Scénarios conceptuels et relations entre cadres .....	269
6.4 DiCoInfo : A Framed Version .....	275
6.4.1 Page d'accueil .....	275
6.4.2 Recherche dans Framed Version .....	276
6.4.3 Affichage des résultats.....	277
<b>CONCLUSION.....</b>	<b>282</b>
<b>BIBLIOGRAPHIE .....</b>	<b>I</b>
<b>ANNEXE 1 : SOURCES DU CORPUS .....</b>	<b>I</b>
<b>ANNEXE 2 : LISTE PARTIELLE DES UTN CANDIDATES.....</b>	<b>III</b>
<b>ANNEXE 3 : LISTE PARTIELLE DES UTV CANDIDATES.....</b>	<b>IV</b>
<b>ANNEXE 4 : UTV RETENUES .....</b>	<b>V</b>
<b>ANNEXE 5 : UTN RETENUES.....</b>	<b>VII</b>
<b>ANNEXE 6 : RÔLES SÉMANTIQUES.....</b>	<b>X</b>

**ANNEXE 7 : CORRESPONDANCE ENTRE LES UTV ET LES UTN..... XIII**

## Liste des tableaux

Tableau 1 : Verbes appartenant au Transfer_scenario.....	127
Tableau 2 : Sujets abordés dans les textes du corpus.....	164
Tableau 3 : Correspondance des quatre paramètres communicatifs aux sources de notre corpus .....	168
Tableau 4 : Résumé des principales caractéristiques de notre corpus .....	173
Tableau 5 : Détails du corpus général.....	179
Tableau 6 : Étiquette VB de Stanford et ses variations .....	182
Tableau 7 : Comparaison entre MADA et AraComLex .....	186
Tableau 8 : Unités éliminées.....	189
Tableau 9 : Extraction des UTV sans seuil de spécificité.....	190
Tableau 10 : Verbes support .....	192
Tableau 11 : Tableau récapitulatif des UTP analysées .....	200
Tableau 12 : Distinction sémantique de l'UTN <i>tavbyt</i> (تثبيت) (installation).....	202
Tableau 13 : Partie des rôles sémantiques utilisés .....	216
Tableau 14 : Relations entre cadres utilisés .....	228
Tableau 15 : Comparaison entre les UTV les plus fréquentes et leurs dérivés nominaux.....	238
Tableau 16 : Liens paradigmatiques observés entre les UTV.....	239
Tableau 17 : Relations paradigmatiques pour les UTN .....	242
Tableau 18 : Dix premiers candidats retenus par TermoStat.....	246
Tableau 19 : Exemples d'UTV et d'UTN correspondantes.....	248
Tableau 20 : UTV sans correspondance dans DiCoInfo.....	249
Tableau 21 : UTN sans correspondance dans DiCoInfo.....	250
Tableau 22 : Quatre groupes de cadres.....	258
Tableau 23 : Répartition des UTP de l'ASM dans les cadres correspondants.....	259
Tableau 24 : Répartition des UTP des trois langues dans les cadres correspondants.....	260
Tableau 25 : Description du cadre Compatibility .....	261
Tableau 26 : Description du cadre Program_creation .....	262
Tableau 27 : UTV et UTN dans cadre Seeking .....	263

Tableau 28 : Relation d'antonymie dans le cadre Manipulating_components.....	265
Tableau 29 : Relation de synonymie dans le cadre Transfer .....	266
Tableau 30 : Relation de correspondance entre UTP dans le cadre Installing_resume .....	268
Tableau 31 : Scénario Computer_setup .....	271
Tableau 32 : Placing_scenario.....	273
Tableau 33 : Activity_scenario.....	274



## Liste des figures

Figure 1 : Dix langues les plus utilisées sur la Toile (Internet World Stats 2016) .....	7
Figure 2 : Morphologie d'un mot en ASM .....	11
Figure 3 : Catégorisation des arguments (Gross et Mathieu-Colas 2001) .....	29
Figure 4 : Prédicat d'action et sa structure argumentale (Gross et Mathieu-Colas 2001) .....	30
Figure 5 : Verbes spécialisés en informatique (Mathieu-Colas 2002 : 62).....	31
Figure 6 : Verbes de langue générale avec un sens spécialisé (Mathieu-Colas 2002 :62) .....	32
Figure 7 : Classe 1 de la classification du domaine bancaire (Condamines 1993 : 51).....	38
Figure 8 : Réseau lexico-terminologique du verbe <i>programmer</i> (L'Homme 2012 : 100) .....	43
Figure 9 : Le verbe <i>télécharger</i> dans le DiCoInfo (2015) .....	45
Figure 10 : Verbe avec deux entrées dans le DiCoInfo (2015).....	46
Figure 11 : Nom prédicatif <i>téléchargement</i> dans le DiCoInfo (2015).....	47
Figure 12 : Catégories des verbes dans la FBT (Araúz et al. 2012, cité dans Faber 2014 : 22)50	
Figure 13 : <i>ENVIRONMENTAL EVENT</i> dans la FBT (Buendia et al. 2014 : 62) .....	50
Figure 14 : Verbes de sous-domaine Cause Motion (Sánchez et Buendia 2012 :559).....	52
Figure 15 : Modèle du verbe <i>éjecter</i> d'après la FBT (Sánchez et Buendia 2012 : 557).....	52
Figure 16 : Activation des domaines lexicaux dans Extreme Event (Buendia et al. 2014 :72) 53	
Figure 17 : Résultat d'une interrogation d'un modèle dans EcoLexicon (2016).....	54
Figure 18 : Classification des verbes en discours (Lorente 2007, cité dans Pimentel (2012 : 93)).....	61
Figure 19 : Analyse du cadre <i>to_begin_to_exist</i> (Buendia 2013: 385) .....	66
Figure 20 : Cadre <i>to_begin_to_exist</i> dans EcoLexicon (EcoLexicon 2016) .....	67
Figure 21 : Cadre <i>Compliance</i> dans le domaine juridique (Pimentel 2011 : 159).....	68
Figure 22 : Fiche du verbe <i>to accept</i> dans Propbank (2015) .....	76
Figure 23 : Frameset du verbe <i>qabila</i> (قبِلَ) (accepter) dans APB (Arabic Propbank 2015) .....	79
Figure 24 : Statistiques de la version 3.1 de WordNet (2015).....	81
Figure 25 : Quatre types de relations d'implication selon Fellbaum (Fellbaum 1990 : 292) ....	82
Figure 26 : Statistiques de la version 2.0 d'AWN .....	83

Figure 27 : Correspondance entre AWN et WordNet via SUMO (Elkateb et al. 2006a et Black et al. 2006b).	85
Figure 28 : Arabic WordNet Browser	86
Figure 29 : Requête dans AWN	87
Figure 30 : Analyse d'Aramorph dans le navigateur d'AWN	88
Figure 31 : Classe <i>create</i> dans VerbNet (VerbNet 2015)	90
Figure 32 : Deux groupes obtenus à la suite de la traduction de la classe <i>Manner_speaking</i> dans VerbNet (Mousser 2010)	91
Figure 33 : Class <i>Change of State</i> dans AVN (Mousser 2011: 357)	93
Figure 34 : Sibling class de <i>Change of State</i> dans AVN (Mousser 2011 : 357)	93
Figure 35 : Cadre <i>nasaxa</i> (نسخ) ( <i>copier</i> ) dans AVN (AVN 2015)	94
Figure 36 : Recherche d'un terme arabe dans WIPO (2015)	99
Figure 37 : Contexte pour un terme dans WIPO (2015)	100
Figure 38 : Représentation graphique des concepts dans WIPO (2015)	101
Figure 39 : Extension <i>Yamli</i> dans ArT (2015)	104
Figure 40 : Partie des résultats pour le terme <i>barmajap</i> (برمجة) ( <i>programmation</i> ) dans ArT (2015)	104
Figure 41 : Définition d'un terme dans ArT (2015)	105
Figure 42 : Interface web d'UMD (2015)	106
Figure 43 : Verbe <i>yuEAlij</i> (يعالج) ( <i>traiter</i> ) dans UMD (2015)	107
Figure 44 : Sous-graphes représentant des classes dans une ontologie de l'informatique (Al-Safadi et al. 2011)	111
Figure 45 : Partie de niveau élevé de ArabicOntology (Jarrar 2011)	114
Figure 46 : Structuration de la phrase selon Tesnière (Tesnière 1959 : 11)	119
Figure 47 : Structuration linéaire de la phrase d'après la grammaire traditionnelle (ibid. : 3)	119
Figure 48 : Structure profonde composée de plusieurs cas	122
Figure 49 : Cadre <i>Placing</i> dans FrameNet (FrameNet 2015)	131
Figure 50 : Liste des unités lexicales du cadre <i>Placing</i> (FrameNet 2015)	132
Figure 51 : Tableau récapitulatif des FE et leur réalisation syntaxique pour le verbe <i>file</i> (FrameNet 2015)	133

Figure 52 : Exemple d'une liste d'occurrences d'une UL (FrameNet 2015) .....	133
Figure 53 : Modèles de valence pour l'UL <i>file</i> (FrameNet 2015) .....	135
Figure 54 : Phrases annotées pour l'UL <i>file</i> (FrameNet 2015).....	136
Figure 55 : Relations entre cadres dans FrameNet (FrameNet 2015).....	137
Figure 56 : Traits sémantiques et fréquence du FE Self-mover pour les verbes anglais et arabes (Alshehri 2014 : 10).....	139
Figure 57 : Fréquence des autres FE du cadre Self_motion (Alshehri 2014 : 10) .....	140
Figure 58 : Traits sémantiques du FE Self_mover pour tous les verbes analysés (Alshehri 2014 : 12).....	141
Figure 59 : Fréquence des non-core FE pour tous les verbes analysés (Alshehri 2014 : 13) .	141
Figure 60 : Structure de base de l'AFN (Ghneim et al. 2009).....	143
Figure 61 : Interface utilisateur de l'application d'AFN (Ghneim et al. 2009) .....	144
Figure 62 : Classification des verbes en fonction de leurs formes (Atwel et Sharaf 2009)....	147
Figure 63 : Concordances du verbe <i>eat</i> dans le Coran (Atwel et Sharaf 2009).....	148
Figure 64 : Cadres dans BioFrameNet avec des contextes annotés (Harabagiu et Bejan 2010) .....	150
Figure 65 : Entrée du verbe <i>recycler</i> dans le DiCoEnviro (2015) .....	151
Figure 66 : Correspondance entre l'UT <i>emit</i> et l'UL <i>emit</i> du cadre Emitting dans FrameNet .....	153
Figure 67 : Affichage du cadre Emitting dans DiCoEnviro <i>framed version</i> (2015) .....	155
Figure 68 : Cycle_of_existence_scenario dans DiCoEnviro Framed Version (2015).....	156
Figure 69 : Partie d'une phrase analysée avec MADA.....	176
Figure 70 : Texte mal converti, problème d'encodage .....	180
Figure 71 : Texte mal converti, problème de lisibilité.....	181
Figure 72 : Sortie d'un texte étiqueté avec Stanford.....	181
Figure 73 : Phrase analysée avec <i>AraComLex</i> .....	183
Figure 74 : Résultats de l'extraction des UTN.....	187
Figure 75 : Résultats de l'extraction des UTV.....	188
Figure 76 : Résultats pour la requête <i>taSaf~aHa</i> (تصفح) (naviguer) dans Almaany.....	194
Figure 77 : Structure d'encodage dans le DiCoInfo .....	210

Figure 78 : Vocabulaire dans une fiche dans XML .....	210
Figure 79 : Structure actancielle dans XML .....	212
Figure 80 : Réalisations linguistiques des actants dans XML .....	213
Figure 81 : Contexte annoté dans XML.....	214
Figure 82 : Balise <equivalences> dans XML.....	217
Figure 83 : Comparaison des termes dans DiCoInfo avec les UL dans FrameNet.....	221
Figure 84 : Structure d'encodage pour les cadres.....	229
Figure 85 : Correspondance entre actants et core-FE sur la page de comparaison.....	230
Figure 86 : Encodage d'exemple pour le cadre Processing_materials .....	232
Figure 87 : Interface du DiCoInfo version ASM.....	252
Figure 88 : Liste alphabétique des UTP en caractères arabes.....	253
Figure 89 : Listes alphabétiques des équivalents français et des formes translittérées .....	253
Figure 90 : Résultat d'une requête dans le DiCoInfo, version ASM.....	254
Figure 91 : Contextes annotés pour l'UTV <i>Taw~ara</i> (طور) (développer).....	255
Figure 92 : Tableau récapitulatif des réalisations linguistiques des actants .....	256
Figure 93 : Page d'accueil de DiCoInfo : A Framed Version.....	276
Figure 94 : Recherche dans DiCoInfo : A Framed Version.....	276
Figure 95 : Vue d'ensemble du résultat de recherche dans Framed Version.....	277
Figure 96 : Définition d'un cadre dans Framed Version .....	278
Figure 97 : Exemples dans Framed Version.....	278
Figure 98 : Participant 1 et Participant 2 dans la Framed Version .....	279
Figure 99 : Listes des UTP faisant partie d'un cadre.....	280
Figure 100 : Relations entre cadres dans la Framed Version.....	280

## Liste des sigles et des abréviations

- UT : Unité terminologique
- UTP : Unité terminologique prédicative
- UTV : Unité terminologique verbale
- UTN : Unité terminologique nominale
- ASM : Arabe Standard Moderne

## Conventions d'écriture

- Terme ASM : (terme translittéré en italique) (forme arabe) (équivalent français), *Ham~ala* (حمل) (télécharger)
- Terme français et anglais : italique, *télécharger*, *download*
- Actants : première lettre en majuscule, Agent
- Nom de cadre : en Century Gothic, Placing

## Translitération (Buckwalter 2016)

ء	‘	ذ	*	م	m
آ		ر	r	ن	n
أ	A/>	ز	z	ه	h
ؤ	&/w	س	s	و	w
إ	i/<	ش	\$	ي	y
ئ	}	ص	S	ى	Y
ا	A	ض	D	َ	a
ب	B	ط	T	ُ	N
ة	P	ظ	Z	ِ	i
ت	T	ع	E	ِ	K
ث	V	غ	g	ُ	u
ج	J	ف	f	ُ	N
ح	H	ق	q	َ	0
خ	X	ك	k	َ	~
د	D	ل	l		

*À la mémoire de : Dr. Zouheir Ghazzawi et Prof. Edward W. Saïd*

*et...à toi, Noam A. Chomsky!*

## Remerciements

Je voudrais tout d'abord remercier grandement ma directrice de thèse, Marie-Claude L'Homme, pour son aide, sa patience, ses efforts, son écoute et son encouragement. Je ne peux qu'exprimer toute mon admiration pour son expertise dans le domaine de la terminologie et pour ses qualités personnelles. Marie-Claude...je te remercie du fond du cœur !

Je remercie également, ma co-directrice, Fatiha Sadat, pour le temps qu'elle m'a consacré, ses conseils et ses commentaires.

J'adresse un remerciement spécial à Benoit Robichaud pour tous les efforts qu'il a mis dans la réalisation informatique des deux ressources. Je sais que la tâche n'était pas du tout facile !

Je remercie Patrick Drouin pour le temps qu'il m'a accordé, pour son aide et pour son sens de l'humour (Vive TermoStat!).

Je remercie mes professeurs, mes collègues et mes amis à l'OLST.

Je remercie Najat Rahman, professeure au département de littératures, pour son aide généreuse.

Finalement, je remercie ma douce moitié, Hala El-Khatib, de m'avoir partagé le chemin. Je remercie les Ghazzawis et les El-Khatibs, notamment, ma mère, ma belle-mère, mes frères, mes belles-sœurs et mes beaux-frères. Je remercie ma tante, Hala Ghazzawi, d'avoir toujours été à mes côtés.



# 1. Introduction

Dans la présente recherche, nous présentons une description des termes prédicatifs de l'informatique selon les principes méthodologiques de la théorie de la Sémantique des cadres (Fillmore 1976, 1977, 1982, 1985). L'objectif de notre étude est de proposer une méthodologie de compilation d'une ressource terminologique pour les unités terminologiques prédicatives (UTP) verbales et déverbales de l'ASM (*arabe standard moderne*). Cette méthodologie s'inspire d'une ressource terminologique multilingue existante appelée le DiCoInfo (L'Homme 2008). De plus, nous proposons une modélisation des UTP arabes<sup>1</sup> selon le projet de FrameNet (Ruppenhofer et al. 2010). La langue visée par notre recherche est la langue arabe, plus précisément l'*arabe standard moderne* (ASM).

## 1.1 Problématique de la recherche

Le contenu spécialisé en langue arabe sur la Toile est de plus en plus important. La facilité d'accès aux services web et les possibilités qu'offrent ces derniers ont permis aux internautes de créer des sites, des blogues et des forums où ils peuvent s'exprimer sur différents sujets et publier leurs travaux. De plus, les réseaux sociaux et la presse électronique ont contribué à la diffusion de ce contenu spécialisé, surtout avec la présence d'infrastructures adaptées à cette fin. Ce développement est accompagné également d'un nombre croissant de ressources lexicales, générales et spécialisées, conçues de différentes façons et accessibles gratuitement. En outre, l'ASM de spécialité fait désormais partie de plusieurs dictionnaires et ressources lexicales et terminologiques multilingues. Nous mentionnons à titre d'exemple, UNTERM (<https://unterm.un.org/UNTERM/portal/welcome>), FAOTERM (<http://www.fao.org/faoterm/en/>) et UNESCOTERM (<http://termweb.unesco.org/>).

Cependant, dans ces ressources, la description proposée pour ces termes a encore des limites. La forme nominale est dominante et les autres parties du discours sont peu présentes. Par exemple, dans UNTERM, nous trouvons *imprimante* mais pas *imprimer* ou *impression* (en

---

<sup>1</sup> Nous avons également associé quelques UTP anglaises et françaises équivalentes à ces cadres à titre indicatif et afin de faire quelques observations comparatives entre ces langues et la langue arabe.

tant qu'unité prédicative). Cette domination de la forme nominale relève de la conception classique de la terminologie. Certains spécialistes en terminologie (Rey 1979 et Sager 1995, cités dans L'Homme 2012) associent les concepts aux noms, tandis que les autres parties du discours, comme les verbes et les adjectifs, ne sont considérés comme des unités terminologiques que lorsqu'elles sont associées à une forme nominale.

En outre, les termes simples sont souvent moins nombreux dans ces ressources que les termes complexes. Par exemple, on a tendance à retenir *barmajap kA}niyyap Al-tawaj~uh* (برمجة كائنية التوجه) (programmation orientée objet) plutôt que *barmajap* (برمجة) (programmation) et *TibAEap vulAviyyap Al-AbEAd* (طباعة ثلاثية الأبعاد) (impression 3D) plutôt que *TibAEap* (طباعة) (impression). Cette préférence pour les termes complexes pose deux problèmes. Premièrement, les termes simples déverbaux et verbaux sont ignorés. Pour nous, cela constitue une lacune dans les ressources spécialisées. Deuxièmement, un terme simple prédicatif permet de mieux renseigner sur les actions dans les domaines de spécialité, surtout pour un domaine comme l'informatique.

Finalement, les ressources actuelles ne montrent pas la réalisation concrète des termes qu'elles décrivent. Ces ressources sont rarement basées sur corpus et, par conséquent, elles ne proposent pas de contextes pour enrichir leur description. Le contexte est un élément important, puisque « toutes les informations nécessaires à la reconnaissance du sens d'un mot sont définies par ses environnements stricts » (Gross 2008 : 112). Dans le cas des UTP, le contexte a une importance capitale, surtout dans le cas de l'ASM, car le taux de polysémie des unités prédicatives est élevé (voir section 2.3).

## 1.2 Objectifs de la recherche

Pour notre recherche, nous avons fixé les objectifs suivants. Notre premier objectif est la création d'une version en ASM du DiCoInfo. Cette version a la particularité d'être consacrée uniquement aux UTP, à savoir les unités terminologiques verbales (UTV) et les unités terminologiques nominales (UTN). Pour ce faire, nous nous inspirons de la méthodologie adoptée dans le DiCoInfo, dans le but de voir comment cette méthodologie peut convenir à une langue typologiquement différente. L'ASM dispose d'un système morphologique riche et complexe (voir section 2.3). Pour cette raison, intégrer cette langue à

une telle ressource peut constituer un défi. De plus, la méthodologie adoptée nous permet de valider la nature spécialisée des verbes et des noms déverbaux et de donner une description de ces unités dans le domaine de l'informatique.

Notre deuxième objectif est le regroupement des UTP dans des cadres sémantiques. Ce regroupement se fait selon les principes de la théorie de la Sémantique des cadres de Fillmore (1976, 1977, 1982, 1985). Les UTP analysées font l'objet d'une réflexion mettant en évidence leurs propriétés sémantiques. Par conséquent, nous validons notre réflexion en appliquant des critères qui nous permettent de faire ce regroupement.

Notre troisième objectif est la création d'une ressource terminologique comprenant les cadres qui renferment les UTP décrites. Dans cette ressource, nous incluons certaines unités anglaises et françaises équivalentes du DiCoInfo. Cette ressource s'appelle « DiCoInfo : A Framed Version ». Elle est réalisée en fonction du modèle FrameNet (Ruppenhofer et al. 2010). Dans la découverte des cadres sémantiques, nous nous basons, en partie, sur les cadres existants dans FrameNet que nous adaptons au domaine de l'informatique.

La raison pour laquelle nous adoptons ce modèle pour nos UTP est qu'il semble bien adapté pour rendre compte d'une modélisation des concepts dans un domaine de spécialité. Le modèle de FrameNet a été adapté à plusieurs domaines (section 4.4.3). Or, notre recherche constitue la première tentative de l'adapter au domaine de l'informatique.

### **1.3 Hypothèses de la recherche**

Dans notre recherche, nous examinons trois hypothèses.

*Hypothèse 1 : L'ASM s'adapte à une méthodologie de travail terminologique particulière*

L'ASM fait déjà partie de plusieurs dictionnaires et ressources terminologiques multilingues. Cependant, la question qui se pose dans notre cas est l'intégration de l'ASM à une ressource dotée d'une méthodologie de travail terminographique bien précise. La ressource visée par notre hypothèse est le « Dictionnaire fondamental de l'informatique et de l'Internet », le DiCoInfo. Cette ressource est trilingue (anglais, français et espagnol) et est

compilée selon les principes de la Lexicologie explicative et combinatoire (Mel'čuk et al. 1995).

*Hypothèse 2 : L'ASM a des particularités qu'il faut prendre en compte*

La deuxième hypothèse concerne les particularités de l'ASM par rapport aux autres langues faisant partie du DiCoInfo. Étant donné que l'ASM est une langue sémitique, elle a des propriétés qu'on ne retrouve pas dans les langues européennes, comme l'anglais et le français. Ces différences peuvent se manifester dans l'existence de certains verbes ou noms déverbaux qui n'ont probablement pas d'équivalents dans les autres langues.

*Hypothèse 3 : L'ASM de spécialité a des points communs avec les autres langues de spécialité*

La troisième hypothèse concerne les points communs partagés par la terminologie de l'informatique en ASM et la terminologie exprimée par les langues comprises dans le DiCoInfo, à savoir l'anglais et le français. Cette question concerne la façon dont les concepts sont exprimés par les UTP. L'informatique est un domaine technique et, par conséquent, les actions, les événements et les processus dans le domaine sont probablement les mêmes dans toutes les langues.

## **1.4 Structure de la thèse**

Notre thèse comporte six chapitres. Dans ce premier chapitre, nous présentons notre travail de recherche en spécifiant la problématique, les hypothèses de la recherche et nos objectifs.

Dans chapitre 2, nous présentons la langue arabe, ses variantes, ses particularités et son système morphologique. Nous parlons également de la terminologie dans l'espace arabophone, surtout en ce qui concerne ses problèmes et les modes de création terminologique.

Le chapitre 3 est consacré à l'état de l'art. Nous présentons les premiers travaux portant sur les verbes en discours spécialisé et nous procédons jusqu'aux récentes tendances en terminologie, en ce qui concerne le traitement des unités prédicatives. Nous présentons également les ressources, générales et spécialisées, de l'ASM. De plus, nous tentons de situer notre travail et notre apport personnel par rapport aux travaux présentés.

Le chapitre 4 aborde le cadre théorique de notre recherche. Nous présentons la théorie de la Sémantique des cadres de Fillmore (1976, 1977, 1982, 1985). Ensuite, nous parlons de l'application de cette théorie à la langue générale, le projet de FrameNet (Ruppenhofer et al. 2010), et à la langue de spécialité. Nous nous attardons également sur l'application de cette théorie à l'ASM général et de spécialité en présentant les projets qui se sont inspirés du projet FrameNet.

Le chapitre 5 traite de notre méthodologie de travail. Dans ce chapitre, nous discutons de notre corpus, ses caractéristiques et sa provenance en montrant les difficultés que nous avons rencontrées lors du traitement de ces textes. Ensuite, nous abordons en détail les étapes suivies dans l'extraction des termes à partir du corpus ainsi que les outils utilisés. Nous présentons aussi les étapes de la validation terminologique de nos UTP extraites et les critères de la découverte des cadres sémantiques.

Le chapitre 6 présente les résultats obtenus à la suite de l'application de notre méthodologie. Nous présentons les termes étudiés et décrits et les cadres sémantiques découverts et créés. Nous présentons également nos deux ressources, la version en ASM du DiCoInfo (<http://olst.ling.umontreal.ca/dicoinfo/arabe/index.php?ui=fr>) et le DiCoInfo : A Framed Version (<http://olst.ling.umontreal.ca/dicoinfo/framed/>).

Le chapitre 7 renferme la conclusion. Dans ce chapitre, nous résumons notre recherche, nous validons nos hypothèses et nous présentons quelques pistes pouvant être exploitées dans l'avenir.

## 2. Quelques questions relatives à l'arabe standard moderne et au travail terminologique dans cette langue

Dans le présent chapitre, nous explorons certaines questions importantes en ce qui concerne notre choix de langue, ses propriétés et le travail terminologique dans le monde arabophone. Nous commençons d'abord par donner quelques informations sur l'ASM, son importance, son expansion et sa morphologie. Ensuite, nous expliquons pourquoi nous avons choisi cette langue.

### 2.1 Importance de la langue arabe

Owens (2013 : 1), dans *The Oxford Handbook of Arabic Linguistics*, constate que : « Arguably, for the linguist, Arabic is the most interesting language in the world ». La langue arabe<sup>2</sup> fait partie des langues les plus utilisées sur la Toile. Le site Internet World Stats<sup>3</sup> classe la langue arabe parmi les dix premières langues les plus utilisées sur Internet (figure 1). Elle vient en 4<sup>e</sup> position après l'anglais, le chinois et l'espagnol. Cette importance peut s'expliquer par différents facteurs :

- D'abord, la langue arabe est une langue parlée par presque 350 millions d'habitants;
- L'espace arabophone est un espace grandissant, sur le plan démographique;
- Ensuite, dans cet espace géographique, les arabophones<sup>4</sup> ont de plus en plus de facilité à accéder aux services Internet, ce qui leur a permis d'être de plus en plus présents sur la Toile;

---

<sup>2</sup> Comme souligné dans l'introduction, dans notre recherche nous utilisons « ASM » pour désigner la langue arabe standard moderne, qui est la variante faisant l'objet de notre travail. Quant à « arabe », il s'agit d'un terme générique que nous utilisons pour désigner toutes les variantes de la langue arabe, ce qui comprend les variantes régionales.

<sup>3</sup> <http://www.internetworldstats.com/stats7.htm>

<sup>4</sup> Par « arabophone », nous désignons toute personne ayant l'ASM comme langue de communication, d'éducation et d'échanges culturels.

- En outre, bien que le contenu arabe sur la Toile ne dépasse pas le seuil de 3 % du contenu mondial, il est en constante évolution grâce aux initiatives lancées par des gouvernements et des entreprises privées. Par exemple, la New York University Abu Dhabi (NYUAD), en partenariat avec d'autres universités, a lancé un projet pilote (Arabic Collections Online) qui a pour but la numérisation de plusieurs milliers de livres rédigés en ASM et les rendre disponibles en ligne<sup>5</sup>.

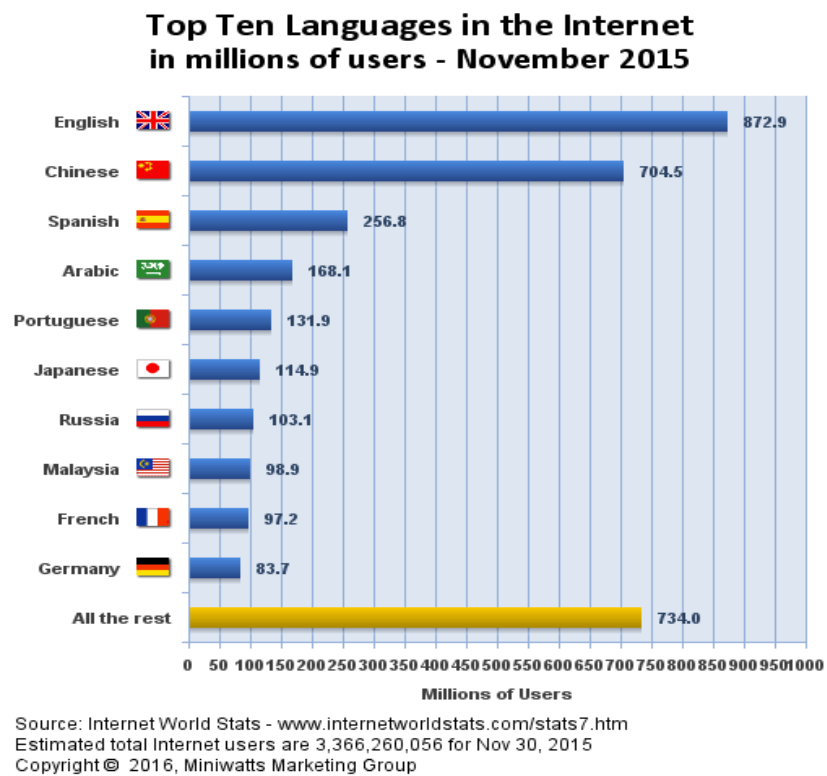
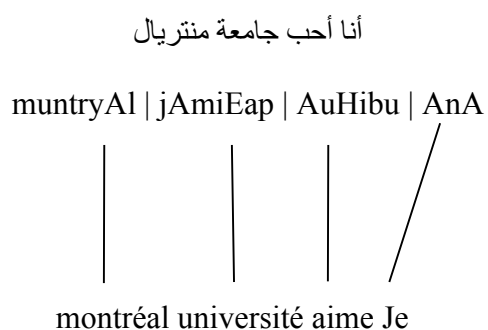


Figure 1 : Dix langues les plus utilisées sur la Toile (Internet World Stats 2016)

<sup>5</sup> <http://dlib.nyu.edu/aco/about/>

## 2.2 Pourquoi l'ASM

La langue arabe appartient à la famille des langues afro-asiatiques, plus particulièrement sa branche des langues sémitiques. Dans cette branche, on trouve d'autres langues comme l'araméen, l'hébreu, l'ougaritique et le phénicien. L'alphabet de la langue arabe dispose de 28 lettres et le système d'écriture va de droite à gauche, comme l'illustre l'exemple suivant pour la phrase : *J'aime l'université de Montréal.*



Ce système d'écriture a soulevé des problèmes pour la conception de nos ressources et pour l'annotation des contextes. Nous y reviendrons dans le chapitre abordant la méthodologie.

La langue arabe comprend un vaste ensemble de variantes. L'ASM est la seule variante formelle qui constitue la langue officielle de l'éducation et des médias dans le monde arabophone :

The Arabic language is a collection of multiple variants among which one particular variant has a special status as the formal written standard of the media, culture and education across the Arab World. The other variants are informal spoken dialects that are the media of communication for daily life (Habash 2010 : 1).

Nous pouvons distinguer trois variantes principales de la langue arabe. La première variante est l'arabe classique (Classical Arabic, CA), qui est la langue des textes anciens, notamment le Coran et la Bible. La deuxième variante est l'ASM (l'arabe standard moderne



ou en arabe (اللغة العربية الفصحى), qui est une variante basée sur la CA, mais plus moderne<sup>6</sup>. Il convient de signaler que ces deux variantes sont utilisées à l'écrit. La troisième variante<sup>7</sup> comprend les langues arabes parlées (colloquial arabic). Il s'agit ici des variations utilisées dans les régions où l'ASM est une langue officielle :

The Arabic dialects, in contrast, are the true native language forms. They are generally restricted in use for informal daily communication. They are not taught in schools or even standardized although there is a rich popular dialect culture of folktales, songs, movies, and TV shows. Dialects are primarily spoken not written. However, this is changing as more Arabs are gaining access to electronic media of communication such as emails and newsgroups. Arabic dialects are loosely related to Classical Arabic. They are the result of the interaction between different ancient dialects of Classical Arabic and other languages that existed in, neighbored and/or colonized what is today the Arab World (Habash 2010 : 1-2).

Par exemple, dans la région du Moyen-Orient (Jordanie, Liban, Syrie et Palestine), les variantes linguistiques sont un mélange de la langue arabe classique, de l'araméen, du turc, de l'anglais et du français. Une phrase comme *j'ai entré le mot de passe, mais ça n'a pas fonctionné*, peut se dire de différentes façons :

[1] دخلت الباسورد، بس ما مشي الحال

dax~alt AlbAswurd, bas mA mi\$y AIHA1

[2] أدخلت كلمة السر، لكن لم يمش الحال

Adxalt kalimap Alsir, lAkin lam yam\$y AIHA1

Le premier exemple représente la variante de la langue arabe du Moyen-Orient, tandis que le deuxième relève de l'ASM. Les différences constatées entre les deux exemples sont d'ordre lexical, morphologique et syntaxique. Sur le plan lexical, dans le contexte [1], nous remarquons la présence du terme anglais *password* (bAswurd, باسورد). En ASM (contexte [2]), l'équivalent est *kalimap sir* (كلمة سر) (littéralement, mot de passe). Nous remarquons également la présence de *bas* (بس) (mais), qui est un mot courant dans cette variante, mais pas dans l'ASM (en ASM c'est *lakin* (لكن)). Sur le plan morphologique, le verbe *Adxala* (أدخل) (entrer) présente une particularité. Dans le contexte [1], le verbe est morphologiquement différent, le A

---

<sup>6</sup> L'ASM représente une évolution de la CA, dans le sens où elle (ASM) ne comporte pas de mots obsolètes.

<sup>7</sup> Une autre variante est la langue arabe moyenne (Middle Arabic). Certains auteurs (Blau 1981) trouvent que cette variante est « the vernacular elements which penetrated mediaeval Arabic texts » (Blau 1981 : 87).

est manquant. Finalement, sur le plan syntaxique, la négation dans le contexte [1] ne correspond pas à la négation dans le contexte [2].

Nous avons fixé notre choix sur l'ASM. Cette variante de la langue arabe est particulièrement importante. L'ASM, en tant que variante langagière, se distingue des autres par ce qui suit :

- Le contenu sur le web est majoritairement en ASM. Par « contenu », nous désignons les livres électroniques, les sites web des journaux, les blogs attachés aux sites d'information, etc.;
- Il est difficile (voire impossible dans certains cas) de constituer un corpus dans une variante spécifique à une région, surtout lorsqu'il s'agit de textes spécialisés; et
- Les variantes régionales (les langues arabes parlées) ne sont pas standardisées : par conséquent, il existe moins de textes sur le web portant sur des sujets techniques ou scientifiques. Quant à la CA, son emploi est limité aux textes anciens, surtout les textes religieux et littéraires.

Par conséquent, l'ASM reste le choix le plus adéquat pour l'étude, l'analyse et la description de la terminologie d'un domaine de spécialité. Comme nous le verrons à la section 5.1, notre corpus est entièrement constitué de textes en ASM. Cependant, malgré les avantages que présente l'ASM par rapport aux autres variantes, l'élaboration d'un corpus spécialisé en ASM soulève des difficultés comme nous le montrons à la section 5.2.4.

## 2.3 Morphologie de l'ASM

La morphologie de l'ASM est considérée comme riche et complexe (Habash 2010, Soudi et al. 2007, Sadat et Mohamed 2013). La raison de cette complexité provient du fait que l'ASM dispose d'un système morphologique où les morphèmes sont concaténés à la racine. Ce système morphologique s'appelle *root-and-pattern* (d'après McCarthy 1981).

Dans ce système, la racine (*root*) est une séquence de consonnes organisées dans un ordre spécifique. Cet ensemble identifie le sens général d'un mot (l'idée d'écrire ou de lire,

par exemple). En ce qui concerne les autres informations comme le temps et la partie du discours, elles se trouvent dans le radical (*stem*) et elles sont représentées par les voyelles courtes et les syllabes (Testen 2016). Ces voyelles courtes et syllabes sont les caractéristiques distinctives d'un radical. Donc, dans le cas où un ensemble de radicaux partagent ces mêmes caractéristiques, on dit qu'ils ont le même schème (*pattern*). Ainsi, la racine *k-t-b* (ك-ت-ب) (l'idée d'écrire) peut avoir le radical *katab* (passé composé) (écrit), comme dans le verbe *katabtu* (كتبت) (j'ai écrit), d'où vient le schème **faEala** (فعل).

À titre d'exemple, nous examinons la forme *wasayaktubwnhA* (وسيكتبونها) qui veut dire *Et ils vont l'écrire*. Dans cette forme, nous pouvons trouver une conjonction (w), une particule désignant le futur (s), un verbe (yaktubwn) et un objet pronominal singulier féminin (hA). Cette « phrase », grammaticalement complète, peut être clairement décrite comme dans la figure 2 (Sadat et Mohamed 2013 : 622).

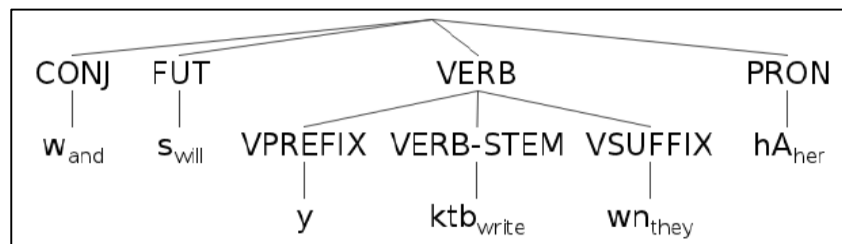


Figure 2 : Morphologie d'un mot en ASM

En outre, le système d'écriture en ASM n'inclut pas un certain nombre d'informations phonologiques (Sproat 2007 : viii). À titre d'exemple, un mot comme *drs* (درس) (il a étudié), prononcé *darasa*, n'inclut pas les voyelles courtes, ou les diacritiques. Également, dans un mot comme *ktb* (كتب) (faire écrire), que l'on peut translitérer comme *kat~aba*, la gémation est absente. C'est pour cette raison que Sproat (2007 : viii) constate que : « the advent of Unicode has failed to standardize Arabic encoding, so that in dealing with real texts, one has to be prepared to do a fair amount of low level normalization ».

Ce système morphologique complexe a un impact direct sur le traitement automatique de l'ASM. Souidi et al. (2007 : 3) constatent que :

The morphology of Arabic poses special challenges to computational natural language processing systems. The exceptional degree of ambiguity in the writing system, the rich morphology, and the highly complex word formation process of roots and patterns all contribute to making computational approaches to Arabic very challenging.

De ce fait, choisir, trouver ou concevoir un système de traitement automatique, que ce soit un extracteur des termes, un analyseur morphosyntaxique ou un étiqueteur, pour l'ASM de spécialité n'est pas une tâche facile. L'ASM dispose d'un système morphologique riche, et l'ambiguïté issue de cette richesse exige que l'on dispose d'un système très efficace. Par exemple, un mot comme *rafE* (رفع) (téléverser) est difficile à analyser par un système d'analyse morphologique. Cette forme est ambiguë, puisqu'elle peut signifier :

- [1] *rafaEa* (رفع) (éliminer);
- [2] *rafE* (رفع) (batir);
- [3] *rafE* (رفع) (promotion);
- [4] *rafuEa* (رفع) (affiner); et
- [5] *raf-aEa* (رفع) (promouvoir).

Donc, pour qu'un système d'analyse morphologique soit efficace, il doit rendre compte de cette ambiguïté, surtout lorsqu'il s'agit d'un contexte spécialisé. L'emploi d'une forme dans un contexte spécialisé accroît le taux de polysémie, puisque de nouveaux sens peuvent s'ajouter. Dans le chapitre consacré à la méthodologie, d'autres difficultés seront évoquées. Cependant, malgré la complexité de la tâche, la langue arabe ne cesse pas d'attirer l'attention des spécialistes, qu'ils soient linguistes, linguistes informaticiens ou terminologues. Des projets sur la langue arabe (ASM, CA ou parlée) sont en constant développement (voir la section 3.2).

L'un des aspects qui intéresse de plus en plus les spécialistes est la linguistique de corpus. Les linguistes arabophones (et arabisants) sont conscients de l'importance du corpus dans le travail sur la langue. Pour cette raison, nous constatons que la langue arabe fait maintenant partie de plusieurs systèmes de stockage et de traitement de la langue comme SketchEngine (Arts et al. 2014) pour l'ASM et ALC (Arabic Learner Corpus) (Alfaifi et al. 2014) pour l'apprentissage de l'ASM, The Living Arabic Project et Database of Arabic

Dialects pour la langue arabe parlée (les variantes régionales). Notre travail s'inscrit dans un mouvement cherchant à rendre disponibles des ressources linguistiques en ASM.

### 2.3.1 La question de la translittération

La question de la translittération est d'une importance capitale en raison du fait qu'un certain nombre d'outils de traitement automatique de l'arabe s'appuient sur des textes arabes translittérés. Dans cette sous-section, nous abordons cette question brièvement sous deux angles : le point de vue adopté dans notre travail et les outils de traitement.

Pour ce qui est de notre travail, les exemples sont tous présentés sous trois formes : forme translittérée, caractères arabes et traduction française. Il existe pour la langue arabe plusieurs systèmes de translittération. Nous mentionnons à titre d'exemple le système de translittération fondé sur la norme DIN-31635. Ce système comporte un nombre élevé de caractères spéciaux, comme  $\underline{t}$  pour  $\text{ث}$ ,  $\underline{g}$  pour  $\text{ج}$ ,  $\underline{h}$  pour  $\text{ح}$  et  $\underline{h}$  pour  $\text{خ}$ . Cela complique la rédaction, puisque ces caractères ne sont pas accessibles directement sur le clavier. Un autre système est celui adopté dans l'Encyclopédie de l'Islam (Bearman et al. 2002). Dans ce système, nous remarquons l'utilisation des digrammes. Ces digrammes indiquent qu'il s'agit d'une translittération d'une seule lettre en arabe et non pas deux, comme  $kh$  pour  $\text{خ}$  et  $sh$  pour  $\text{ش}$ .

Dans notre cas, nous avons adopté le système Buckwalter (voir la table de Translittération au début de la thèse). La raison pour ce choix est que ce système est largement utilisé dans la communauté linguistique.

En ce qui concerne les outils de traitement, les outils de traitement automatique de l'arabe se basent généralement sur la translittération Buckwalter (Habash et al. 2007). Habash et al. (ibid. : 15) constatent que :

The Buckwalter Arabic transliteration [2] is a transliteration system that follows the standard encoding choices made for representing Arabic characters for computers. The Buckwalter transliteration has been used in many publications in natural language processing and in resources developed at the Linguistic Data Consortium (LDC). The main advantages of the Buckwalter transliteration are that it is a strict transliteration (i.e., one-to-one) and that it is written in ASCII characters.

Ce système est particulièrement intéressant puisqu'il n'emploie pas des caractères spéciaux que l'on ne trouve pas sur les claviers (par exemple ڍ, ٲ et Š). Toutefois, lorsque ce système est utilisé, il faut reconvertir les textes en caractères arabes pour les lire. Dans la section 3.2, nous abordons la question de la translittération dans les ressources lexicales de l'ASM et nous montrons comment ces ressources utilisent la translittération Buckwalter.

## **2.4 Travail terminologique dans l'espace arabophone**

Dans cette section, nous nous attardons sur certains aspects du travail terminologique en ASM de spécialité. Nous commençons par présenter quelques particularités de la terminologie en ASM dans l'espace arabophone, puis nous abordons les problèmes de l'industrie terminologique.

### **2.4.1 Particularités et problèmes de la terminologie technique en ASM**

Lelubre (2008 : 1) constate que :

Les terminologies scientifiques et techniques en arabe sont toutes tributaires, d'une façon ou d'une autre, des terminologies établies en anglais ou en français, qui constituent les terminologies de référence. Ce cas n'a rien d'exceptionnel concernant bien d'autres langues, avec ceci que les terminologies sources ne sont pas forcément l'anglais ou le français, comme c'est le cas pour l'arabe.

Ce constat est particulièrement vrai pour les domaines techniques<sup>8</sup>, tels que l'informatique, qui est, d'ailleurs, un domaine par excellence où règne l'anglais. La terminologie en ASM des domaines techniques dépend largement de la traduction des termes anglais de ces domaines. Dans les dictionnaires, on peut constater que la nomenclature est toujours bilingue ou trilingue, avec l'anglais comme langue source. Cette situation de la terminologie technique en ASM n'est pas exemplaire. La terminologie doit faire face à de nombreux problèmes, surtout lorsque le terminologue ou le traducteur souhaitent traduire des textes techniques ou concevoir des glossaires pour des termes. Dans ce qui suit, nous attardons sur certains problèmes.

---

<sup>8</sup> Nous sommes en désaccord par rapport à certains domaines, comme la médecine, les mathématiques et la physique, où il existe un héritage arabophone important (Wulff 2004).

Khuwaileh (2010) fait remarquer que parmi les problèmes de la terminologie des domaines techniques (dans son cas, l'informatique) en ASM, se trouve celui de l'usage exclusif de l'anglais par les grandes entreprises, surtout lorsqu'il s'agit de la diffusion de la technologie. Bien que ces entreprises, comme Oracle et Google, aient des branches dans certains pays arabophones comme les Émirats Arabes Unis, la diffusion ne se fait qu'en anglais. Même si cette situation se révèle exacte, d'après nos observations, la globalisation a rendu un grand service à l'ASM<sup>9</sup>. De nos jours, la tendance est de diffuser dans le plus grand nombre de langues possible. Dans presque tous les sites Internet destinés au grand public, nous trouvons l'ASM comme option.

Le deuxième problème évoqué par l'auteur est un problème lié à la langue même. Il constate que (ibid. : 17) :

Renner (1998) in his article *Beyond Borders* argues that Arabic is not coping with new linguistic developments necessary for processing computer advancement and creating electronic computer repertoire of technical and semi-technical vocabulary. It follows from these circumstances that Arabic has become imbalanced when it comes to technical texts rendered from modern languages like English into Arabic. Consequently, rendering technical texts from English into Arabic will become extremely difficult because doing so requires not only cultural similarities between the source language and the target language, but also the two languages must be equally served in terms of technical vocabulary and structures.

Ce constat est appuyé par d'autres chercheurs (Darwish 1988 et Hamzé 2005). Par exemple, Darwish (1988 : 43-44) fait observer qu'après la période coloniale au Moyen-Orient et en Afrique du Nord, les élites de ces deux régions étaient formées principalement en Angleterre et en France. Par conséquent, les spécialistes et les experts communiquaient en français ou en anglais. De plus, ils avaient même tendance à voir le monde scientifique selon une perspective occidentale. Cela étant dit, la langue arabe n'a pas repris son rôle comme langue de sciences et de technologies, comme c'était le cas avant l'empire ottoman. Cette situation a eu un impact sur le statut de la langue dans les pays arabophones qui sont devenus des pays importateurs de terminologie.

---

<sup>9</sup> Puisque la variante utilisée dans ces sites est l'ASM, nous évitons d'employer le terme générique « arabe » pour parler de la langue faisant l'objet de cette question.

Dans son analyse de la traduction des affixes et des formants gréco-latins, Hamzé (2005 : 49-50) constate que le transfert de la terminologie de l'anglais et du français vers la langue arabe se fait souvent à une vitesse élevée et d'une façon désuète<sup>10</sup>. De plus, malgré le fait que des listes d'équivalents sont proposées pour faciliter la traduction de ces affixes, certains chercheurs suggèrent que le transfert se fasse de façon systématique. Par contre, l'auteur souligne le fait que la situation dans les autres langues n'est pas idéale. Il trouve que la création terminologique en français, par exemple, n'est pas systématique. Il donne l'exemple des termes *ophthamologue* et *ophthamologiste*, *neurologue* et *neurologiste* et *radiologue* et *radiologiste*. Pour cette raison, il conclut que :

However, on no account are these true difficulties and problems to mask the reality and to let us believe that vocabulary of exporting languages is a model to follow, not without the same defects. It is just simply that exporting languages' difficulties are double. These difficulties are related to the formation of vocabulary, even scientific vocabulary, as well as to issues of synonymy, polysemy and homonymy.

Dans la même veine, Saraireh (2001) retrace la source de l'incohérence de la terminologie technique en ASM. D'abord, l'auteur propose trois classes pour les termes issus de la traduction anglais-arabe. La première classe comprend les termes de la langue source (LS) qui ont des équivalents dans la langue cible (LC). La deuxième classe comprend des termes de la LS qui ont des équivalents partiels dans la LC. Finalement, la troisième classe comprend des termes de la LS qui n'ont pas d'équivalents dans la LC. Le problème peut être retracé dans la troisième classe. Dans cette classe le traducteur se réfère à plusieurs techniques afin d'établir la relation entre le concept et le signifiant. Par conséquent, le problème qui revient est celui de la standardisation dans la création de nouveaux termes dans le domaine technique. L'auteur (ibid. : 10) constate que trois facteurs posent problème :

First, the lagging behind of Arabic institutions in Arabicizing the ever-incoming terms, and, if dealt with, the poor circulation of the Arabic equivalents in proper time and ways (such as textbooks for students at different levels). Second, the lack of coordination among Arabic Language Academies in their efforts in dealing with foreign terminology [...]. Third, the wide gap between language planners and language users.

---

<sup>10</sup> Ici, il est question du manque réel de normalisation des termes.



Le troisième facteur est celui que nous considérons être le plus problématique. Dans les dictionnaires techniques, les termes sont standardisés, sauf qu'en pratique l'utilisateur (traducteur, terminologue, rédacteur professionnel, etc.) opte pour les termes qu'il estime être les plus faciles à utiliser. Par exemple, considérons le verbe *vab~ata* (يثبت) (installer). Dans le Dictionary of Information Technology Terms, issu de la Syrian Computer Society (2000), ce terme est présenté comme *yursy* (يرسي) (installer ou littéralement *monter*). D'après une recherche que nous avons effectuée à l'aide du moteur de recherche Google (et d'après les données de notre corpus, voir section 5.2.4.9 et 5.2.4.10), ce terme n'est pas en usage. Les seuls verbes que nous avons trouvés qui désignent ce processus en informatique sont *vab~ata* (يثبت) (installer) et *naS~aba* (نصب) (installer).

C'est pour cette raison que nous jugeons très important que le terminographe tienne compte de l'usage des termes. De plus, il importe lors de la création terminologique de prendre en considération les trois besoins de tout travail terminologique, à savoir la description, la transmission et la normalisation (Rey 1992 : 55). Nous sommes conscient du fait que la langue de spécialité est « plus qu'un registre (soutenu, courant ou familier) » (Kocourek 1991 : 40-41) et c'est un « système libre de ressources sur tous les plans de la langue qui possède plusieurs registres et plus que des caractéristiques lexicales ». À ce titre, nous soulignons l'importance du corpus dans tout travail terminologique et terminographique. Le corpus permet de mieux voir et comprendre l'usage et la circulation des termes dans une communauté linguistique donnée.

Malgré les problèmes discutés dans cette section, la situation actuelle de la langue arabe sur la scène terminologique a nettement progressé (même si elle est encore loin d'être idéale). Comme nous le verrons dans plusieurs sections de notre thèse, les ressources dédiées à l'ASM de spécialité sont de plus en plus nombreuses et les projets portant sur la langue arabe se multiplient. L'objectif commun est d'harmoniser et de systématiser la langue, y compris le transfert de la technologie. Certes, la standardisation et les difficultés de transfert sont toujours problématiques, même avec la production textuelle croissante en langue arabe sur la Toile.

## 2.4.2 Création terminologique des termes techniques : méthodes

La terminologie technique en ASM est formée selon deux procédés (Lelubre 2009), à savoir la néonymie de forme et la néonymie de sens. Ces deux procédés sont assujettis à quatre types de création (Elmgrab 2011 et Solimando 2015) : la dérivation, la périphrase, l'arabisation et le mot-valise.

La dérivation (ou en ASM *AlAi\$tiqAq* الإشتقاق) est considérée comme un procédé très important et efficace en ASM. On peut en distinguer deux types : la dérivation simple et la métathèse<sup>11</sup> (Elmgareb 2011 : 494). La dérivation simple est le procédé le plus utilisé dans la création lexicale et il s'appuie sur la racine (*root*) et le schème (*pattern*) du mot (voir 2.3). Par exemple, à partir de la racine T-b-E (طب-ع) (l'idée d'imprimer), on peut avoir toute une série de formes, comme *TabaEa* (طبع) (imprimer), *TAbiEap* (طابعة) (imprimante) et *TibAEap* (طباعة) (impression).

Quant à la périphrase (ou *AlAi\$tiqAq biAltArjamap* بالإشتقاق بالترجمة), il s'agit d'introduire de nouveaux termes en traduisant le sens des termes empruntés à d'autres langues. À titre d'exemple :

- [1] *clavier (keyboard)*: *lawHap mafAtiyH* (لوحة مفاتيح), où *key* est l'équivalent de *miftAH* (مفتاح) et *board* est l'équivalent de *lawHap* (لوحة);
- [2] *lecteur CD-ROM (CD-ROM drive)*: *saw~aqap AqrAS maDguwTap* (قارئ أقراص مضغوطة), où *drive* est l'équivalent de *saw~Aqap* (سواقة) et *CD* (Compact Disk) est l'équivalent de *qurS maDgwT* (قرص مضغوط);
- [3] *réinstaller (reinstall)*: *yuEyd ta\$giyl* (يعيد تشغيل). Puisque le préfixe *re-* n'a pas d'équivalent, la langue le remplace par le verbe *yuEyd* (يعيد), ce qui signifie *refaire*. Dans cet exemple, l'équivalent en ASM est un terme composé, *refaire l'installation*;
- [4] *puce (chipset)*: *ruqAqap* (رقاقة). Le terme *puce* existe dans la langue générale. Donc, en langue de spécialité, cette forme a acquis un sens spécialisé pour signifier *puce informatique*.

La création terminologique par périphrase est un procédé très utile. De plus, la périphrase permet de résoudre des problèmes importants lorsque le concept à traduire est

---

<sup>11</sup> L'auteur ne donne pas d'exemples pour la création par métathèse. Ce procédé relève d'une dérivation où un échange entre deux phonèmes donne lieu à un nouveau mot. Cette dérivation est moins courante dans une langue de spécialité, surtout en informatique.

inexistant dans la langue (et la culture) d'arrivée. Cependant, Elmgrab (2011 : 495-496) constate que la périphrase a des inconvénients. C'est un problème général en terminologie arabe des domaines techniques de trouver plusieurs termes désignant un seul concept. Comme expliqué à la section 2.4.1, ce problème est dû à l'absence d'une instance officielle en charge de la standardisation (et de la création) des terminologies dans tout l'espace arabophone. Lelubre (2008 : 1) relie ce problème au phénomène de la variation régionale et ajoute que :

[...] il n'existe pas à l'échelle du monde arabe une instance reconnue en matière de création terminologique; les terminologies sont établies pays par pays, ou groupes de pays par groupes de pays, dans un cadre officiel - comme c'est souvent le cas pour l'enseignement secondaire, où l'enseignement scientifique et technique se fait en arabe pour la plupart des pays. Une autre raison en est le statut même de l'arabe scientifique et technique comme langue de communication, qui semble, hormis l'enseignement primaire et secondaire, et en tout cas pour le domaine de la physique, ne trouve sa place que dans la littérature de vulgarisation. Littérature où l'on peut s'attendre à ce que les créations personnelles soient bien moins contrôlées que dans un cadre de communication scientifique de niveau plus élevé, où la terminologie est régulée par les instances scientifiques elles-mêmes, comme c'est le cas en physique pour l'anglais et le français - ... et où elle fait d'ailleurs loi pour le discours vulgarisé.

Le troisième type de création terminologique est l'arabisation (ou *AltaEryb* التعريب). Par *arabisation*, on entend « the adaptation of non-Arabic terms to Arabic by applying the rules of the phonological and sometimes morphological systems of the language to the terms » (Elmgrab 2011 : 496). Au sujet de l'arabisation, Stetkevych (1970: 56) constate que : « one of the most important factors which contributed to the rapid modernization of the Arabic language was the assimilation of vocabulary of foreign origin ».

En informatique, il existe un certain nombre de termes formés par ce moyen. Nous mentionnons à titre d'exemple, *kamyra* (كاميرا) (caméra) et *fydyw* (فيديو) (vidéo). D'autres termes formés au moyen de l'arabisation existent, mais ils ne sont pas standardisés. Par exemple, le terme *ordinateur* est rendu par *kumbywtar* (كمبيوتر), de l'anglais *computer*. Ce terme a un dérivé adjectival, *kumbywtary* (كمبيوترية) (relatif à l'ordinateur), de l'anglais *computerized*. Parmi les termes arabisés les plus répandus mais non standardisés, nous mentionnons *rAmAt* (رامات) (de l'anglais RAM) et *hArd* (هارد) (de l'anglais *hard*). Bien entendu, l'arabisation subit des règles qui encadrent l'adoption de nouveaux termes dans la langue. Par exemple, pour qu'un terme étranger soit admis en ASM, ce terme doit respecter le système phonétique et morphologique

de la langue. Donc, on dit *Awksyd* (أوكسيد) (oxyde), *mwaksad* (مؤكسد) (oxydé) et *muwaksid* (مؤكسد) (oxydant).

Le dernier type de création terminologique est le mot-valise (ou *AlnaHt* (النحت)). Il s'agit de combiner deux unités lexicales pour en former une et l'associer à un concept. Par exemple, *kahrw* (كهرو) (électro<sup>12</sup>) + *magnaTysy* (مغناطيسي) (magnétique) forme le terme *kahrumagnATysy* (كهرومغناطيسي) (électromagnétique). Le mot-valise est plus observable dans des constructions où un mot est attaché à l'unité, comme dans *lAsilkiy* (لاسلكي) (sans fil), dans *\$abakap lAsilkiyyap* (شبكة لاسلكية) (réseau sans fil). Dans cette construction, le mot *lA* (لا) (sans) est collé à l'unité pour former un seul terme. Cependant, en informatique, cet emploi n'est pas très courant puisque des verbes comme *réinstaller* et *redémarrer* sont rendus en ASM au moyen de termes complexes non combinés, *yuEyd Altavbiyt* (يعيد التثبيت) et *yuEyd AlAiqIAE* (يعيد الإقلاع), respectivement.

## 2.5 Défis et problèmes

Dans ce chapitre, nous avons parlé de la morphologie de l'ASM. Nous avons abordé ses particularités et les problèmes que cela pose aux linguistes, surtout lorsqu'il s'agit de concevoir un système d'analyse morphologique. Nous avons également discuté quelques questions relatives à la terminologie dans l'espace arabophone. Nous avons montré les problèmes de la terminologie et les méthodes de la création terminologique.

Dans le cadre de notre travail, nous avons à faire face aux mêmes défis présentés dans ce chapitre. Puisque nous travaillons sur les unités prédicatives en informatique, le taux d'ambiguïté est important. La forme verbale (voir la section 5.2.4.11), en particulier, pose des problèmes lors de l'identification par un analyseur morpho-syntaxique. De plus, pour qu'un système d'extraction automatique puisse fonctionner, il doit rendre compte des propriétés sémantiques de ces unités. Nous avons montré dans la section 2.3, comment un verbe comme *rafaEa* (رفع) (téléverser) peut être associé à une multitude de formes (suite à l'absence des diacritiques) à cause de son ambiguïté morphologique.

---

<sup>12</sup> *Électro* est un préfixe en français. En ASM, l'équivalent est un mot *kahruba'y* (كهربائي) (électrique).

En ce qui concerne les outils de traitement automatique, il nous semble que la terminologie dans l'espace arabophone doit faire face à la pénurie d'outils de traitement automatique conçus pour la terminologie. Cette question est particulièrement préoccupante pour nous, puisque nous souhaitons soumettre les termes prédictifs à une méthodologie faisant appel à des outils de traitement automatique. Le travail terminologique sur la langue arabe a encore des progrès à réaliser dans ce sens. La compilation des termes se fait manuellement et les outils disponibles ne sont pas adaptés à la terminologie.

De même, le processus de constitution d'un corpus risque, lui aussi, de faire face à un certain nombre de défis. Ces défis concernent les outils de conversion des textes dans un format qu'un système de traitement automatique peut comprendre. En outre, d'autres problèmes surviennent lorsqu'il est question d'intégrer l'ASM à des logiciels et à des applications conçus à des fins de traitement automatique. Ces problèmes concernent l'encodage des caractères arabes qui ne sont pas pris en compte par ces outils. Nous exposons ces problèmes dans les sections 5.2.4.2 et 5.5.6 et nous montrons les difficultés à surmonter.

### **3. État de l'art**

Le présent chapitre décrit différents travaux qui explorent des phénomènes reliés à notre recherche. D'abord, en ce qui concerne les UTP, nous décrivons des travaux portant sur la question en mettant l'accent sur les méthodologies adoptées dans le traitement de ces d'unités, les cadres théoriques dans lesquels elles ont été abordées et leur inclusion dans des ressources spécialisées. Une autre question qui nous intéresse dans ce chapitre est celle des ressources lexicales multilingues qui tiennent compte de l'ASM et les ressources spécifiquement conçues pour cette langue. Nous tentons de voir de quelle manière l'ASM a été abordé et traité d'un point de vue lexical et terminologique.

Ce chapitre est divisé en deux parties. Dans la première partie, nous nous attardons sur les travaux qui ont abordé les unités prédicatives en discours spécialisé (section 3.1). Nous parlons de l'application de la notion de « classes d'objets » aux domaines médical et informatique, de l'application de la théorie de la grammaire des cas à la terminologie bancaire, de l'optique lexico-sémantique et de la terminologie des cadres (*Frame-based terminology*). Nous concluons avec une dernière approche qui propose une classification des verbes en discours spécialisé. Dans la deuxième partie, nous nous attardons sur les travaux portant sur l'ASM (section 3.2). Nous discutons, plus particulièrement, des ressources lexicales disponibles pour la langue générale et la langue de spécialité. Cette section est composée des sous-sections suivantes : les ressources portant sur la langue générale, à savoir Arabic Propbank (APB), Arabic WordNet (AWN) et Arabic VerbNet (AVN); les ressources portant sur une langue de spécialité, à savoir WIPO, ARABTERM (ArT) et The Unified Medical Dictionary (UMD); et les ressources ontologiques qui sont au nombre de trois.

#### **3.1 Partie 1 : Les travaux sur les unités terminologiques prédicatives en terminologie**

Les unités prédicatives (UP) sont des unités qui attirent de plus en plus l'attention des spécialistes en terminologie. L'importance accordée à de telles unités provient du fait qu'elles : 1) désignent des activités dans un domaine; et 2) sont les unités les plus susceptibles de montrer la structure actancielle dans un domaine de spécialité. La structure actancielle est

importante pour nous, puisque nous souhaitons montrer les particularités que les UP peuvent avoir dans un contexte spécialisé.

Dans ce qui suit, nous décrivons les travaux que nous jugeons être les plus significatifs en la matière. Nous nous penchons, d'abord, sur les premières études menées sur les UP en discours spécialisés et, par la suite, nous montrons de quelle manière les auteurs ont abordé la question de ces unités en terminologie au moyen de différentes approches. Nous avons choisi de présenter les travaux de Gross (1994) et Gross et Mathieu-Colas (2001), de Condamines (1993), de L'Homme (1995, 1998, 2002, 2003, 2004 et 2012), de Faber (Faber et al. 2005 et 2009b) et de Lorente (2000 et 2007). Nous signalons que la plupart de ces auteurs se sont penchés sur l'analyse des UTV (unités terminologiques verbales).

Les travaux menés sur les UP en discours spécialisé par les auteurs mentionnés ci-dessus varient en fonction des méthodes et des théories choisies. Gross (1994) et Gross et Mathieu-Colas (2001) font état d'une analyse basée sur les classes d'objets. D'après ces auteurs, la description d'une langue a pour premier objectif de « dresser la liste de tous les prédicats de cette langue et d'attribuer à chacun d'eux les substantifs qu'il peut avoir dans ses positions argumentales » (Gross et Mathieu-Colas 2001 : 70).

Condamines (1993) propose une analyse sémantico-syntaxique qui repose sur la grammaire des cas de Fillmore (1968). Les arguments représentés par des syntagmes nominaux sont étiquetés au moyen de rôles thématiques. Quant aux verbes, ils sont classifiés en fonction de leurs propriétés dans des classes sémantiques. Pour chaque classe sémantique, des rôles thématiques sont prédéterminés. Le domaine bancaire sert d'exemple à des fins de traitement automatique, particulièrement l'acquisition semi-automatique d'un vocabulaire de spécialité (Condamines 1993 : 25).

Pour sa part, L'Homme (1998, 2002, 2003, 2004 et 2012) aborde le verbe sous l'angle de la Lexicologie explicative et combinatoire (LEC) de Mel'čuk et al. (1995). L'Homme propose une méthodologie de sélection et de repérage de verbes en se basant sur un certain nombre de critères qui comprennent, entre autres, la structure actancielle du verbe, ses possibilités combinatoires avec d'autres termes et sa parenté morphologique avec une forme nominale. L'Homme a par la suite étendu ses analyses pour d'autres types d'unités prédicatives, à savoir les noms (L'Homme 2015), en proposant une méthodologie de description adaptée à ce genre d'unités.

### 3.1.1 Classes d'objets

Les travaux sur le prédicat (verbe, nom et adjectif) chez Gross (1994, 2003 et 2008), Le Pesant et Mathieu-Colas (1998) et Gross et Vivès (2001) sont motivés par un souci d'élaborer des dictionnaires électroniques qui rendent compte d'informations relatives à la construction élémentaire de la phrase, comme la séquence des arguments, les valeurs aspectuelles des prédicats et les verbes supports qui précèdent les noms prédicatifs. De plus, la description doit rendre possible : 1) la discrimination des emplois différents des unités lexicales; 2) la désambiguïsation; 3) la résolution de problèmes liés au figement; et 4) le traitement automatique des textes. De ce point de vue, trois éléments de la langue sont désormais combinés pour former un ensemble, à savoir la sémantique, la syntaxe et le lexique.

Dans cette perspective, Gross (2008) estime que le contexte linguistique joue un rôle déterminant quant au sens d'un terme. Selon lui, dans un texte ou discours « les mots ne sont que très rarement ambigus. L'ambiguïté peut apparaître en cas de télescopage de constructions syntaxiques, comme le faisait remarquer avec insistance les premiers travaux de la grammaire générative (Cf. N. Chomsky 1956 : *Les Américains jugent les Vietnamiens coupables*) » (Gross 2008 : 111). Ainsi, il fait les observations suivantes (ibid. : 111) :

Si un mot ne peut pas être défini en lui-même, c'est-à-dire hors contexte, mais seulement dans un environnement syntaxique donné, alors le lexique ne peut pas être séparé de la syntaxe, c'est-à-dire de la combinatoire des mots. La sémantique n'est pas autonome non plus : elle est le résultat des éléments lexicaux organisés d'une façon déterminée (distribution). Qu'il est soit ainsi est confirmé par les auteurs de dictionnaires eux-mêmes qui, timidement et sans aucune méthode, notent pour un prédicat donné, le ou les arguments qui permettent de séparer un emploi d'un autre. Pas de sémantique sans syntaxe donc, c'est-à-dire sans contexte.

Un lien direct est donc établi entre le lexique, la syntaxe et la sémantique. Le contexte, quant à lui, occupe le centre de toute description linguistique du lexique, combinant la sémantique, la syntaxe et le lexique.

Par conséquent, l'idée centrale derrière la notion des « classes d'objets » est que l'unité minimale de sens est la phrase simple et non pas le mot. L'analyse de la phrase se fait en termes de prédicat et d'arguments, soit le schéma « P (Argument 0, Argument 1, Argument 2) » (Gross et Vivès 2001 : 38-39). Le prédicat est considéré comme l'élément essentiel, ou le noyau, et les arguments sont les éléments s'articulant autour de ce noyau. En d'autres termes,



le « prédicat est la constante et les arguments sont des variables » (Gross et Vivès 2001 : 39). Pour pouvoir distinguer le sens d'un prédicat, le linguiste fait appel aux traits syntaxico-sémantiques rattachés aux arguments comme Abstrait, Humain, Concret, etc. Or, ces traits ont l'inconvénient d'être moins précis dans la discrimination de sens d'un prédicat. Considérons les deux exemples suivants pour le verbe *essuyer* :

*Paul a essuyé les assiettes*  
*Paul a essuyé la tache*

Selon Gross (1994 : 17), si l'on considère que le verbe *essuyer* admet un sujet de type Humain et un objet de type Concret, la deuxième phrase ne rend pas compte avec la précision nécessaire du sens du verbe, puisqu'il s'agit d'un autre sens.

De ce point de vue, la façon la plus adéquate de faire la distinction dans un dictionnaire électronique est de préciser les traits afin de rendre compte des propriétés des arguments. Donc, le verbe *essuyer* peut être décrit comme suit :

*essuyer*#1/N0:hum/Nl:inc<vaisselle>  
*essuyer*#2/N0:hum/Nl:inc<meuble>  
*essuyer*#2/N0:hum/Nl:inc<saleté>

Ainsi, les classes d'objets servent à « sous-catégoriser les traits en sous-classes sémantiques » (ibid. : 18). Dans les exemples reproduits plus haut, nous remarquons que le verbe s'associe à trois sous-catégories d'arguments de type *inanimé concret (inc)*, à savoir *vaisselle*, *meuble* et *saleté*. L'argument N0 est de type *humain* dans les trois cas.

Dans l'élaboration du système des classes d'objets, Gross (1994) s'est basé, dans un premier temps, sur des traits syntaxico-sémantiques. Il a dressé une liste de huit traits : Humain, Animal, Végétal, Inanimé Concret, Inanimé Abstrait, Locatif, Temps et Évènement. Par la suite, pour chacun de ces traits, il a donné une description détaillée, précisant les sous-catégories qui relèvent de chacun de ces traits. Puisque le prédicat est l'élément qui motive l'activation des traits (ibid. : 20), chaque trait se distingue par ses propriétés syntaxiques. Dans ce qui suit, nous montrons comment les arguments sont encodés, d'une part, puis les prédicats, d'autre part.

En fonction des classes d'objets, le trait Humain peut avoir un certain nombre de classes dont nous mentionnons une partie ci-dessous (Gross 1994 : 20) :

*adepte* : protestant, taoïste  
*ape* (appellatif) : monsieur, sire  
*col* (...) : foule, troupe  
*déf* (défauts) : menteur, voleur, imbécile, traître  
*doct* (doctrine) : catholicisme, jansénisme  
*écr* (écrit) : livre, Bible  
*fon* (fonction) : adjoint, arbitre  
*grade* (...) : capitaine, colonel  
*ins* (instrumentiste) : violoniste, pianiste  
etc.

Pour ce trait, 54 classes d'objets ont été dégagées. Par exemple, la classe « déf » (défauts) désigne des qualités négatives du trait Humain, comme *menteur*, *voleur*, *imbécile* et *traître*.

Gross (ibid. : 20-21) précise que ces classes ne sont pas des traits, mais des unités distributionnelles élémentaires, puisqu'elles définissent le prédicat avec précision. Pour le traitement automatique, ces classes doivent être décrites syntaxiquement. Par exemple, la classe « défauts » peut être décrite de la façon suivante :

- a) les éléments fonctionnent en position prédicative :  
*Luc est un traître*
- b) ils peuvent être au vocatif :  
*Traître !*
- c) avec la variante :  
*Espèce de traître !*
- d) ou encore :  
*Traître que tu es !*
- e) l'ordre des éléments peut être inversé :  
*Ce traître de Luc*
- f) les éléments sont des compléments naturels en position de second complément  
*du verbe traiter N1 de N2 :*  
*Nous avons traité Luc de traître*

De cette façon, un analyseur syntaxique peut facilement reconnaître les relations argument-prédicat dans un texte. Enfin, dans un dictionnaire, les données sont codées en fonction des classes dégagées comme dans ce qui suit (ibid. : 21) :

*ébardeur/T : hum/C : pro/Op : /S : /D : métall.*  
*ébardeur/T : hum/C : pro/Op : /S : /D : reliure*  
*ébauche/Tihum/Ciécr/Opi/Siesquisse/Dilitt.*  
*ébéniste/T : hum/C : pro/Op : /S : menuisier/D : ébén .*  
*éboureur/T : hum/C : pro/Op : /S : boueux/D : urban .*  
*éboureur/T : hum/C : pro/Op : /S : ébourrer/D : tanner.*  
*ébroïcien/T'.hum/C-.loc/Opi/Sid'Evreux/Dtgéogr.*

Dans ces exemples, *T* désigne le trait, *C* la classe d'objet, *Op* est le verbe associé, *S* est le synonyme et *D* est le domaine.

Les verbes subissent un traitement qui met l'accent sur leurs propriétés transformationnelles. Pour que ces verbes soient admis dans un système d'encodage, il faut voir leurs différents emplois. Gross (1994 : 26-27) donne les exemples suivants pour le verbe *prendre* :

*Luc a pris un steak*  
*Luc a pris de l'aspirine*

*la prise d'un médicament*  
*\*la prise d'un steak*

Sur le plan syntaxique, le verbe ne présente aucune distinction dans la première paire. Dans la deuxième paire, on peut remarquer une distinction claire, la *prise de médicament* n'équivaut pas à la *prise d'aliment*. Donc, le verbe doit appartenir à deux classes différentes. Tout comme les arguments, les verbes reçoivent un type de codage propre à eux. Ainsi, le verbe *prendre* peut être représenté de la façon suivante :

*prendre/N0:hum/Nl:inc<alim>/TR:2*  
*prendre/ N0:hum/Nl :inc<médic>/TR:1*

En ce qui concerne les prédicats, une distinction est faite entre quatre types (Gross et Mathieu-Colas 2001 : 71). Le premier type de prédicat accepte tous les arguments, un exemple de ce type est le verbe *réfléchir à*. Ce verbe n'impose pas de restriction à l'égard de son complément. Donc, on peut dire *réfléchir à ce qu'on va dire*, *réfléchir à ses paroles*, *réfléchir à quelque chose*, etc. Le deuxième type choisit parmi six traits syntactico-sémantiques, à savoir : Humain (*Paul a travaillé*), Animal (*tous les animaux ont crevé*), Végétal (*cette plante a poussé*), Concret (*nous avons déplacé le placard*), Locatif (*toutes les routes mènent à Rome*) et Temps (*la cérémonie a duré deux heures*). Le troisième type

concerne des verbes qui choisissent des substantifs appartenant aux classes d'objets. Ici, les classes d'objets désignent, par exemple, la fonction qu'un trait peut avoir. La fonction dans le cas du trait Humain peut être *voisin* ou *président*. Enfin, pour le quatrième type, la structure argumentale du verbe n'admet qu'un seul substantif, comme avec le verbe *abaisser* dans *abaisser une perpendiculaire*.

Les classes d'objets que nous avons évoquées pour décrire le troisième type de prédicat sont particulièrement caractérisées par le fait qu'elles sont essentielles pour la désambiguïsation du sens du prédicat dans la phrase. Gross (2008 : 11) définit les classes d'objets comme « un ensemble de substantifs, sémantiquement homogènes, qui détermine une rupture d'interprétation d'un prédicat donné, en délimitant un emploi spécifique ». De ce fait, les classes d'objets permettent : a) de spécifier les différents sens des prédicats ; b) d'énumérer les synonymes de chacun entre eux ; et c) de rendre possible la traduction du prédicat en d'autres langues (Gross 2003 : 40-41).

### **3.1.1.1 Les classes d'objets en langue de spécialité**

En terminologie, la notion de « classes d'objets » a été appliquée au domaine médical (Gross et Mathieu-Colas 2001), au domaine du soccer (Gross et Guenther 2002) et au domaine juridique (Lerat 2002a, 2002b, 2005). Partant du principe que la langue technique ne se résume pas à une liste de termes, les auteurs mettent l'accent sur le rôle du prédicat et de ses arguments dans le sens d'un mot. Selon ces auteurs, un verbe est susceptible d'avoir une variété de sens, mais ce qui permet de distinguer ses différents sens sont les arguments avec lesquels il se combine dans la phrase. C'est pourquoi Gross et Mathieu-Colas (2001 : 70) estiment que les entrées dans un dictionnaire spécialisé doivent être des phrases plutôt que des mots. Donc, la phrase est composée d'un prédicat accompagné de ses arguments, ce qui forme un schéma d'arguments. Pour comprendre le sens du prédicat, il faut définir les classes d'arguments qui l'entourent.

Tout comme dans la langue générale, l'application des classes d'objets à la langue de spécialité s'inscrit dans une démarche qui répond aux besoins de la terminologie en termes de description, de traitement automatique, de recherche d'information, de traduction automatique, etc. Dans cette perspective, Gross et Mathieu-Colas (*ibid.* : 70) trouvent que la plupart des lexiques « ne sont pas en mesure de rendre compte du fonctionnement d'une langue de

spécialité avec la précision nécessaire à des fins terminologiques ». Les auteurs estiment que l'application de la notion des « classes d'objets » répond à ces besoins. Nous détaillons dans ce qui suit trois méthodologies basées sur les classes d'objets dans le domaine médical, informatique et juridique.

### 3.1.1.2 Le domaine médical

Dans leur analyse de la terminologie médicale, Gross et Mathieu-Colas (2001) établissent une méthodologie consistant à dresser deux listes : une liste de tous les arguments et une autre renfermant les prédicats. En ce qui concerne la liste d'arguments, on les catégorise en fonction des traits et des classes d'objets que les prédicats sélectionnent. La figure 3 montre un exemple de la langue médicale.



Figure 3 : Catégorisation des arguments (Gross et Mathieu-Colas 2001)

La lettre *G* désigne la catégorie grammaticale, *T* le trait syntaxique et *C* la classe d'objet.

L'argument *virus de la grippe* appartient au trait « micro-organisme » en discours médical. Ce trait admet la classe d'objet « virus ». D'après Gross et Mathieu-Colas (ibid. : 76), les traits admis dans l'analyse du vocabulaire de la médecine sont : Hum=Humain, Ani=Animal, Vég=Végétal, Inc=Inanimé concret, Loc=Locatif, Tps=Temps et Micro=Micro-organisme.

En ce qui a trait à la liste des prédicats, on les décrit en spécifiant le nombre et la construction des arguments. Il faut souligner que la distinction des sens d'un verbe passe par l'établissement de ses classes d'arguments : « quand donc on veut définir la signification d'un prédicat donné, il faut être en mesure d'établir des classes d'arguments. On appellera *emploi* une occurrence de prédicat avec des classes spécifiques de substantifs en position d'arguments » (ibid. : 70). Dans la liste des prédicats, les prédicats font l'objet d'une distinction d'ordre sémantique, à savoir « état », « événement » et « action ». Cette distinction a pour but de reconnaître la construction syntaxique. Considérons la figure 4.

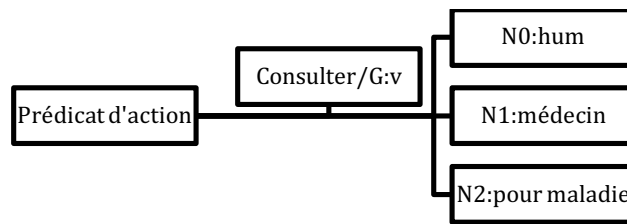


Figure 4 : Prédicat d'action et sa structure argumentale (Gross et Mathieu-Colas 2001)

La figure 4 montre que le prédicat de type *action* est représenté par le terme verbal *consulter* tiré de la terminologie médicale. Ce verbe admet dans sa structure argumentale trois arguments : N0 (qui est de genre Humain, comme le *patient*) ; N1 (le médecin) ; et N2 (la raison de la consultation). Cette analyse tient compte de la spécificité du domaine. Les classes d'objets, en fin de compte, relèvent d'une logique de domaine (Mathieu-Colas 2006 : 12).

Enfin, à la suite d'une analyse de la langue de spécialité, Gross et Mathieu-Colas (2001 : 74) distinguent trois types de vocabulaire :

1. vocabulaire de langue générale ;
2. vocabulaire lié au domaine ;
3. vocabulaire de langue générale, mais avec un sens lié au domaine.

Le premier type concerne toutes les unités lexicales qui appartiennent à la langue générale. Le deuxième type comprend des unités lexicales propres au domaine, par exemple les termes *traiter*, *prescrire* et *opérer*. Le dernier type a trait aux unités lexicales de la langue générale, mais avec un sens terminologisé, comme les verbes *admettre* (un malade dans un hôpital) et *prendre* (un médicament).

### 3.1.1.3 Le domaine de l'informatique

Se basant sur les travaux de Lerat (1995, 2000, 2002a), Mathieu-Colas (2002) a repris cette distinction en l'appliquant à plusieurs domaines de spécialité, dont l'informatique. Donc, il propose cinq types de verbes dans un discours spécialisé (ibid. : 57-58) :

- a) des verbes généraux;
- b) des verbes strictement spécialisés;
- c) des emplois spécialisés des verbes généraux;
- d) des verbes spécialisés repris par la langue générale; et
- e) des verbes *polyspécialisés*.

En ce qui concerne le type (a), il s'agit de verbes qui se trouvent dans la langue générale sans aucun sens spécialisé, comme *correspondre*. Le type (b) comprend des verbes monosémiques qui appartiennent à un domaine de spécialité, au restreint sens du terme, comme *cliquer*, pour l'informatique. Le type (c) regroupe les verbes polysémiques mais qui ont pris un sens particulier dans la spécialité, comme *naviguer* et *attaquer*, dans le domaine de l'informatique. Le type (d) comprend les verbes qui attestent d'un emploi figuré dans la langue générale. Ici, on donne l'exemple de *vacciner* (domaine médical) qui est utilisé dans la langue générale au figuré. Le dernier type désigne les verbes polysémiques, non utilisés dans la langue générale, qui appartiennent à plusieurs domaines de spécialité. Par exemple, le verbe *instrumenter* est utilisé dans des domaines comme la musique, la pétrochimie, la construction et les travaux publics.

Mathieu-Colas (2002) utilise la notion d'« emploi » et les « schémas d'arguments » comme un outil permettant de distinguer les sens d'un verbe spécialisé. Il représente les verbes spécialisés utilisés en informatique comme montré dans la figure 5.

<b>N<sub>0</sub></b>	<b>Verbes</b>	<b>P<sub>1</sub></b>	<b>N<sub>1</sub></b>	<b>P<sub>2</sub></b>	<b>N<sub>2</sub></b>	<b>Dérivés</b>
hum	<b>cliquer</b>	sur	loc:partie de l'écran			<i>clic</i>
hum:programmeur	<b>déboguer</b>		in:programme			<i>débogage</i>
hum	<b>formater</b>		inc:disque			<i>formatage</i>
hum:programmeur	<b>incrémenter</b>		in:variable			<i>incrémentation</i>
hum	<b>télécharger</b>		in:données	sur	in:machine	<i>téléchargement</i>

Figure 5 : Verbes spécialisés en informatique (Mathieu-Colas 2002 : 62)

Dans la figure 5, nous pouvons remarquer que le schéma d'arguments pour un verbe comme *télécharger* est composé des éléments suivants : **N<sub>0</sub>** : hum, **N<sub>1</sub>** : in : *données*, **P<sub>2</sub>** : *sur*,

**N2** : in : *machine*<sup>13</sup>. Dans le cas des verbes de langue générale mais avec un sens spécialisé, le schéma d'arguments sert à distinguer ce dernier. Considérons la figure 6.

<b>N<sub>0</sub></b>	<b>Verbes</b>	<b>P<sub>1</sub></b>	<b>N<sub>1</sub></b>	<b>P<sub>2</sub></b>	<b>N<sub>2</sub></b>	<b>Dérivés</b>
in:programme	<b>accéder</b>	à	loc:mémoire			<i>accès</i>
in:programme	<b>adresser</b>		loc:mémoire			<i>adressage</i>
hum	<b>agrandir</b>		loc:fenêtre			<i>agrandissement</i>
hum	<b>appuyer</b>	sur	inc:touche			
in:virus	<b>attaquer</b>		inc:machine			<i>attaque</i>
in:programme	<b>compiler</b>		in:instructions			<i>compilation, -eur</i>
hum	<b>créer</b>		in:fichier			<i>création</i>

Figure 6 : Verbes de langue générale avec un sens spécialisé (Mathieu-Colas 2002 :62)

Par exemple, le verbe *attaquer* appartient à la langue générale. Dans un contexte spécialisé, tel que l'informatique, le verbe a un sens particulier propre au domaine. Le schéma d'arguments spécifie ce sens spécialisé en mettant à l'avant-scène les éléments qui sont impliqués dans la réalisation de ce sens. Donc, le schéma pour ce verbe est formulé comme suit : **N0** : in : *virus* et **N1** : inc : *machine*.

### 3.1.1.4 Le domaine juridique

Pour le domaine juridique, Lerat (2002b : 157) estime que :

Entre le niveau très particulier du mot et le niveau très général de la phrase, un niveau d'analyse particulièrement fécond est celui de la phrase simple réduite à un schéma propositionnel dont le coeur est une expression prédicative autour de laquelle se distribuent des arguments (sujet et objets).

Lerat se sert également des classes d'objets dans l'analyse et la description de la langue juridique afin de discriminer les sens qu'un verbe peut avoir. Lerat (2002b : 158) donne l'exemple du trait Abstrait. À ce trait, il associe la classe <somme d'argent>, comme dans l'exemple :

<sup>13</sup> Dans le cadre de notre projet, nous dégagons quatre arguments pour *télécharger*.



*Une <somme d'argent> est quelque chose que l'on peut à la fois devoir / payer / percevoir / rembourser / virer.*

Dans cet exemple, nous observons qu'on associe un certain nombre de verbes à cette classe, comme *payer, percevoir, rembourser* et *virer*.

D'après Lerat (2002b : 159) « tout emploi d'un mot spécialisé renvoie à un scénario de réalité lexicalisée au moyen d'un schéma d'arguments » et, par conséquent, « l'implication d'arguments ou même du prédicat apparaît comme un procédé économique normal dans la communication ». De ce fait, les arguments d'un verbe prédicatif déterminent la construction d'un scénario en langue de spécialité. Pour en rendre compte, Lerat met l'accent sur les expressions prédicatives en discours, d'une part, parce que ces expressions représentent les caractéristiques d'un métier, et, d'autre part, parce qu'autour d'elles se distribuent les arguments typiques (Lerat 2006 : 91).

### **3.1.2 Étude des propriétés syntaxico-sémantiques du verbe par l'établissement de classes**

Suivant Fillmore (1968) et Gross (1975 et 1981), Condamines (1993), dans son analyse et sa classification des prédicats et l'analyse de leur structure actancielle dans le domaine bancaire, accorde le rôle principal au verbe. D'après cette analyse, le verbe est le composant essentiel de la phrase, tandis que le nom a le rôle d'actant auquel on attribue un rôle thématique. Condamines (1999 : 42) constate que « en particulier, puisque les domaines spécialisés sont liés à des pratiques professionnelles, la prise en compte des verbes est primordiale pour accéder à une meilleure description de l'action, sous la forme de prédicats associés à des arguments ».

Dans ce qui suit, nous évoquons les trois bases théoriques sur lesquelles reposent les travaux de Condamines sur les verbes spécialisés : la notion des classes d'objets (Gross 1994), la grammaire des cas (Fillmore 1968 et 1977) et le lexique grammatical (Gross 1981). L'objectif est de montrer l'intérêt d'employer chacune de ces bases.

En ce qui concerne la notion des « classes d'objets » de Gross (1994), nous avons abordé cette question dans la section 3.1.1 et nous avons montré que cette approche vise à proposer une description de la langue permettant son traitement automatique. L'un des objectifs principaux de l'analyse proposée par Condamines est également l'automatisation du

traitement de la langue. Le choix de décrire le vocabulaire d'un domaine de spécialité suivant les classes d'objets vient du fait que cette approche répond aux questions fondamentales quant à la structuration du lexique en vue de son automatiser.

La grammaire des cas (Fillmore 1968 et 1977), quant à elle, propose une structuration de la langue (voir section 4.3) selon laquelle les phrases sont décrites en termes de cas. Le rôle primordial est accordé au verbe, puisque cet élément de la phrase détermine, syntaxiquement et sémantiquement, la distribution des arguments, leur nature et leur nombre. Enfin, une des principales caractéristiques de la grammaire des cas est qu'elle combine une description des situations et les représentations syntaxiques. Elle donne des rôles sémantico-syntaxiques aux participants de la situation représentée par la phrase. C'est de là que vient l'importance d'avoir un système de structuration pour un ensemble de termes appartenant à un domaine donné.

En ce qui concerne le lexique-grammaire, son importance réside dans le fait qu'il propose une analyse formelle de la phrase basée sur le verbe et ses arguments. Par exemple, Gross (1981) accorde une grande importance à l'analyse linguistique formelle qui est fondée sur l'association des éléments syntaxiques, représentés par ce qu'il appelle « actants », aux éléments sémantiques, dans ce cas « arguments ». À la suite de son analyse des phrases simples en français, Gross considère que le lexique ne peut pas être séparé de la grammaire, d'où l'importance de considérer la phrase comme l'élément de base de l'analyse. De plus, le verbe n'est pas dissociable de la phrase et son sens est discriminé en fonction de ses arguments.

### **3.1.2.1 Traitement des verbes**

Dans son traitement du verbe en discours spécialisé, Condamines (1993 : 27) précise que le point de départ de sa recherche est le traitement sémantique et syntaxique de différents types d'unités lexicales, mais le verbe est considéré comme « le noyau de la phrase ». Elle part de l'hypothèse selon laquelle les contraintes qui sont liées « au sémantisme du verbe sont à l'origine des contraintes lexicales de la plupart des autres éléments de la phrase, tout au moins de ses arguments » (ibid. : 27). De ce point de vue, le verbe subit deux traitements, à savoir un traitement d'ordre sémantique et un autre de nature syntaxique.

Pour le traitement syntaxique, l'analyse des verbes prend la forme d'une classification basée sur les compléments essentiels que le verbe accepte. Les compléments sont traités selon

leur nombre, la nature de leur construction et leur nature. À titre d'exemple, en ce qui concerne le nombre, certains verbes acceptent 0 complément (exemple, *briller*), d'autres verbes admettent 1 complément (exemple, *écouter*) ou 2 compléments (exemple, *donner*). Quant à la nature de leur construction, on classifie les verbes qui acceptent des compléments directs ou indirects. En ce qui concerne la nature des compléments, on distingue les verbes qui acceptent des compléments de type syntagme nominal, comme dans *organiser une réunion*. À la suite de cette classification, Condamines (1993 : 28) propose vingt-quatre classes de verbes.

Sur le plan sémantique, Condamines (ibid. : 32) constate que le traitement sémantique fait appel à l'intuition du linguiste. De plus, la classification au moyen de classes, telle que conçue par Vendler (1967), semble insuffisante. Pour cette raison, Condamines propose quatorze classes (ibid. : 32-34) : modalité, aspectuel, performatif, télique/atélique, activité, état, changement, mouvement, communication, mental, corporel, sensation, relationnel et sentiment. Ces classes ont pour but de montrer les fonctions différentes des verbes étudiés. Prenons l'exemple de la classe « changement » : un verbe dans cette classe doit indiquer un procès ayant une valeur (-) ou (+) et il peut avoir lieu dans le temps ou dans l'espace. Nous prenons l'exemple du verbe *ouvrir* pour l'appliquer à cette classe. Ce verbe exprime un changement, puisque *ouvrir un compte* indique un changement d'état et il reçoit la valeur « + télique ».

Afin d'attribuer des rôles sémantiques à certains arguments des verbes, il est essentiel de prendre en compte les prépositions. Tout comme les verbes, les prépositions subissent le même genre de traitement sur deux plans, syntaxique et sémantique. L'analyse est faite sur un ensemble de 20 prépositions. Le traitement syntaxique est basé sur l'analyse des contextes syntaxiques droit et gauche pour chaque préposition. Pour le traitement sémantique, Condamines (ibid. : 40) organise les prépositions en trois grandes classes : localisation du procès (temporelle et spatiale), circonstance du procès et intervenants dans le procès.

À la suite de l'étude des verbes et des prépositions, des rôles dits *thématiques* peuvent être définis pour les arguments. Pour Condamines, les rôles thématiques (ou *thêta-rôle*) sont une idée inspirée de Fillmore (1968). Il s'agit de « la relation sémantique qui s'instaure entre un verbe et chacun de ses arguments » (Condamines 1993 : 40). Ce qui est intéressant ici est de voir dans quelle mesure les arguments peuvent exprimer la notion des « intervenants dans

le procès », c'est-à-dire le rôle présumé des groupes nominaux. Pour ce faire, on établit une liste de neuf rôles thématiques :

1. Agent (intervenant de façon directe et active dans le procès);
2. Patient (intervenant de façon passive dans le procès);
3. Cointervenant (c'est le cas de co-agent, co-patient ou co-objet);
4. Objet (affecté par le procès);
5. Propos (thème d'une énonciation);
6. Origine (provenance);
7. Destination (aboutissement);
8. Instrument (moyen d'agir par l'agent);
9. Destinataire (d'une attribution ou d'une énonciation).

Ainsi, une phrase telle que *Monsieur X a retiré de l'argent* peut être analysée de la façon suivante :

- Verbe : classe syntaxique numéro 6 (un verbe avec un seul complément de nature prépositionnelle nominale, préposition+nom)
- Sujet : AGT
- Complément (prépositionnel): OBJ

Le rôle thématique AGT est défini comme étant une entité intervenant de façon active et directement dans le procès. Également, ce rôle doit admettre le trait (+) animé. Pour le rôle thématique OBJ, il s'agit d'une entité affectée par le procès ou qui le subit. Le trait pour un tel rôle est (-) animé. Enfin, le verbe appartient à classe sémantique « mouvement ».

Finalement, les noms, eux aussi, sont classés selon les rôles thématiques qu'ils peuvent avoir dans la phrase et par rapport aux verbes. Comme les verbes et les prépositions, les noms reçoivent les mêmes traitements sémantico-syntaxiques. Condamines ne fait pas état d'une approche pour le traitement syntaxique, mais elle fait appel à la sémantique lexicale pour le traitement sémantique (Cruse 1986). Ainsi, les noms sont classés en fonction des relations qu'ils entretiennent entre eux. L'ensemble de ces relations constitue ce que Condamines

appelle « les relations essentielles ». On trouve, donc, des relations hiérarchiques (taxonomiques et méronymiques) et des relations non hiérarchiques (synonymie, opposition, etc.). Toutefois, les noms reçoivent un traitement basé sur leur comportement fonctionnel, c'est-à-dire que l'analyse se fait en fonction des rôles syntaxico-sémantiques de ces noms dans le contexte.

### **3.1.2.2 Application au domaine bancaire**

Dans l'application de ces classifications au domaine bancaire, Condamines situe le domaine du point de vue du client, soit la relation client – banque. Cette orientation se rapproche de celle de la théorie de la Sémantique des cadres, en ce qui concerne *la perspective* (voir la section 4.3). Toutefois, l'auteure admet que le traitement du sens inverse n'est pas exclu, à savoir banque – client. De plus, elle constate que le rôle déterminant est au domaine et au contexte. De cette manière, elle présente deux niveaux d'analyse et de classification fine, un niveau linguistique et un niveau extralinguistique.

Pour commencer, la classification se divise en quatre groupes :

1. objets bancaires;
2. actants;
3. opérations bancaires;
4. localisateurs concrets.

Dans la figure 7, nous donnons l'exemple de la classe 1, *objets bancaires*.

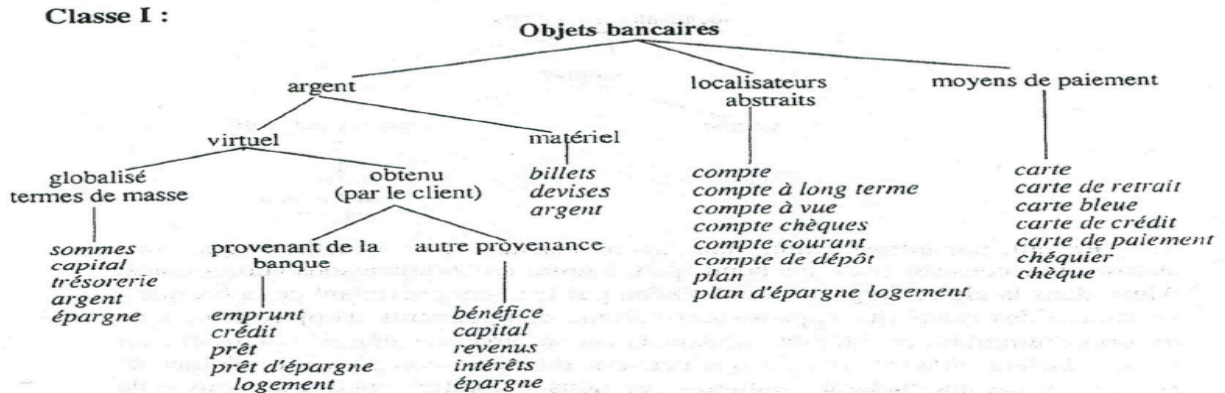


Figure 7 : Classe 1 de la classification du domaine bancaire (Condamines 1993 : 51)

Cette classification du domaine en quatre groupes permet de décrire les combinaisons verbe/nom qui se trouvent dans le domaine. Par exemple, dans la figure, les sous-classes *argent*, *virtuel*, *obtenu (par le client)*, *provenant de la banque* ont des noms qui apparaissent avec comme compléments d'objets directs du verbe *accorder*, comme *emprunt*, *crédit*, *prêt*, etc.

À la suite de cette classification, cinq catégories de verbes sont proposées (ibid. : 55). Ces catégories prennent en compte le sens du verbe et la relation que ce dernier entretient avec ses arguments dans la phrase. Nous soulignons que, dans un contexte, le glissement de sens d'un verbe se traduit souvent par un changement syntaxique, c'est-à-dire que l'on remarquera à la suite de ce changement, une structure argumentale différente. Les catégories sont :

1. les verbes liés au domaine qui sont déjà connus : par exemple, *payer*;
2. les verbes liés au domaine, mais qui sont connus avec un autre sens : par exemple, *ouvrir* et *retirer*;
3. les verbes spécifiques au domaine mais qui sont inconnus : par exemple, *débiter* et *gérer*;
4. les verbes non liés au domaine et qui ne sont pas connus : par exemple, *effectuer*;
5. les verbes non spécifiques au domaine : par exemple, *devoir* et *permettre*.

Parmi les catégories présentées, la seule qui ne fait pas partie du traitement est la catégorie numéro 5. Ce choix est motivé par le fait que ces verbes relèvent de la langue générale.

Dans ce qui suit, nous présentons un exemple d'un traitement du verbe *ouvrir* en tenant compte de ce qui a été présenté jusqu'à maintenant.

Sur le plan syntaxique, le verbe *ouvrir* admet deux compléments, l'un direct nominal et l'autre prépositionnel nominal. Les compléments sont permutable et l'argument prépositionnel peut être omis. Les prépositions qui appartiennent à la structure de la phrase du verbe sont *pour* et *en faveur de*. À titre d'exemple, on peut avoir *un client a ouvert un compte et la banque a ouvert un compte pour le client*.

Sur le plan sémantique, le verbe est traité par rapport aux relations de type antonymique et synonymique. Par exemple, le verbe *ouvrir* a l'antonyme *fermer*. De plus, le verbe *fermer* peut avoir un synonyme, à savoir *clôturer*. En ce qui concerne les 14 classes sémantiques définies pour les verbes, le verbe *ouvrir* appartient aux classes +Télique et Changement d'état. Pour ce qui est des rôles thématiques des arguments, les rôles prédéterminés sont AGT (agent), OBJ (objet), PAT (patient), ORI (origine) et DES (destinataire). La présence de la préposition *pour* permet la sélection des rôles DES et INS (instrument). Par exemple :

*La banque* <sub>AGT</sub> *ouvre* *un compte* <sub>OBJ</sub> *au client* <sub>DES</sub>

*Un client* <sub>AGT</sub> *ouvre* *un compte-chèques* <sub>OBJ</sub>.

Nous avons vu jusqu'ici deux approches aux verbes spécialisés qui mettent l'accent sur la classification des verbes en discours spécialisé et l'attribution de rôles sémantiques ou de classes d'objets aux arguments. L'approche suivante met à contribution la sémantique lexicale dans l'analyse en proposant une description des verbes spécialisés qui montre leur capacité combinatoire avec d'autres termes dans le contexte et les relations qu'ils peuvent avoir avec d'autres formes avec lesquelles il existe une parenté morphologique ou sémantique.

### 3.1.3 L'optique lexico-sémantique

Suivant la Lexicologie explicative et combinatoire (LEC) (Mel'čuk et al. 1995), L'Homme (1995, 1998, L'Homme et Bodson 1998, 2002, 2003, 2004, 2012) aborde les verbes

spécialisés sous un angle lexico-sémantique. Le choix de la Lexicologie explicative et combinatoire comme cadre d'analyse et de description pour les verbes spécialisés réside dans le fait que cette théorie fournit « un modèle élaboré pour rendre compte du lexique en général. En outre, c'est un modèle formalisé, ce qui le rend particulièrement intéressant sur le plan de l'informatisation » (L'Homme 1998 : 63). Ainsi, ce modèle permet de combiner les deux objectifs suivants :

1. Comprendre l'utilisation peu familière d'un verbe donné (décodage);
2. Produire un verbe correct dans un contexte (encodage).

L'Homme (ibid. : 63) postule que le choix des unités lexicales et leur description doivent permettre au consulteur d'un dictionnaire terminologique d'effectuer ces deux tâches.

L'auteure fait remarquer que le verbe ne reçoit pas le même traitement que le nom en terminologie. Pour cette raison, il faut tenir compte du fait que le verbe est un prédicat qui fait appel à des participants et, pour l'étudier, il faut prendre en compte sa structure actancielle (d'après Mel'čuk et al. 1995). En effet, l'identification des actants peut constituer un critère important quant au repérage des verbes spécialisés dans un domaine. De plus, l'importance de la prise en compte des actants permet de « sortir de l'impasse (imposée d'une certaine manière par l'optique conceptuelle) que seuls les verbes associés étroitement à un nom peuvent prétendre au statut terminologique » (L'Homme 2012 : 99).

Dans cette optique, L'Homme et Bodson (1998) proposent un modèle lexicographique et terminographique pour l'étude du verbe spécialisé. Ce modèle intègre une description de la structure actancielle du verbe (Mel'čuk et al. 1995). La structure actancielle nous informe des possibilités combinatoires que le verbe peut avoir avec les noms. Ces derniers sont considérés (dans ce contexte) comme les participants à l'action exprimée par le verbe. Considérons les verbes suivants (L'Homme et Bodson 1998 : 2-3) :



*Charger (1)*

*Structure actancielle :*

*utilisateur,*

*information exécutable; et*

*mémoire de travail.*

*Turner (2)*

*Structure actancielle : logiciel et ordinateur*

Nous remarquons dans ces exemples que le verbe *charger* se combine avec un certain nombre de termes dans le domaine. Ces termes nous informent sur le comportement du verbe et surtout sur ses possibilités combinatoires.

Dans les deux sections précédentes, nous avons montré comment Gross (1994) et Condamines (1993) mettent l'accent sur l'importance de prendre en compte les arguments du verbe. De ce fait, nous constatons que la combinaison verbe/nom semble être un champ fertile pour la description et l'analyse du verbe, puisque celui-ci est analysé de point de vue de sa combinaison avec le nom, qui en est le participant. L'Homme (2001 et 2002) et Claveau et L'Homme (2004) ont étudié cet aspect de la description des verbes spécialisés en terminologie.

L'Homme (2004 : 64-66) a élaboré des critères qui aident les terminologues dans la tâche de repérage des verbes spécialisés. Bien entendu, les critères ne sont pas uniquement destinés aux verbes : on peut les appliquer à toutes les parties du discours, noms, adjectifs et adverbes. Ces critères sont :

1. L'unité lexicale (UL) a un sens lié au domaine : ce critère s'applique aux verbes dont l'utilisation dans un contexte relève d'un emploi propre à un domaine de spécialité, comme le verbe *télécharger* qui est propre au domaine de l'informatique;
2. La nature des actants sémantiques de l'unité prédicative : si les actants du verbe correspondent à des termes, alors il est fort probable que le verbe soit spécialisé. À titre d'exemple, *adresser une mémoire*, où le verbe *adresser* acquiert un sens spécialisé lorsque combiné avec *mémoire* pour désigner la façon dont sont accédées les données sur une mémoire;
3. Les parentés morphologiques que l'UL peuvent avoir avec les termes : si une unité est déjà admise comme terme dans un domaine et ses actants sont

eux-mêmes des termes (critère 2), alors il est fort probable que les unités morphologiquement apparentées soient considérées comme des termes. Par exemple, l'UL *compilateur* est un terme en informatique. Elle a une parenté morphologique avec les unités *compiler* et *compilation*. Alors, elles peuvent être admises comme termes, puisqu'elles véhiculent des sens spécialisés dans le domaine;

4. Les relations paradigmatiques que le verbe a avec d'autres termes : ce dernier critère concerne les liens lexicaux qu'un verbe peut avoir dans le domaine avec d'autres unités déjà admises comme termes. À titre d'exemple, le verbe *télécharger* peut avoir l'antonyme *téléverser*.

Dans la section 5.3, nous nous attardons davantage sur ces critères en les appliquant sur des termes extraits de notre corpus.

À partir des critères présentés, L'Homme (2012) revient sur les critères 2 et 3 afin de donner une nouvelle dimension à l'étude du verbe spécialisé. L'auteure trouve que, souvent, le verbe est étudié du point de vue de son environnement actanciel (ou argumental) ou de la relation morphologique qu'il peut avoir avec un nom déjà admis comme terme dans le domaine. Donc, l'auteure propose de « déplacer le point focal » sur le verbe même. Ce déplacement amène le terminologue à découvrir le réseau lexico-terminologique (ibid. : 100) que les verbes peuvent créer avec d'autres unités lexicales. Afin d'aboutir à un tel réseau, L'Homme (ibid. : 100) suggère la prise en compte aussi des liens paradigmatiques. À titre d'exemple, le verbe *programmer* peut appartenir à un réseau lexical qui comprend, d'une part, les participants à l'action, *informaticien*, *logiciel* et *langage*, et d'autre part, les synonymes et les quasi-synonymes, *écrire* et *développer*, et la forme nominale et la nominalisation du terme, *programmation* et *programme* (dans le cas où une parenté morphologique existe avec le verbe) (figure 8).

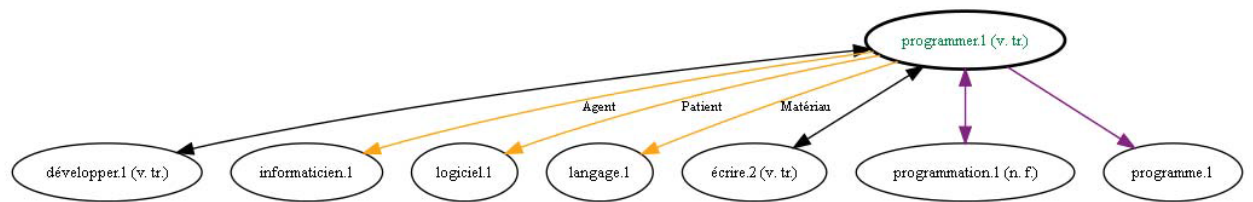


Figure 8 : Réseau lexico-terminologique du verbe *programmer* (L’Homme 2012 : 100)

Dans le cas du verbe en question, *programmer* en tant qu’unité prédicative a une structure actancielle composée de trois éléments, à savoir Agent (informaticien), Patient (logiciel) et Matériau (langage).

Enfin, l’optique lexico-sémantique fait l’objet d’une application en ligne. Il s’agit d’un dictionnaire électronique multilingue (anglais, français et espagnol) libre d’accès qui s’appelle le *DiCoInfo, Dictionnaire fondamental de l’informatique et de l’Internet*<sup>14</sup>. Dans ce dictionnaire spécialisé, les unités terminologiques sont décrites en tenant compte de leur comportement linguistique. L’Homme (2008) constate que les renseignements linguistiques sur les termes dans les dictionnaires et les banques de terminologie sont souvent absents. C’est pour cette raison que L’Homme (ibid. : 79-80) évoque un certain nombre de points qui interpelle le terminologue sur ce sujet.

Le premier point est la prise en compte des parties du discours autres que le nom. Dans les dictionnaires terminologiques, l’utilisateur peut trouver des verbes comme *programmer*, mais un verbe comme *consulter* qui peut apparaître dans une phrase comme *l’internaute consulte une page web* risque d’être absent. L’importance de répertoire de tels verbes dans des dictionnaires spécialisés provient du fait que leurs combinaisons avec d’autres termes dans le domaine constituent un cas qui mérite d’être étudié (ce que le point quatre souligne).

<sup>14</sup> L’optique fait également l’objet d’un autre dictionnaire électronique en ligne pour le domaine de l’environnement qui s’appelle DiCoEnviro. Toutefois, dans ce chapitre, nous essayons de cibler, dans la mesure du possible, les travaux portant sur le domaine de l’informatique.

Le deuxième point est la distinction sémantique fine qui doit être réalisée afin de discriminer les différents sens qu'une unité prédicative peut avoir dans le domaine. L'Homme (2008 : 79) évoque le cas du terme *configuration* en informatique. Ce terme, qui est la nominalisation du verbe *configurer*, désigne « 1. Activité qui consiste à préparer une composante matérielle ou logicielle pour qu'elle fonctionne en conformité avec une autre composante informatique; 2. Ensemble des propriétés rattachées à l'état d'une composante matérielle ou logicielle qui a fait l'objet d'une préparation pour qu'elle fonctionne en conformité avec une autre composante informatique ».

Le troisième point concerne la description des comportements syntaxiques des termes. Dans un dictionnaire spécialisé, il importe de fournir les enseignements sur les structures syntaxiques des termes qui peuvent aider les utilisateurs à mieux comprendre le sens. Par exemple, un verbe comme *démarrer* peut être transitif, c'est-à-dire qu'il admet un complément d'objet direct, *un utilisateur démarre l'ordinateur*, et intransitif comme dans *l'ordinateur démarre*.

Le quatrième point est la combinatoire des termes. Il s'agit ici de montrer aux utilisateurs, comme les rédacteurs professionnels et les traducteurs, l'environnement dans lequel s'utilise un terme. Par exemple, un verbe comme *éditer* se combine en informatique avec des termes comme *texte*. Il en va de même pour le verbe *formater* qui se combine avec *fichier*, *disque*, etc. Ce genre de renseignements assiste l'utilisateur dans la réalisation de certaines tâches, par exemple rédiger un texte portant sur un sujet bien spécifique.

En tenant compte de ces quatre points et en appliquant les critères lexico-sémantiques, les unités terminologiques verbales du DiCoInfo sont traitées sur deux plans, à savoir le plan sémantique et le plan syntaxique. Nous montrons à la figure 9 comment un verbe comme *télécharger* est présenté.

i
✖
Définition
Structure actancielle

Un **utilisateur** déplace un **fichier** ou un **logiciel** à partir d'un **réseau** ou d'un **ordinateur** distant et place le **fichier** ou le **logiciel** sur l'**ordinateur** de l'**utilisateur**.

i

Voir aussi

[transférer](#) <sub>1</sub>

i
Contextes
Contextes annotés
Sommaire

Les polices de caractères demandées ne **sont pas toujours téléchargées** dans le bon format . [ Source : DELAVI1]

En attendant les versions définitives de ces logiciels d'ici un mois ou deux, les versions bêta peuvent **être téléchargées** gratuitement du site Netscape . [ Source : INFOR96]

Il existe deux méthodes de visualisation de séquences vidéos : la méthode en différé et la diffusion en direct. Avec la première méthode, il faut que tout le fichier **ait été téléchargé** sur son disque dur avant que l'on puisse le visionner. [ Source : GIRI]

Famille de mots
Autres liens
i

[anglais](#) : [download](#) <sub>1</sub>  
[espagnol](#) : [descargar](#) <sub>1</sub>  
 Rédacteur(s) : MCLH  
 Date de mise à jour : 07/11/2007

Figure 9 : Le verbe *télécharger* dans le DiCoInfo (DiCoInfo 2015)

Une entrée typique dans le DiCoInfo consiste à présenter le terme suivi de sa partie du discours. Dans le cas de l'exemple de la figure 9, le verbe *télécharger* est présenté comme un verbe transitif (v. tr.). Nous remarquons que le verbe est associé un numéro (numéro 1) qui indique son acception. Puis l'entrée est divisée en quatre sections :

1. Une section dédiée à la définition et à la structure actancielle. Dans la définition, on trouve les éléments définitoires du terme, tels les actants. Quant à la structure actancielle, on y trouve les actants représentés par des rôles sémantiques et des termes typiques : Agent (utilisateur) ~ Patient (fichier) de Source (ordinateur) à Destination (ordinateur);
2. Une section dédiée aux liens paradigmatiques plus connus (quasi-synonymes, antonymes, etc.) : *transférer*<sub>1</sub>;
3. Une section dédiée aux contextes annotés, on y trouve également un onglet qui résume l'ensemble des groupes syntaxiques des actants. Les contextes sont annotés, c'est-à-dire que les actants ainsi que les circonstants

apparaissant dans les contextes où se trouvent également les termes sont soulignés au moyen de couleurs;

4. Une section dédiée aux liens lexicaux. Ici certaines informations lexicales sont données, comme les dérivés du terme. Par exemple, le verbe *télécharger* a une forme nominale et adjectivale, *téléchargement* et *téléchargeable*. Il convient de préciser que dans d'autres entrées, les liens lexicaux sont beaucoup plus nombreux et variés.

Dans le cas où le terme a plus qu'un sens, le DiCoInfo propose plusieurs entrées. Considérons la figure 10.

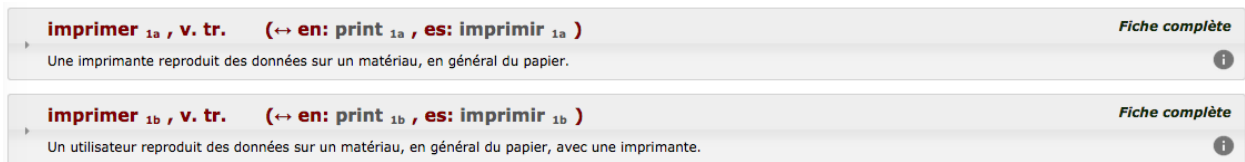


Figure 10 : Verbe avec deux entrées dans le DiCoInfo (DiCoInfo 2015)

Dans la figure 10, le verbe *imprimer* a deux entrées dans le dictionnaire, à savoir *imprimer*<sub>1a</sub> et *imprimer*<sub>1b</sub>. Le terme *imprimer*<sub>1a</sub> a la structure suivante : *imprimante* [Instrument] ~ *données* [Patient], tandis que *imprimer*<sub>1b</sub> a la structure *utilisateur* [Agent] ~ *données* [Patient] avec *imprimante* [Instrument].

Finalement, ce qui a été fait sur les verbes en discours spécialisé a été étendu à d'autres unités prédicatives en terminologie. L'Homme propose une méthode de description et de représentation des unités prédicatives fondée sur les principes de la LEC (Mel'čuk et al. 1995) et la Sémantique des cadres de Fillmore (1976, 1982 et 1985). La particularité que présente la LEC est le traitement unifié des toutes les unités prédicatives (L'Homme 2015 : 8). Par exemple, un terme comme *téléchargement*, qui est une UTP, reçoit le même traitement que le verbe *télécharger* dans le DiCoInfo. À cette unité sont associés : une structure actancielle montrant les participants principaux, des contextes annotés et autres informations lexicales (figure 11).



Figure 11 : Nom prédicatif *téléchargement* dans le DiCoInfo (DiCoInfo 2015)

### 3.1.4 Frame-based terminology (FBT)

Une approche récente à l'étude des termes est *la Terminologie des cadres*, ou *Frame-based Terminology* (FBT), de Faber (et al. 2005, 2009b et 2014). Bien que cette approche ne soit pas une approche dédiée, au départ, à l'étude des verbes spécialisés, il nous semble intéressant de l'aborder pour les raisons suivantes. Il s'agit d'une approche inspirée de la Sémantique des cadres (notre cadre théorique). De plus, ce qui a été fait sur les verbes spécialisés en faisant appel à cette approche mérite d'être discuté, puisqu'on propose de traiter les verbes en fonction des cadres sémantiques. Finalement, l'approche fait l'objet d'une application en ligne. Ceci est particulièrement important pour nous, car nous élaborons une méthodologie qui a pour but de créer une ressource terminologique.

La FBT est une : « cognitive approach to terminology, which directly links specialised knowledge representation to cognitive linguistics and semantics » (Faber 2014 : 15). Cette approche est basée sur la Sémantique des cadres (*Frame Semantics*) de Fillmore (1985 : 222-254; 2006 : 373-400; Fillmore et al. 2003a : 298-332). La particularité de la FBT réside dans le fait qu'elle fait appel à d'autres fondements théoriques comme le modèle de la grammaire lexicale (*Lexical Grammar Model*) (Faber et Mairal 1999), le lexique génératif (Pustejovsky 1991) et la cognition située (Barsalou 2003 et 2008).

La FBT (Faber 2009b : 123) prend en compte les éléments suivants :

1. l'organisation conceptuelle;
2. la nature multidimensionnelle des unités terminologiques; et
3. l'extraction des informations sémantiques et syntaxiques de corpus multilingues.

En ce qui concerne l'organisation conceptuelle, la FBT se démarque par l'accent qu'elle met sur un nouveau modèle de représentation des connaissances spécialisées d'un domaine de spécialité. Parmi les questions que cette approche pose est la façon la plus adéquate de représenter les concepts. Par exemple, comme constaté au début de cette sous-section, l'un des fondements théoriques de la FBT est la Sémantique des cadres. Pour la FBT, la Sémantique des cadres propose une façon de structurer les domaines de spécialité et de créer des représentations des concepts indépendants de la langue (ibid. : 121). Faber (ibid. : 121) fait d'ailleurs remarquer que la terminologie sociocognitive et la théorie communicative de la terminologie n'ont pas fait état d'une réflexion sur l'élaboration, la conception et l'organisation d'une telle représentation ou structure.

Par conséquent, nous constatons qu'en FBT l'idée de cadre, ou *frame*, est un composant d'un réseau plus large comprenant le domaine, le cadre et plusieurs niveaux de spécificités caractérisés par un héritage hiérarchique. De plus, la FBT combine l'idée de cadres à une autre théorie complémentaire qui est celle de la cognition située (Barsalou 2003 et 2008). Cette théorie a pour but de situer la représentation des concepts dans un contexte d'événement ou de processus. D'après Faber et al. (2005), un système de représentation plus crédible sera celui qui se base sur le processus (*process-oriented*) « with goal-derived categories that provide mappings from roles in action sequences to instantiations in the environment » (Faber et al. 2005 : 2). Ainsi :

In order to reflect all dimensions of meaning, concepts must be situated within the context of an event specific to the specialized area. In this sense the notion of frame is a useful construct not only to define concepts (Strehlow 1997), but also to situate them within a context in which categories are related to each other (Faber 2004) (ibid. : 3).



Il convient de mentionner que les termes (qui sont considérés comme des unités de connaissance spécialisée) activent des cadres sémantiques propres au domaine de spécialité. Ces cadres s'accordent avec le domaine ainsi que les connaissances de l'utilisateur. Ainsi, le cadre est le : « context in which FBT specifies the semantic, syntactic, and pragmatic behaviour of specialised language units » (Faber 2014 : 15). Il est basé sur les *micro-théories* suivantes :

1. La micro-théorie sémantique : elle concerne la représentation des unités de connaissance;
2. La micro-théorie syntaxique : elle sert à rendre compte d'évènements qui sont représentés selon la structure prédicat/argument (Faber 2014 : 21); et
3. La micro-théorie pragmatique : elle comprend la théorie des contextes qui peut être linguistique, culturelle ou géographique.

La micro-théorie syntaxique est celle qui nous intéresse le plus pour les raisons qui suivent. D'abord, elle traite les verbes spécialisés du point de vue de la FBT. Ensuite, c'est dans cette micro-théorie que nous trouvons une classification des verbes dans les domaines de spécialité concernées par la FBT. Enfin, elle rend compte de la structure événementielle représentée par la structure prédicat/argument.

#### *La micro-théorie syntaxique*

Cette micro-théorie syntaxique sert à rendre compte d'évènements qui sont représentés selon une structure prédicat/argument (Faber 2014 : 21). Dans la FBT, les verbes font l'objet d'une catégorisation. Tout comme chez Condamines (1993), Faber catégorise les verbes selon la théorie de Vendler (1967) : toutefois, Faber enrichit son approche en ajoutant d'autres catégories qui s'inspirent des travaux de Van Valin (2005), à savoir *semelfactive* et *active accomplishments*. Il est à noter que la catégorie *activité* est remplacée par la catégorie *processus* (figure 12).

Predicate types	Example
State	Atmospheric conditions were favourable.
Process	The wind is blowing.
Semelfactive	Waves hit the cliffs.
Achievement	Waves break in the surf zone.
Accomplishment	The cliff eroded.
Active accomplishment	The ocean flooded shore lagoons.
Causative	The tsunami caused the flooding.

Figure 12 : Catégories des verbes dans la FBT (Araúz et al. 2012, cité dans Faber 2014 : 22)

Ainsi, comme les domaines de spécialité sont décrits en termes d'évènements (qui prennent la forme d'un *cadre* dans la FBT), ceux-ci sont basés sur l'ensemble des actions, des processus, des états, etc. évoqués par les concepts. Par exemple, le ENVIRONMENTAL EVENT (EE) (Faber 2011) est un modèle (ou *template* d'après la terminologie de la FBT) qui représente un évènement prototypique du domaine de l'environnement (figure 13).

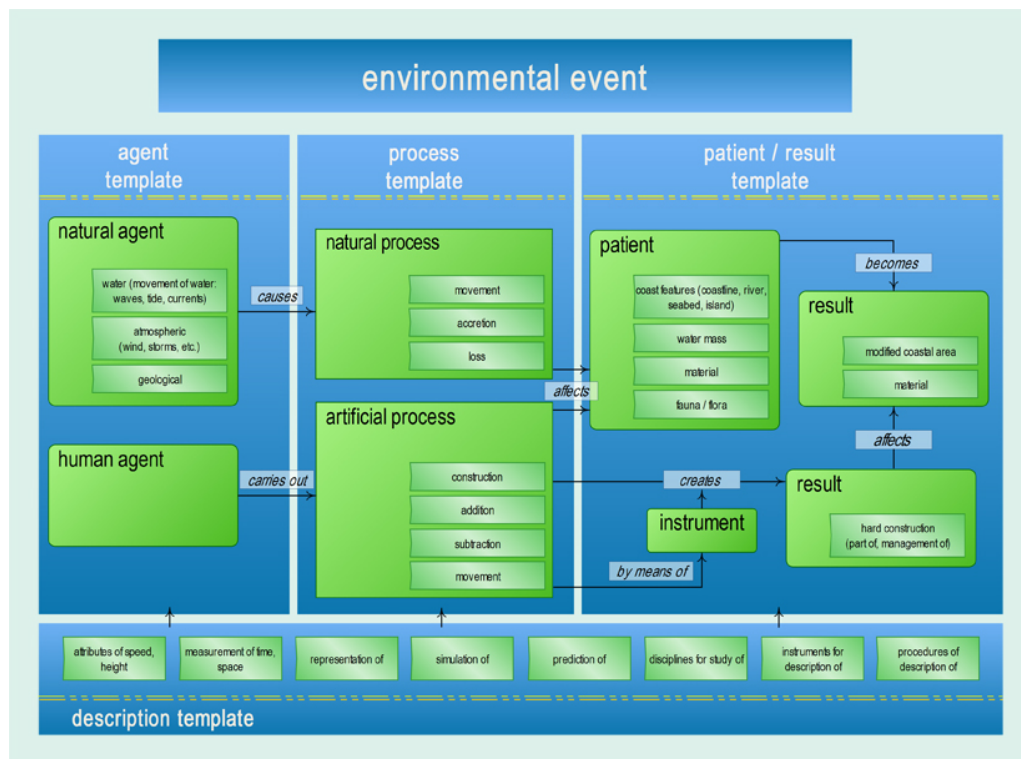


Figure 13 : ENVIRONMENTAL EVENT dans la FBT (Buendia et al. 2014 : 62)

La figure 13 montre que l'évènement est basé sur des prédicats généraux : *cause*, *carry out*, *affect*, *become*, *create*, etc. Les connaissances génériques occupent la tête du modèle

(*template*), suivies par les connaissances les plus spécifiques. Nous remarquons que le modèle se subdivise en quatre sections : celui de l'Agent, du Processus ou de l'État, du Patient et du Résultat. En ce qui a trait à l'Agent, il est présenté comme l'initiateur d'un processus dynamique, soit par une relation de *cause* ou de *carry out*. De plus, il peut être humain ou naturel. Pour ce qui est de Processus/État, il peut avoir deux natures différentes : naturel et artificiel. Les deux types du processus sont liés au Patient/Résultat par relation *affect*, *create* et *by means of*. Pour Patient/Résultat, on trouve un autre composant, à savoir Instrument. Enfin, les relations exprimées dans le modèle sont de nature hiérarchique.

#### **3.1.4.1 Les verbes spécialisés dans la FBT**

Faber et al. (2005) estime que, pour comprendre les discours spécialisés, d'autres unités lexicales (que le nom) doivent être prises en compte, à savoir les verbes qui jouent un rôle crucial : « as semantic predicates, verbs generally determine the overall form and meaning of sentences, which are the linguistic representation of one or various propositions » (Faber et al. 2005 : 5). Par exemple, dans le domaine de l'ingénierie côtière, le verbe *to nourish* désigne le processus de donner à quelque chose (côte ou plage) ce qu'il faut (du sable) afin que ce dernier reste sain. Ce qui distingue le verbe *to nourish* en discours spécialisé sont les limites de son sens par rapport à la langue générale. Cela est dû à la nature des arguments.

Dans cette perspective, le traitement des verbes spécialisés dans la FBT fait appel à une autre base théorique, à savoir le modèle de la grammaire lexicale (Faber et Mairal 1999). D'après cette approche, les verbes dans le lexique d'une langue sont organisés en fonction de dix domaines lexicaux. Ces domaines sont : Existence, Change, Possession, Speech, Emotion, Action, Cognition, Movement, Physical Perception et Manipulation. Chacun de ces domaines comprend des sous-domaines lexicaux, organisés de façon hiérarchique. La figure 14 montre les verbes de sous-domaines lexicaux Cause Motion du domaine de Motion dans le domaine de la volcanologie.

<p><b>1. dégager:</b> libérer une substance <i>Les volcans dégagent du soufre.</i></p> <p><b>1.1. émettre:</b> dégager quelque chose. <i>Le volcan de l'île de Vulcano émet des fumerolles.</i></p> <p><b>1.1.1. laisser échapper:</b> émettre ce qui était retenu. <i>Le volcan laisse échapper des gaz.</i></p> <p><b>1.1.2. exhaler:</b> émettre des substances gazeuses. <i>Le volcan a exhalé 20 millions de tonnes de CO2.</i></p> <p><b>1.1.3. rejeter:</b> émettre hors de soi avec force. <i>Le volcan de l'Eyjafjallajokull rejette des cendres.</i></p> <p><b>1.1.2.1 cracher:</b> rejeter hors de la bouche. <i>Le volcan a craché un panache de cendres.</i></p> <p><b>1.1.2.2 éjecter:</b> rejeter avec force. <i>Le volcan a éjecté d'énormes masses de laves et de cendres.</i></p>	<p><b>1. release:</b> to allow something to flow freely. <i>Volcanoes release massive amounts of co2.</i></p> <p><b>2. emit:</b> to throw out. <i>The majority of volcanoes emit sulphur in the form of gaseous SO 2.</i></p> <p><b>2.2. expel:</b> to emit liquid, gaseous or solid material. <i>A volcano expels more than molten material or lava when it erupts.</i></p> <p><b>2.1.1. eject:</b> to expel out with force. <i>The volcano ejects thousands of tons of boiling, poisonous mud.</i></p> <p><b>2.1.1.1. spit:</b> to eject from the mouth. <i>For more than nine hours the volcano spit vigorous ash in a large plume.</i></p> <p><b>2.1.1.2. erupt:</b> to eject (gas and solid material) out of the volcano. <i>In 1902, the La Soufriere volcano erupted, killing 2,000 people.</i></p> <p><b>2.1.1.2.1. spew:</b> to erupt with force <i>A volcano expels more than molten material or lava when it erupts.</i></p>
---	---

Figure 14 : Verbes de sous-domaine Cause Motion (Sánchez et Buendia 2012 :559)

Dans la figure, nous observons que les verbes sont organisés de telle façon que le verbe occupant la tête de la hiérarchie est le verbe le plus générique. Ce verbe est suivi par des verbes avec lesquels il entretient des relations d'hyponymie. Pour le verbe *éjecter*, le modèle peut avoir la structure suivante, figure 15.

<b>ÉJECTER (FRENCH)</b>					
Lexical domain: <b>MOVEMENT</b> . Frame: <b>Cause Motion</b>					
semantic tag	macrorole	thematic relation	syntactic function	argument example	context
GEOGRAPHICAL ACCIDENT	ACTOR	Agent	S	- volcan	Le <b>volcan</b> Kizimen a éjecté une colonne de cendres à une altitude de 7,4 kilomètres.
NATURAL RESOURCE	UNDER-GOER	Theme	COD	<ul style="list-style-type: none"> <li>- des gaz et du magma</li> <li>- des débris</li> <li>- de la lave</li> <li>- de la matière volcanique</li> <li>- des cendres volcaniques</li> <li>- des bombes volcaniques</li> <li>- des fragments de roche</li> </ul>	<ul style="list-style-type: none"> <li>- Le volcan éjecte <b>gaz et magma</b> à grande vitesse.</li> <li>- Le volcan éjecte <b>des cendres</b> dans l'atmosphère ce qui perturbe le trafic aérien mondial.</li> </ul>

Figure 15 : Modèle du verbe *éjecter* d'après la FBT (Sánchez et Buendia 2012 : 557)

Le verbe *éjecter* appartient au domaine lexical Movement et au sous-domaine lexical de Cause Motion. Nous remarquons que le modèle comprend cinq sections. La section dédiée à l'étiquetage sémantique (*semantic tag*) regroupe les noms qui partagent les mêmes propriétés sémantiques. Les propriétés sémantiques d'un nom sont basées sur la notion des classes d'objets de Gross (1994), les fonctions lexicales de Mel'čuk<sup>15</sup> et l'étiquetage sémantique de Polguère (2003). L'étiquetage sémantique sert à délimiter les sens des arguments du verbe.

<sup>15</sup> Les auteurs ne font pas état de l'utilisation des fonctions lexicales dans leurs travaux. Cependant, dans une note, ils constatent que les fonctions lexicales sont utilisées pour décrire les combinaisons nom/verbe.

Dans le cas du verbe *éjecter*, le rôle Agent est restreint à Geographical Accident, tandis que son Thème est restreint à Natural Resource. Les sections *macrorole* et *thematic relation* décrivent les rôles sémantiques des arguments du verbe. Les relations thématiques décrivent le comportement sémantique des arguments du verbe, tandis que les macro-rôles sont des généralisations des relations thématiques. La section *syntactic function* est dédiée à la fonction syntaxique des arguments. La lettre « S » désigne le sujet de l'Agent, alors que « COD » désigne le complément d'objet direct du Thème. Finalement, dans la section « argument example », les réalisations linguistiques des arguments sont données; et dans la section « context » on trouve des contextes pour le verbe en question. Ces contextes sont extraits d'un corpus.

Finalement, une correspondance entre les domaines lexicaux et les relations thématiques peut être établie (figure 16).

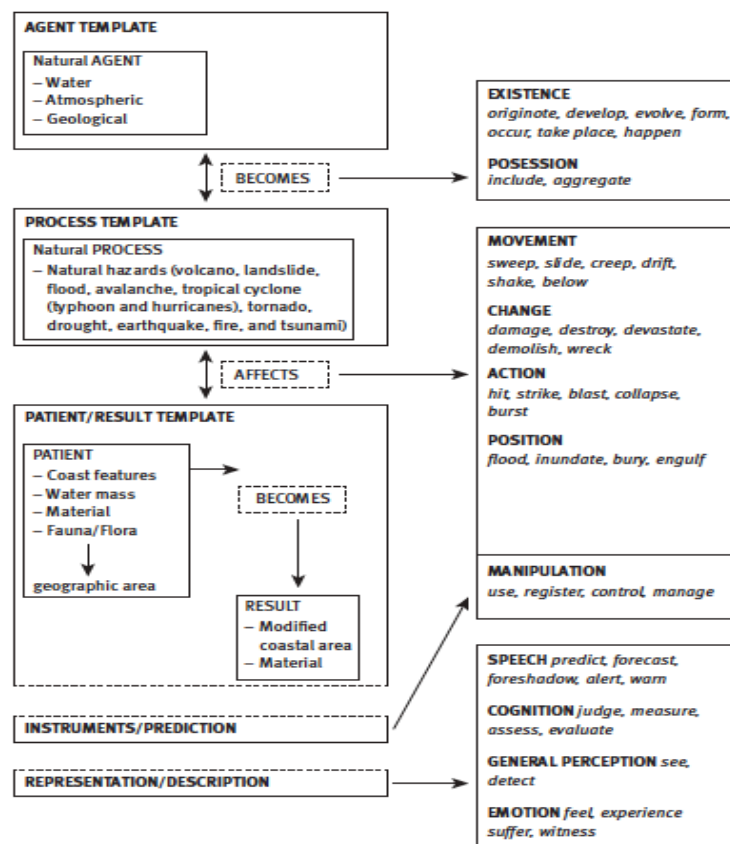


Figure 16 : Activation des domaines lexicaux dans Extreme Event (Buendia et al. 2014 :72)

La FBT fait l'objet d'une application en ligne. Il s'agit d'une base de données ontologique appelée *EcoLexicon* (Faber et Buendia Castro 2014). Cette base de données comprend 3527 concepts et 18 617 termes en anglais, français, espagnol, allemand, russe et grec moderne. *EcoLexicon* est librement disponible. Le concept recherché apparaît au centre avec les équivalents en d'autres langues ainsi que les relations avec d'autres concepts du même réseau. En cliquant sur le concept un réseau plus large apparaît (figure 17).

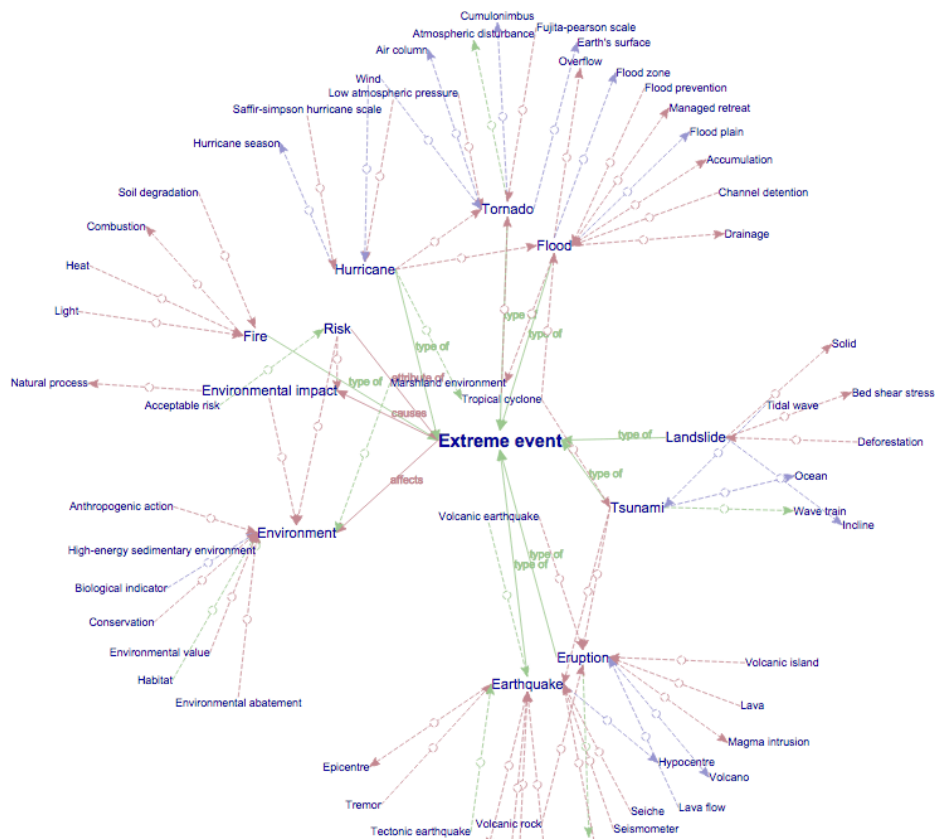


Figure 17 : Résultat d'une interrogation d'un modèle dans *EcoLexicon* (*EcoLexicon* 2016)

Les informations dans *EcoLexicon* sont présentées de la façon suivante. D'abord, le concept ou le terme est placé au centre en caractères gras. Les relations avec d'autres concepts apparaissent sous forme d'un réseau. Chaque relation dans ce réseau est indiquée au moyen d'une couleur distincte. Par exemple, dans la figure, *Extreme Event* est relié à *Environment* par une relation *affect* et à *Earthquake* par une relation *type of*.

En somme, nous avons constaté que la FBT a un système complexe de représentation et de description terminologique. Comme nous l'avons dit au début de cette section, cette approche est basée sur quatre théories, à savoir la Sémantique des cadres (Fillmore 1985), la cognition située (Barsalou 2003 et 2008), le modèle grammaire lexicale (Faber et Mairal 1999) et le lexique génératif (Pustejovsky 1991). À notre avis, un tel système risque de porter à confusion surtout lorsqu'il s'agit d'un travail terminologique orienté vers des besoins particuliers. De plus, dans l'application pratique de la FBT, le projet EcoLexicon, nous ne sommes pas parvenu à trouver une structuration des UT basée sur des cadres et peu d'UTP sont réellement décrites dans la ressource.

### **3.1.5 Autres classifications des verbes spécialisés**

Dans cette sous-section nous présentons une autre approche aux verbes spécialisés visant leur classification. Après une analyse basée sur une approche pragmatico-communicative des textes spécialisés portant sur l'environnement et le droit environnemental en espagnol, Lorente<sup>16</sup> (2000, 2002, 2007 et Lorente et al. 2008) et Lorente et Bevilacqua (2000) constatent que les verbes spécialisés qui apparaissent dans ces textes peuvent être décrits au moyen d'une typologie et faire l'objet d'une validation terminologique en fonction de certains critères. De plus, partant d'un principe pragmatique, Lorente constate que la prédominance d'un groupe verbal donné dans un texte détermine la nature de ce texte et la façon dont le contenu est traité (Lorente 2000, cité dans Buendia 2013 : 370).

La classification des verbes a fait l'objet de deux publications par Lorente (2000 et 2007). En 2000, Lorente organise les verbes selon quatre classes : les verbes discursifs, les verbes connectifs, les verbes phraséologiques et les verbes terminologiques. En 2007, l'auteure revoit sa classification. Cette fois-ci, les verbes s'organisent de la manière suivante : les verbes quasitermes, les verbes phraséologiques, les verbes de relation logique et les verbes performatifs du discours. Dans ce qui suit, nous exposons les deux classifications, par la suite, nous décrivons les critères de validation terminologique proposés par Lorente et, finalement,

---

<sup>16</sup> Nous n'avons pas trouvé de publications de Lorente en langue française ou anglaise. Donc, dans cette sous-section, nous nous basons sur ce qui a été écrit sur les travaux de l'auteure dans d'autres publications, notamment dans Valente (2002), Joan Casademont (2008), Buendia (2013) et Pimentel (2012).

nous comparons les différentes classifications dont nous avons discuté dans ce chapitre, à savoir celle de Gross (1994), de Condamines (1993) et finalement de Lorente (2000 et 2007).

### 3.1.5.1 Les classifications

#### *Première classification*

#### *Les verbes discursifs*

Ces sont des verbes liés aux éléments textuels suivants :

1. la fonction des textes dans lesquels ils apparaissent comme *describir* (*décrire*), *narrar* (*narrer*);
2. les actes de parole comme *decir* (*dire*), *comunicar* (*communiquer*);
3. la structure du discours comme *continuar* (*continuer*), *concluir* (*conclure*);  
et
4. l'objectif du discours du texte comme *presentar* (*présenter*), *convencer* (*convaincre*).

Ces verbes n'appartiennent pas à un domaine donné et ils ne véhiculent pas de sens spécialisé. Ils constituent le métalangage que les spécialistes utilisent pour communiquer. Certains de ces verbes peuvent être de nature hybride, dans le sens où ils sont reliés à la méthodologie de recherche suivie par les experts. À titre d'exemple, on trouve ici des verbes comme *analyser*, *déduire* et *classifier*. Lorente estime que ces verbes ne doivent pas être inclus dans les ressources terminologiques (Lorente 2000, cité dans Buendia 2013 : 371).

#### *Les verbes connectifs*

Ces sont des verbes qui relient deux désignations linguistiques. Il s'agit des verbes copules dont la fonction est d'attribuer des qualités ou des valeurs ou d'exprimer l'équivalence, l'égalité, des similitudes et des relations de dépendances. Par exemple, le verbe *être* est l'un des verbes les plus fréquents qui appartiennent à cette classe et il est utilisé surtout dans les définitions des concepts. Tout comme les verbes discursifs, les verbes de cette classe ne véhiculent pas de sens spécialisé et ils appartiennent à la langue générale.



### *Les verbes phraséologiques*

Les verbes de cette classe expriment l'action, le processus et l'état dans les textes de spécialité. Ils appartiennent à la langue générale, sauf que leur combinaison avec les termes en discours les rend spécialisés. Par exemple, le verbe *générer* n'a pas de sens spécialisé puisqu'il existe dans des contextes de langue générale. Toutefois, sa cooccurrence avec un terme comme *énergie* lui procure un sens spécialisé propre à un domaine. Les verbes supports et les constructions métaphoriques peuvent faire partie de ces verbes aussi. Pour Lorente, ces verbes peuvent apparaître dans des ressources terminologiques, si le sens généré est spécialisé.

### *Les verbes terminologiques*

Ces verbes ont la particularité d'appartenir à un domaine de spécialité. Par exemple, le verbe *eutrofizar* (produire l'eutrophisation) appartient au domaine de l'écologie. Ces verbes ont normalement des dérivations nominales et adjectivales qui sont des termes aussi.

### *Deuxième classification*

Lorente (2007) revisite les classes proposées et les redéfinit comme suit.

### *Les verbes performatifs du discours*

Cette classe remplace la classe des verbes discursifs. Lorente ne constate pas de différence dans cette classe.

### *Les verbes de relation logique*

Cette classe comprend des verbes qui font partie des unités de connaissance spécialisée et qui expriment de relations logiques et génériques.

### *Les verbes phraséologiques*

Cette classe reste intacte. Le seul changement concerne le fait que ces verbes apparaissent dans les collocations, les phrases figées et la construction avec verbes supports (Lorente 2007, cité dans Pimentel 2012 : 94).

### *Les verbes quasitermes*

Pimentel (2012 : 94) constate que la raison pour laquelle Lorente a changé l'appellation de cette classe n'est pas claire. Elle soutient que la seule raison donnée par Lorente semble en contradiction avec la théorie communicative de la terminologie de Cabré (1999)<sup>17</sup>. En effet, Lorente constate que les relations établies entre le verbe et les termes déterminent la classification des verbes dans les textes spécialisés. Cette relation laisse croire que les verbes ne peuvent pas être classifiés par eux-mêmes sans les relier aux autres termes.

#### **3.1.5.2 Critères de validation**

En ce qui concerne les critères de validation de sens des verbes, Lorente (Lorente et Bevilacqua 2000 et Lorente 2002, cité dans Pimentel 2012 : 87-88 et Valente 2002 : 150-152) propose les critères suivants selon lesquels le sens des verbes peut être discriminé :

1. La fonction discursive : le rôle joué par le verbe en contexte détermine sa fonction selon la typologie des textes, le genre du discours et l'acte de parole. Comme montré dans les classes des verbes, ce critère sert à distinguer les verbes à fonction discursive des autres classes de verbes.
2. La valeur spécialisée : ce critère concerne les verbes qui véhiculent un sens spécialisé lorsqu'ils appartiennent à un domaine donné. Contrairement aux autres classes, ces verbes possèdent une valeur spécialisée sans être combinés à d'autres éléments linguistiques dans un discours.
3. Le contenu sémantique du syntagme verbal : chaque syntagme verbal possède un certain nombre de propriétés sémantiques. Ce critère sert à identifier les verbes selon leurs propriétés en discours. Comme illustré dans la partie traitant des classes des verbes, les verbes quasitermes font partie d'un domaine de spécialité, mais pas les verbes performatifs du discours. Ces verbes ont la particularité d'exprimer des concepts liés à la méthodologie de recherche propre aux chercheurs, scientifiques,

---

<sup>17</sup> Nous signalons que Lorente situe son travail dans la théorie communicative de la terminologie (Cabré 1999 et 2003)

académiciens, etc. Par conséquent, les verbes connectifs peuvent appartenir à trois classes : les verbes *dictum*, comme *llamar, denominar, nombrar* (français *appeler, nommer, mentionner*); les verbes copules, comme *ser, parecer, tener* (français *être, sembler, avoir*) et les verbes de relations logiques, comme *estar formado por, proceder de, originar* (français *être constitué de, venir de, être issu de*). En ce qui concerne les verbes phraséologiques, ils tombent dans quatre catégories : construction inaccusative, comme *la temperatura aumenta; el aire circula* (français *la température augmente; l'air circule*); construction résultative, comme *almacenar energía, generar electricidad* (français *stocker de l'énergie, générer de l'électricité*); construction du processus, comme *fermentar azúcar, convertir el calor en energía útil* (français *fermenter le sucre, convertir la chaleur en énergie utilisable*) et construction ergative, comme *la radiación eleva la temperatura/la temperatura se eleva* (français *la radiation augmente la température/la température augmente*). Finalement, pour les verbes terminologiques, ils appartiennent aux verbes d'action et de processus dans un domaine de spécialité.

4. La formation morphologique : les auteurs partent du principe que la morphologie conditionne le sens et le comportement des mots (Lorente et Bevilacqua 2000 et Lorente 2002, cité dans Pimentel 2012 : 88). Ce critère a pour objectif de prendre en compte la formation lexicale des verbes. Par exemple, les verbes discursifs et les verbes connectifs sont des verbes simples. Les verbes phraséologiques peuvent être simples ou dérivés. Les verbes terminologiques sont morphologiquement complexes.
5. Sous-catégorisation : ce critère concerne les propriétés syntaxiques des verbes de chaque classe. Les verbes discursifs sont normalement transitifs. Les verbes connectifs peuvent être transitifs avec un COD ou intransitifs avec un syntagme prépositionnel et des verbes copules. Les verbes phraséologiques peuvent être transitifs ou intransitifs. Finalement, les verbes terminologiques peuvent être intransitifs ou transitifs avec un COD.

6. La relation sémantique avec le sujet : le sujet d'un verbe est distingué au moyen de sa relation avec le verbe (Agent, Cause, Instrument et Cause). Pour les verbes discursifs, leur sujet est toujours un Agent. C'est aussi le cas des verbes connectifs. Pour les verbes phraséologiques et les verbes terminologiques, le sujet relève de la personne qui contrôle l'action (Agent), de phénomènes naturels, d'événements ou de concepts naturels (Cause), d'objets contrôlés par l'Agent (Instrument) et d'objets concrets (Thème ou Résultat).
7. Sélection lexicale : ce critère est utilisé afin d'analyser le comportement phraséologique et les possibilités combinatoires que le verbe peut avoir avec ses arguments et les restrictions qui peuvent survenir. Les verbes discursifs peuvent avoir des cooccurrents mais sans véhiculer de sens spécialisé. Les verbes connectifs ont des restrictions en ce qui concerne l'utilisation des prépositions. Les verbes phraséologiques et terminologiques ont des restrictions évidentes en ce qui concerne leur combinatoire.

Certains chercheurs ont formulé des critiques à l'égard de ces critères. Par exemple, Valente (2002 : 151-152) constate que Lorente et Bevilacqua mettent les verbes support dans le critère des verbes qui ont une valeur spécialisée. L'auteure trouve que cela constitue une contradiction, puisque les verbes support sont vides de sens. De plus, Valente (ibid. : 153) s'oppose aux deux conditions que Lorente et Bevilacqua formulent pour que le critère de la valeur spécialisée soit valable. Ces conditions sont : la valeur spécialisée qu'un verbe peut avoir passe par sa combinaison avec un terme et son lien morphologique avec un terme de nature nominale. Par rapport à la première condition, Valente estime que la combinaison d'un verbe avec un terme ne le rend pas toujours spécialisé. Valente trouve que cette condition n'est pas toujours valable, étant donné que, dans certains cas, des verbes combinés avec des termes ne véhiculent aucun sens spécialisé. En ce qui concerne la deuxième condition, Valente (ibid. : 153) trouve qu'un lien morphologique peut être un critère d'identification d'une UL spécialisée. Toutefois, elle trouve que cette condition exclut les verbes spécialisés qui n'ont pas de lien morphologique avec un terme nominal.

Basée sur sa lecture de Lorente, Pimentel (2012 : 93), de son côté, explique les deux raisons pour lesquelles Lorente a proposé des modifications sur la classification des verbes. De plus, l’auteure en ajoute une troisième. Il est à noter que la première classification de Lorente est basée sur le diagramme de Venn, c’est-à-dire que les classes, représentées par des catégories, sont incluses les unes à l’intérieur des autres. Dans la deuxième classification, la représentation prend une autre forme. Le modèle est basé maintenant sur un organigramme. Dans ce modèle, des conditions de *oui* ou de *non* forment la base de la classification (figure 18).

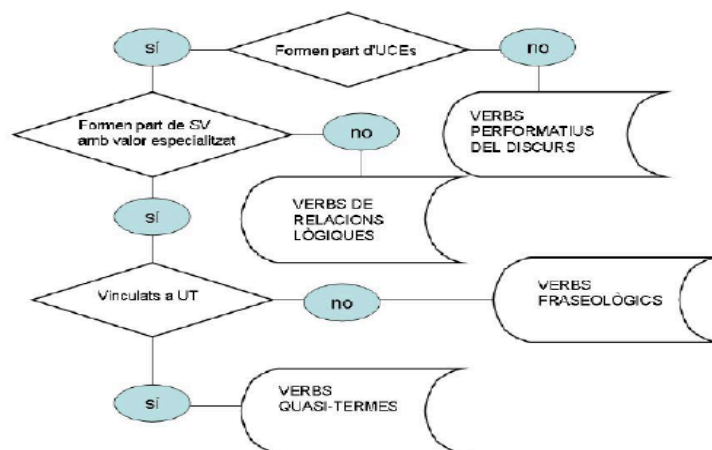


Figure 18 : Classification des verbes en discours (Lorente 2007, cité dans Pimentel (2012 : 93))

Pimentel soutient que la représentation des verbes en catégories incluses dans d’autres catégories laisse croire que les quatre classes partagent des propriétés communes, alors que ce n’est pas toujours le cas. Elle donne l’exemple des verbes terminologiques qui sont inclus dans la classe des verbes phraséologiques, qui sont inclus dans la classe des verbes connectifs.

### 3.1.6 De multiples typologies de classification des verbes : Gross, Condamines, Mathieu-Colas et Lorente

Nous avons montré jusqu'à présent plusieurs approches qui proposent différentes classifications des verbes en fonction du sens qu'ils véhiculent en discours spécialisé. Nous avons montré, également, comment chaque auteur propose de décrire le sens des verbes selon sa propre perspective. Nous allons maintenant tenter de dégager les points communs dans leurs travaux. Dans ce qui suit, nous abordons cette question en analysant les classes des verbes selon chacun des auteurs et en soulignant leurs correspondances aux critères de validation terminologique chez L'Homme (2004) dans les cas où cela s'applique.

#### *Classe 1 : Les unités propres au domaine*

Cette classe concerne les unités prédicatives qui appartiennent au domaine de spécialité. Ici, l'appartenance de ces unités au domaine est évidente, soit à cause du sens véhiculé par l'unité, soit à cause des propriétés sémantiques de l'unité même. Il nous semble que tous les auteurs étudiés dans le présent chapitre se mettent d'accord sur la présence d'un tel type d'unités en langue de spécialité. Gross et Mathieu-Colas (2001) ne donnent pas une appellation spécifique à cette classe, mais ils constatent que ces unités font partie du vocabulaire de spécialité. Nous avons montré dans la section 3.1.1 que les auteurs ont tiré une telle conclusion après avoir analysé le vocabulaire de la médecine en appliquant les traits sémantico-syntaxiques et les classes d'objets. Chez Condamines (1993), cette classe tombe dans la catégorie des verbes qui sont « liés au domaine qui sont déjà connus avec le même sens ». Elle cite l'exemple du verbe *payer* du domaine bancaire. De plus, nous pouvons inclure un autre type de verbes dans la classification de Condamines qui peut appartenir à cette classe. Il s'agit des verbes qui sont « spécifiques au domaine qui sont inconnus ». À titre d'exemple, les verbes *débiter* et *rembourser* sont liés au domaine bancaire. Pour Mathieu-Colas (2002), deux types de verbes correspondent à cette classe, à savoir les verbes dits « strictement spécialisés », soit les verbes monosémiques du domaine, et les verbes « polyspécialisés », soit des verbes polysémiques appartenant à plusieurs domaines. Finalement, pour Lorente, il s'agit de la classe « verbes terminologiques » ou « verbes quasitermes ». Même si elle ne propose pas une classification pour les verbes en discours spécialisé, nous trouvons intéressant de

signaler que L'Homme affirme cette classe des verbes dans la langue de spécialité en accordant le premier critère de validation terminologique à cette classe (L'Homme 2004).

### *Classe 2 : Les unités potentiellement spécialisées*

La deuxième classe regroupe des unités qui ont le potentiel d'appartenir à un domaine en sachant qu'elles proviennent de la langue générale. Pour Gross et Mathieu-Colas, ce sont « des éléments du vocabulaire général mais affectés d'un sens particulier dans le domaine considéré » (2001 : 74). On peut avoir des noms comme *circulation* (en physiologie), des verbes comme *admettre* (un malade dans un hôpital), *prendre* (un médicament) et *se réveiller* (en parlant d'une douleur). Bien entendu, Ce sont des unités de la langue générale mais utilisées dans un domaine avec un sens spécialisé. Pour Condamines, ce sont des verbes « liés au domaine qui sont connus avec un sens différent » (1993 : 56). Elle donne l'exemple de *ouvrir* et *retirer* du domaine bancaire. Ces deux verbes expriment un sens spécifique aux transactions bancaires. Pour Mathieu-Colas, cette classe comprend les verbes polysémiques généraux avec des emplois spécialisés. Pour Lorente, ce sont les *verbes phraséologiques* qui entrent dans la formation de collocations, de phrases figées et de construction avec verbes supports. Ces verbes ne peuvent pas être traités en dehors du contexte spécialisé, c'est leur combinaison avec un terme qui leur donne un sens particulier. Pour L'Homme (2004 : 65), cette classe relève du deuxième critère de validation terminologique. L'unité prédicative a un sens spécialisé que « lorsque qu'elle est accompagnée des actants sémantiques de sens spécialisé » (ibid. : 65). Si le sens de l'unité reste tel quel avec des actants sémantiques non spécialisés, dans ce cas l'unité n'est pas spécialisée. L'auteure donne l'exemple du verbe *adresser* dans des contextes comme *adresser une mémoire* et *adresser un message*.

### *Classe 3 : Les unités du lexique général*

Cette classe comprend les unités qui se trouvent dans un discours spécialisé mais elles ne véhiculent pas de sens propre au domaine. Gross et Mathieu-Colas constatent que ces unités relèvent de la langue générale et elles ne sont pas spécifiques au domaine. Les exemples donnés par les auteurs méritent d'être discutés et nous reviendrons sur cette question plus loin. Condamines classe ses unités sous le type « non spécifique au domaine ». Par exemple, dans le domaine bancaire, Condamines trouve des éléments comme *pouvoir* et *devoir*. De plus, elle

intègre un autre type de verbes dans cette classe. Il s'agit des verbes qui ne sont pas liés au domaine et « qui ne sont pas connus ». Ces verbes ne sont pas connus dans le sens où ils ne font pas partie du français fondamental. Par exemple, le verbe *effectuer* est repéré par un système de traitement automatique, sauf que Condamines propose un traitement de ce genre d'unités. D'après les analyses d'un corpus portant sur le domaine bancaire, ces unités apparaissent avec certains noms prédicatifs. À titre d'exemple, la phrase *effectuer un virement*. Mathieu-Colas classe ces verbes sous l'étiquette des « verbes généraux ». De plus, l'auteur donne un autre type de verbes que nous jugeons être compatible avec cette classe, à savoir les « verbes spécialisés repris par la langue générale ». Ces verbes sont terminologiques sauf qu'ils sont affectés par une *déterminologisation*, c'est-à-dire qu'ils s'emploient dans la langue générale (Meyer et Mackintosh 2000). Enfin, Lorente propose les classes des verbes performatifs du discours et celle des verbes de relation logique. Les verbes performatifs du discours ne font pas partie d'un domaine de spécialité, tandis que les verbes de relation logique (ou verbes connectifs) peuvent être considérés comme des unités de connaissance spécialisée. Étant donné que Lorente considère deux dimensions dans son analyse (sens spécialisé versus connaissance spécialisée), ces verbes peuvent être classifiés dans la classe précédente.

### **3.1.7 Autres travaux sur les unités prédicatives**

Dans cette sous-section, nous présentons d'autres travaux effectués sur les unités prédicatives spécialisées. Ces travaux sont basés sur les approches dont nous avons déjà discuté dans ce chapitre. Nous avons trouvé utile et pertinent de les aborder, puisqu'ils donnent d'autres dimensions intéressantes aux approches sur lesquelles ils sont basés. Dans ce qui suit, nous abordons les travaux suivants : les unités phraséologiques spécialisées du domaine de l'environnement (Buendia 2013), les verbes spécialisés du domaine juridique (Pimentel 2012) et les verbes spécialisés dans le domaine de la médecine (Wandji Tchami 2014).

#### **3.1.7.1 Les unités phraséologiques en discours spécialisé**

Buendia (2013) a abordé les verbes dans le domaine de l'environnement. Son cadre théorique est la terminologie Sociocognitive (Temmerman 2000 et 2001) et la terminologie



des cadres (FBT) (Faber 2014). L'auteure se base aussi sur la classification de Lorente (2000 et 2007) dans l'analyse et la description des unités phraséologiques. L'objectif de la recherche est de formuler un modèle de description des unités phraséologiques afin d'enrichir une base de données terminologique en ligne, EcoLexicon. Le modèle proposé par l'auteure est celui dont nous avons discuté dans la section 3.1.4 traitant la FBT. Donc, les unités phraséologiques sont décrites au moyen de gabarits (*templates*).

Buendia part du principe que les verbes et leurs arguments peuvent être classifiés et structurés selon un ensemble de catégories sémantico-conceptuelles qui reflètent la structuration du domaine même. Par conséquent, lorsque les rôles sémantiques des verbes, leurs macro-rôles et la structure de la phrase qui résulte de ces derniers sont spécifiés, il serait alors possible d'établir des modèles qui représentent ce sens pour le cadre entier. Pour cette raison, l'auteure a choisi de travailler sur le sous-domaine de Extreme Event (ou « événement extrême », qui concerne des phénomènes naturels tel l'ouragan). Ce sous-domaine correspond à un cadre spécifique en FBT. Il est à noter que ce cadre fait partie d'un cadre plus large qui est le *Environment Event* (EE) déjà discuté dans la section 3.1.4.

En s'appuyant sur le modèle de grammaire lexicale (Faber et Mairal 1999), Buendia suggère une classification des verbes en fonction de dix domaines et sous-domaines lexicaux. Ces informations font l'objet d'une intégration dans EcoLexicon. Chaque verbe est encodé dans un modèle (figure 19) qui fournit une description du cadre comportant les éléments suivants (Buendia 2013 : 385) :

1. le domaine lexical activé par le cadre : comme le domaine *Existence* ;
2. le nom du cadre : comme le cadre *to\_begin\_to\_exist*;
3. la définition du cadre : la définition du cadre dans 2 est « [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER] »;
4. les rôles sémantiques : Theme, Geographical location, Time et Manner ;
5. les macro-rôles : Undergoer;
6. les étiquettes : Natural disaster;
7. les réalisations linguistiques : fire season, small islands, earlier et suddenly;
8. les types de syntagmes : NP, PP (in, about), PP (on, at) et AVP; et

9. des exemples annotés.

1 Lexical domain: EXISTENCE					
Frame: to_begin_to_exist					
Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
Semantic role	START <sub>2</sub>	Theme	Geographical location	Time	Manner
Macrorole		Undergoer	_____	_____	_____
Labels		natural disaster, natural disaster season	_____	_____	_____
Linguistic realizations		fire season, hurricane season, hurricane, tornado	small islands, Boone county	earlier, 14 April, 3.30 pm, twenty ninth of September	suddenly
Phrase type		NP	PP (in, about)	PP (on, at)	AVP
Usage examples	<ol style="list-style-type: none"> <li>1. <b>Fire seasons</b> will start <b>earlier</b> and will finish later in the year.</li> <li>2. The most active phase of the <b>eruption</b> started on <b>14 April</b>, resulting in about 0.25 km<sup>3</sup> of ejected tephra.</li> <li>3. The <b>tornado</b> started at <b>3:30 pm</b> about <b>1.2 miles north of Poplar Grove</b> in Boone county and ended at 3:48 pm.</li> <li>4. This tragic <b>hurricane</b> started suddenly in the <b>small islands of Cape Verde</b> on the <b>twenty ninth of September</b>.</li> <li>5. The Atlantic <b>hurricane season</b> starts on <b>June 1</b> and ends on Nov 30 of every year.</li> </ol>				

Figure 19 : Analyse du cadre to\_begin\_to\_exist (Buendia 2013: 385)

À titre d'exemple, le domaine lexical *Existence* (ibid. : 387) comprend les cadres suivants :

*to begin to exist, to begin to exist from sth else, to begin to exist becoming sth else, to cause to exist/happen, to exist in time, to continue to exist (of people), to continue to exist (of natural disaster), to cease to exist, to cease to exist in the perception of others, to cause sb to cease to exist, and to cause to cease to exist (of fire disaster).*

Le cadre *to begin to exist* comprend les verbes anglais suivants : *start<sub>2</sub>, break out, originate, develop, evolve, blow up, form, burst<sub>2</sub>, erupt<sub>1</sub>, explode<sub>2</sub>*. Le cadre comprend, également, les verbes espagnols suivants : *empezar, despertar, comenzar, originarse, desarrollarse* et *formarse*.

Dans EcoLexicon, ce cadre est représenté de la façon suivante (figure 20).

Nuclear meaning	EXISTENCE
Meaning dimension	to_begin_to_exist
Phraseological pattern	NATURAL DISASTER begins to exist in a particular LOCATION, at a given TIME, or in a specified MANNER.
Verbs	<a href="#">blow up</a> , <a href="#">burst2</a> , <a href="#">develop</a> , <a href="#">evolve</a> , <a href="#">form</a> <a href="#">originate</a> , <a href="#">start2</a>

Figure 20 : Cadre to\_begin\_to\_exist dans EcoLexicon (EcoLexicon 2016)

En cliquant sur un verbe, un certain nombre d'informations s'affiche. Par exemple, pour le verbe *blow up*, on peut avoir l'usage, comme dans *A hurricane had blown up in the Gulf* et une note qui donne davantage d'informations sur le cadre, comme *The Natural DISASTER can also be an ATMOSPHERIC AGENT, but it is usually associated with wind events.*

### 3.1.7.2 Les verbes du domaine juridique

En s'appuyant sur la théorie de la Sémantique des cadres, Pimentel (2012) analyse et décrit les verbes spécialisés en anglais et en portugais dans le domaine juridique. Son analyse est basée sur le modèle de FrameNet (Ruppenhofer et al. 2010) et la méthodologie du DiCoInfo (L'Homme 2008). Les résultats de son étude des verbes sont encodés dans une base de données appelée JuriDico. À partir de 200 verbes analysés en anglais et en portugais, 165 paires équivalentes ont été identifiées et regroupées dans 76 cadres.

La méthodologie suivie est celle de *bottom-up*, c'est-à-dire que pour décrire les verbes spécialisés d'un domaine, l'auteure part des occurrences attestées dans un corpus. Les analyses sont faites à partir d'un corpus comparable bilingue, anglais-portugais. La première étape consiste à extraire les verbes au moyen d'un extracteur automatique (TermoStat, Drouin 2003). Par la suite, les verbes subissent une validation terminologique et sont encodés dans une base de données. Pour chaque verbe, un certain nombre de contextes sont retenus. Ces contextes sont annotés afin de montrer l'unité en question et ses participants (actants et circonstants). La dernière étape consiste à regrouper les verbes dans des cadres sémantiques. Ces cadres sont évoqués par des unités prédicatives verbales. Pour chaque cadre, des éléments sont décrits. Ces éléments correspondent aux participants à la situation décrite. Les éléments

de cadre sont traités en fonction de leur présence (obligatoire ou non). Donc, l'unité peut avoir des éléments obligatoires (*core frame elements*) et des éléments non obligatoires (*non-core frame elements*).

À titre d'exemple, un cadre comme Compliance est décrit de la façon suivante (figure 21).

<b>Compliance:</b> A sentient Protagonist or the Act for which they are responsible complies or not with the Law.			
Act or Protagonist		Terms	Law
action, copy, detention, order, roadblock, unavailability, use	child, owner, police, SPCUM, state	<i>infringe<sub>1</sub></i>	copyright, freedom, law, liberty, patent, provision, right, rule, section, security
action, change, detention, search, supression, transfusion, use	Agency, authority, BNS, City, counsel, Crown, government, IRPA	<i>violate<sub>1</sub></i>	agreement, Charter, duty, law, obligation, order, principle, provision, right, section, undertaking

Figure 21 : Cadre *Compliance* dans le domaine juridique (Pimentel 2011 : 159)

Dans la figure, on observe les éléments suivants : la définition du cadre, les éléments du cadre et les unités prédicatives (les termes). Il est à noter, que dans la constitution des cadres, les critères de définition de cadres décrits dans (Ruppenhofer et al. 2010) sont appliqués. Nous reviendrons sur cette question dans le chapitre 5.

### 3.1.7.3 Simplification des textes médicaux au moyen des verbes

Les travaux de Wandji Tchami (Wandji Tchami et al. 2013 et Wandji Tchami 2014) sur les verbes en discours médical cherchent à répondre aux questions relatives à :

1. La simplification des textes en français par l'implémentation d'un outil qui repère les emplois verbaux peu communs au discours des patients. Cet outil doit ensuite proposer des emplois sémantiquement similaires, mais plus adaptés au niveau de spécialisation de ces utilisateurs (Wandji Tchami 2014 : 37).

2. La proposition d'une analyse comparative des corpus médicaux écrits caractérisés par plusieurs niveaux de spécialisation. L'accent est mis sur les verbes utilisés dans ces corpus et leur structure actancielle (actants) (Wandji Tchami et al. 2013).

L'auteure part de l'hypothèse selon laquelle le verbe est l'élément central dans la phrase et, par conséquent, il est le point focal pour décrire la syntaxe et la sémantique des textes spécialisés. Dans son travail, l'auteure s'appuie sur des ressources existantes, notamment le DiCoInfo et FrameNet. Ces ressources permettent de décrire les verbes dans le domaine médical tout en dégagant les différences linguistiques observées dans des corpus de plusieurs niveaux d'expertise (Wandji Tchami et al. 2013). En outre, la théorie de la Sémantique des cadres (Fillmore 1982) est prise comme un cadre de travail.

La méthodologie de travail comprend trois étapes (Wandji Tchami et al. 2013 : 45) :

1. L'annotation automatique des arguments des verbes selon leurs catégories sémantiques : il s'agit d'annoter automatiquement les verbes et les arguments et d'attribuer des étiquettes sémantiques à ces derniers. Pour réaliser cette tâche, une ressource sémantique a été utilisée : il s'agit du *Snomed International Terminology* (Côté 1996). Cette ressource est structurée en fonction de onze axes sémantiques : *topography* ou *anatomical locations*, *social status*, *procedures*, *living organisms*, *professional occupations*, *functions of the organism*, *disorders* et *pathologies*, *chemical products* et *physical agents*. Un certain nombre de termes sont rattachés à chaque axe. Ces termes servent d'unités lexicales et les axes servent à étiqueter ce qui pourrait correspondre à des éléments de cadre (conformément à la Sémantique des cadres).
2. L'analyse des verbes : les informations obtenues lors de la détection et de l'annotation automatique des arguments effectuées au cours de l'étape précédente servent à analyser les verbes. L'analyse a pour but de spécifier

leur nombre et leur type sémantique et, par conséquent, discriminer les emplois spécialisés.

3. La simplification des emplois des verbes : cette étape a pour but de simplifier les emplois des verbes pour qu'ils soient plus abordables aux utilisateurs non spécialistes. La stratégie de simplification s'appuie sur le modèle proposé dans le DiCoInfo qui fournit une définition typique de la construction dans laquelle entre le verbe.

### **3.1.8 Importance du verbe spécialisé et notre apport personnel**

Dans les sections précédentes, nous avons fait un survol d'un certain nombre de travaux que nous jugeons être les plus rapprochés de notre travail. Nous avons montré de quelle façon chacun des auteurs aborde la question des verbes spécialisés en terminologie. Nous avons également discuté des différents cadres théoriques qui sont adoptés pour l'étude des verbes en discours spécialisé. Ces travaux soulèvent un certain nombre de questions à propos de l'importance des verbes en terminologie et les aspects les plus souvent abordés par les spécialistes.

D'abord, la première observation que nous pouvons faire est que l'étude de verbe en terminologie a largement profité de son traitement en langue générale. Considérons la citation suivante de Fellbaum (1990 : 278) :

The verb provides the relational and semantic framework for its sentence. Its predicate-argument structure (or subcategorization frame) specifies the possible syntactic structures of the sentences in which it can occur. The linking of noun arguments with thematic roles or cases, such as INSTRUMENT, determines the different meanings of the events or states denoted by the sentence, and the selectional restrictions specify the semantic properties of the noun classes that can flesh out the frame.

Les ressources lexicales de la langue générale ont servi de modèle pour la langue de spécialité et la terminologie. Cela se manifeste clairement dans les ressources terminologiques (section 3.1) qui sont basées sur des cadres théoriques originalement conçues pour la description des données lexicales. Dans la citation reproduite plus haut, Fellbaum, qui s'intéresse aux verbes anglais dans les réseaux sémantiques, résume (de manière indirecte) les

aspects abordés dans l'étude du verbe en terminologie. Dans ce qui suit, nous nous attardons sur les trois aspects qui sont, à notre avis, les plus intéressants pour le travail terminologique actuel.

Le premier aspect est la structure actancielle. La structure actancielle, d'après Fellbaum, spécifie la structure possible de la phrase dans laquelle apparaît le verbe. En terminologie, la structure actancielle permet de donner des indications sur la manière de délimiter le sens d'un terme. Nous avons montré comment un verbe comme *télécharger* est analysé dans le DiCoInfo et chez Mathieu-Colas (2002). Puisqu'elle est associée à des arguments (ou encore actants), l'UP (en général) est délimitée par des frontières. Cette délimitation permet de mieux comprendre le genre de structure contextuelle à laquelle elle correspond.

Le deuxième aspect est l'importance accordée à l'association des rôles thématiques (au sens large) aux arguments. Dans une structure actancielle où le sujet du verbe *télécharger* est un Agent (la personne en charge du processus) et son objet est un Patient (l'objet téléchargé) : il est permis de croire que ces renseignements aident le terminologue dans son analyse des différentes propriétés sémantiques de l'UP.

Le troisième aspect est la prise en compte des propriétés sémantico-syntaxiques des verbes en discours spécialisé. Cela joue un rôle important dans l'étude du verbe, surtout en vue de son inclusion dans des ressources terminologiques modernes. Compte tenu de l'optique conceptuelle, les ressources terminologiques classiques se distinguent par une limitation en ce qui concerne le choix d'unités incluses, étant majoritairement des unités nominales. Par contre, des ressources modernes admettent maintenant des unités prédicatives, notamment des verbes.

Notre travail va dans le même sens que les travaux présentés dans ce chapitre. Cependant, nous nous distinguons des autres travaux par ce qui suit.

Premièrement, nous proposons une ressource terminologique de l'informatique conçue spécifiquement pour les verbes et les noms déverbaux en langue de spécialité (les UTV et les UTN). En outre, dans la conception d'une telle ressource, des outils informatiques ont été

adaptés pour le traitement des verbes et des noms déverbaux. Par exemple, nous avons réalisé une extraction semi-automatique de ces unités (voir chapitre 4).

Deuxièmement, nous travaillons sur une langue sémitique, ce qui représente un aspect original étant donné que les travaux consultés se penchent sur des langues comme l'anglais et le français. L'ASM se caractérise par les défis qu'il soulève, compte tenu des facteurs suivants : 1) sa morphologie complexe et riche; 2) la difficulté que présente son traitement automatique (voir sections 2.3) ; et 3) le nombre limité de ressources terminologiques basées sur des principes discutés dans cette section, surtout en ce qui concerne la structure actancielle des unités.

Troisièmement, nous proposons un modèle de représentation terminologique fondé sur des cadres sémantiques. Ce modèle est basé sur la Sémantique des cadres dans le sens où il utilise des cadres pour modéliser les UTP de l'informatique. De cette façon, les termes qui partagent les mêmes propriétés syntaxiques et sémantiques sont associés à des scénarios conceptuels. De plus, les cadres sémantiques permettent de mieux comprendre les propriétés des termes, étant donné que pour chaque terme on associe des contextes annotés soulignant les participants principaux à l'action exprimée par le terme. Finalement, les cadres sont reliés de façon à montrer les relations qu'ils entretiennent entre eux dans le domaine.



## 3.2 Partie 2 : les ressources lexicales de l'arabe standard moderne

La deuxième partie de notre chapitre est consacrée à l'ensemble des travaux qui portent sur les ressources lexicales de la langue arabe. Cette partie est divisée en trois sections : la section 3.2.1 décrit les ressources générales, la section 3.2.2 décrit les ressources spécialisées et la section 3.2.3 décrit trois travaux visant l'élaboration d'ontologies. Dans notre choix de ressources générales, nous nous sommes limité aux travaux qui s'inscrivent dans une démarche s'appuyant sur un cadre théorique. C'est pour cette raison que nous ne présentons pas une description exhaustive des ressources lexicales. Nous avons donc choisi de nous attarder sur Arab Propbank (APB), Arab WordNet (AWN) et Arab VerbNet (AVN). Ces ressources ont la particularité d'être *la version en ASM* d'autres ressources conçues d'abord pour la langue anglaise. De plus, ces ressources ont fait l'objet d'articles scientifiques décrivant la méthodologie de travail utilisée pour leur réalisation. Dans notre analyse, nous essayons de nous concentrer le plus possible sur le traitement des UP dans ces ressources.

En ce qui concerne les ressources spécialisées, notre choix a été aussi limité et restrictif. Tout d'abord, nous avons écarté les ressources papier, puisque nous cherchons à discuter des ressources en ligne semblables à celle que nous voulons mettre en place. Ensuite, nous avons choisi, parmi les ressources en ligne, celles qui ont les particularités suivantes : 1) les ressources dont nous avons pu identifier le concepteur; 2) les ressources qui ne se limitent pas à donner des équivalents interlinguistiques; 3) les ressources qui montrent l'utilisation d'un terme dans des contextes; et 4) les ressources qui font état d'une standardisation des termes.

Il est à noter que la facilité d'accéder aux différents services (création des sites web, des forums, des blogues, etc.) qu'offre le web maintenant a permis la création de nombreux dictionnaires généraux et spécialisés en ligne. L'un des problèmes de ces dictionnaires est que, souvent, ils se limitent à donner les équivalents des termes recherchés sans aller plus loin, comme donner des définitions, des contextes ou d'autres renseignements décrivant les propriétés linguistiques des termes. Un autre problème que nous constatons, en ce qui concerne les ressources spécialisées, relève du type d'unité décrit. Souvent ces dictionnaires ne donnent que des unités complexes dans les résultats d'un terme recherché.

Pour les ontologies, nous avons constaté que les travaux sur les ontologies en ASM et en d'autres variantes de la langue arabe sont de nombre important. Par conséquent, nous avons filtré les travaux qui nous intéressent en fonction de leur pertinence par rapport à notre recherche. Nous avons choisi de présenter une ontologie qui concerne les verbes et une autre ontologie qui traite du domaine de l'informatique. La dernière ontologie est un projet qui met l'accent sur une modélisation du lexique.

### 3.2.1 Les ressources générales

#### 3.2.1.1 Propbank

Propbank, ou *The Proposition Bank* (Palmer et al. 2005 : 72), est un projet lexical qui a pour but de fournir un corpus manuellement annoté qui permet le « development of better domain-independent language understanding systems ». L'étude des verbes et de leur structure argumentale dans Propbank est inspirée des travaux sur les verbes de Levin (1993). Le travail consiste à définir un ensemble de rôles sémantiques pour chaque verbe et à annoter chaque occurrence dans les textes du Penn Treebank (Marcus et al. 1993), comme dans les exemples suivants pour le verbe anglais *to offer* (Palmer et al. 2005 : 72) :

... [Arg0 the company] to ... offer [Arg1 a 15% to 20% stake] [Arg2 to the public]. (wsj 0345)  
... [Arg0 Sotheby's] ... offered [Arg2 the Dorrance heirs] [Arg1 a money-back guarantee] (wsj 1928)  
... [Arg1 an amendment] offered [Arg0 by Rep. Peter DeFazio] ... (wsj 0107)  
... [Arg2 Subcontractors] will be offered [Arg1 a settlement] ... (wsj 0187)

Ces renseignements ajoutés aux phrases dans les textes sont importants pour des applications comme l'extraction d'informations, les systèmes de questions-réponses et la traduction automatique. En outre, c'est un moyen de fournir une analyse sémantique plus adéquate. Par exemple, le mot *window* peut apparaître dans des fonctions syntaxiques différentes, comme :

*John broke the window*  
*The window broke*

Dans une analyse syntaxique, *window* serait considéré le COD de la première phrase et le sujet de la deuxième phrase. Cette analyse ne rend pas compte du sens des deux phrases. Donc, une analyse qui illustre le sens de *window* prend la forme suivante :

*Broke (arg0 = John, arg1 = the window)*

*Broke (arg1 = the window)*

Les verbes appartiennent à des classes sémantiques, qui dépendent de leurs propriétés syntaxiques. Ces classes partagent certains composants. Par exemple, dans les deux exemples donnés pour le verbe *break*, les deux phrases sont reliées par une relation d'alternance transitive/intransitive, dite causative/inchoative. L'objectif visé par Propbank dans ce genre d'analyse des verbes est de fournir un niveau de représentation de ces unités et de constituer un corpus annoté, ce qui permet d'effectuer des études empiriques sur des questions liées aux verbes.

L'annotation dans Propbank se fait en fonction des verbes et de leur structure argumentale. Chaque argument est numéroté. Par exemple, les phrases produites pour le verbe *to offer* sont annotées avec l'étiquette « arg » indiquant un argument. L'étiquette « arg0 » désigne l'Agent prototypique, alors que l'étiquette « arg1 » désigne le Patient ou le Thème prototypique. Dans l'exemple suivant nous illustrons l'annotation des arguments à partir du verbe anglais *to accept* (Palmer et al. 2005 : 75).

*Arg0: Acceptor*

*Arg1: Thing accepted*

*Arg2: Accepted-from*

*Arg3: Attribute*

*Ex:[Arg0 He] [ArgM-MOD would] [ArgM-NEG n't] accept [Arg1 anything of value] [Arg2 from those he was writing about]. (wsj 0186)*

Dans l'exemple, la structure des arguments correspond à un sens spécifique du verbe *to accept*. Cela constitue ce qu'on appelle un *roleset*. Le *roleset* peut être associé à un ensemble de cadres syntaxiques qui comprennent les variations syntaxiques possibles dans les limites de ce *roleset*. Cette association entre le *roleset* et les cadres syntaxiques s'appelle *Frameset*. Dans le contexte du verbe *to accept*, le verbe appartient au *Frameset accept.01* (rendant compte du sens de « take willingly »).

Ainsi, la fiche d'un verbe dans le site officiel de Propbank a la forme suivante (figure 22).

**Predicate: *accept***

**Roleset id:** *accept.01* , *take willingly*, **Source:** , **vncls:** *13.5.2 29.2-1-1 77*, **framnet:** *Receiving Respond\_to\_proposal*

*accept.01:* Frames file for 'accept' based on survey of initial sentences of big corpus. VerbNet entries approve-77, characterize-29.2-1-1, and obtain-13.5.2.

**Roles:**

**Arg0-PAG:** *acceptor* (vnrole: 29.2-1-1-agent, 13.5.2-agent, 77-agent)  
**Arg1-PPT:** *thing accepted* (vnrole: 29.2-1-1-theme, 13.5.2-theme, 77-theme)  
**Arg2-DIR:** *accepted-from* (vnrole: 13.5.2-source)  
**Arg3-PRD:** *attribute* (vnrole: 29.2-1-1-attribute)

**Example: transitive agentive**

The companies prevent [them]-1 from \*trace\*-1 accepting our superior cash offer.

**Arg0:** \*trace\*-1 (= them)  
**Rel:** accepting  
**Arg1:** our superior cash offer

Figure 22 : Fiche du verbe *to accept* dans Propbank (Propbank 2015)

### 3.2.1.2 The Arabic Propbank (APB)

Propbank existe pour l'anglais, le chinois, l'allemand et l'ASM. Pour l'ASM, le projet de Propbank (APB) fait face à un certain nombre de défis, compte tenu de la morphologie riche et complexe de cette langue. Ce projet (Diab et al. 2008) est basé sur *The Arabic Treebank, ATB* (Maamouri et al. 2004) et l'analyseur morphologique *AraMorph* (Buckwalter 2002). L'annotation dans la version ASM de Propbank suit la même méthodologie que dans la version anglaise. Ainsi, une phrase comme (Diab et al. 2008) :

*m\$rwE AlAmm AlmtHdp frD mhlp nhAyp l AtAHp AlfrSp AmAm qbrS*  
 مشروع الأمم المتحدة فرض مهلة نهائية لإتاحة الفرصة أمام قبرص  
 (Le projet des Nations Unies a imposé une période finale de grâce comme une dernière opportunité à Chypre).

peut être annotée de la façon suivante :

*[m\$rwE AlAmm AlmtHdp]*<sub>ARG0</sub> *[frD]*<sub>PREDICATE</sub> *[mhlp nhA}yp]*<sub>ARG1</sub> *[l AtAHp AlfrSp AmAm qbrS]*<sub>ARGM-PRP</sub>

Dans cette phrase, la structure argumentale prend la forme : *Arg0, prédicat, Arg1* et *Argm-prp*, où *m\$rwE AlAmm AlmtHdp* (مشروع الأمم المتحدة) (le projet des Nations Unies) est le

Arg0 (sujet de la phrase), *mhlp nhA}yp* (مهلة نهائية) (une période finale de grâce) est Arg1 (l'objet) et *l AtAHp AlfrSp AmAm qbrS* (لإتاحة الفرصة أمام قبرص) (une dernière opportunité à Chypre) est un Argm-PRP (argument de But, ou *purpose argument*).

Tout comme Propbank pour l'anglais, l'APB se base sur les principes expliqués dans la sous-section précédente. L'annotation est fondée sur les *rolesets* et les *Framesets*. Donc, un verbe comme *AqAm* (أقام) (dans le sens de *mettre en place* ou *séjourner*) peut avoir les *Framesets* suivants :

*w tm Akt\$Af Awl mnzl AqAmt fyh EA}lp bAx*

و تم اكتشاف أول منزل أقامت فيه عائلة باخ

(La première maison à laquelle la famille de Bach a résidé a été découverte)

**PRÉDICAT:** *AqAmt* (résider)

**ARG0** est « *experiencer* »: *EA}lp bAx* (la famille de Bach)

**ARG1** est « *location* »: *fyh* (à laquelle)

*w kAnt mjmwEp AslAmyp mslHp AqAmt HJzA whmyA EIY AlTryq Aldwly rqm wAHd*

و كانت مجموعة إسلامية مسلحة أقامت حاجزا وهميا على الطريق الدولي رقم واحد

(un groupe islamiste armé a mis en place un faux barrage sur la route internationale numéro 1)

**PRÉDICAT:** *AqAmt* (mettre en place)

**ARG0** est « *organizer* »: *mjmwEp AslAmyp mslHp* (groupe islamiste armé)

**ARG1** est « *event* »: *HJzA whmyA* (faux barrage)

**ARGM-LOC:** *EIY AlTryq Aldwly rqm wAHd* (sur la route nationale numéro 1)

La méthodologie suivie dans la création de l'APB comprend les étapes suivantes (Diab et al. 2008 : 3469) :

1. Une fois le prédicat choisi, les annotateurs examinent un échantillon de presque 50 occurrences dans le corpus afin d'avoir une idée de sa structure syntaxique. Dans le cas où le prédicat a une faible fréquence, les annotateurs consultent des dictionnaires et le web;
2. La création d'un Frameset est basée sur les propriétés sémantico-syntaxiques des verbes; les arguments sont choisis en fonction du sens du verbe. Si le verbe est polysémique, chacun des sens correspond à un Frameset différent;
3. Chaque occurrence du verbe est étiquetée avec ses arguments en fonction des Frameset.

Dans la création de l'APB, les concepteurs ont dû faire face à un certain nombre de contraintes. Comme nous l'avons signalé à plusieurs reprises dans notre thèse, la langue arabe se distingue par un système morphologique riche et complexe. Par conséquent, parmi les contraintes se trouve celle de la représentation des verbes dans les fiches. Deux stratégies ont été envisagées : soit utiliser le lemme, soit faire appel à la racine du verbe<sup>18</sup> (Diab et al. 2008 : 3469). Finalement, le choix a porté sur le lemme du verbe pour la simple raison que le lemme encode d'une manière exacte les informations morfo-sémantiques, contrairement à la racine à laquelle plusieurs lemmes peuvent correspondre. Cependant, les concepteurs font remarquer que les racines sont ajoutées pour chaque Frameset. Cela pourrait aider à relier les Frameset d'une même racine.

Une autre contrainte dans la conception de l'APB est celle de la voix du verbe. Puisqu'en ASM les voyelles courtes (les diacritiques ou encore la vocalisation) ne sont pas marquées à l'écrit, un verbe à la voix passive est problématique lors de l'annotation. Diab et al. (ibid. : 3470) donnent l'exemple suivant :

فتح الباب بالمفتاح  
*ftH AlbAb bAlmftAH*  
*(il a ouvert la porte avec la clef)*

Dans cet exemple, le verbe (puisque non vocalisé) peut signifier : 1) quelqu'un a ouvert la porte avec la clef; et 2) la porte a été ouverte avec la clef. Si le verbe est à la voix passive, alors la structure argumentale de la phrase comprend les éléments suivants : *ftH* (فتح) (a été ouverte) est le prédicat, *AlbAb* (الباب) (la porte) est Arg1 et *bAlmftAH* (بالمفتاح) (avec la clef) est Arg2. Dans le cas inverse, le verbe a la structure : *ftH* (فتح) (a ouvert) est le prédicat, l'Arg0 est le sujet compris dans la structure du verbe, donc marqué par « NONE » dans ATB, Arg1 est *AlbAb* (الباب) (la porte) et Arg2 est *bAlmftAH* (بالمفتاح) (avec la clef).

En plus des contraintes que nous avons présentées, d'autres difficultés surviennent lors de l'annotation de certaines constructions morfo-syntaxiques. Bien entendu, ces difficultés

---

<sup>18</sup> Dans la tradition dictionnaire de la langue arabe, les entrées dans les dictionnaires sont les racines des mots et non pas leur lemme.

relèvent, comme discuté au début de cette sous-section, du système morphologique complexe de l'ASM. Parmi ces difficultés se trouve l'annotation des compléments des verbes comme dans l'exemple *{iEotabara (اعتبر) (considérer)}*, surtout dans des constructions comme *yaEotabiruwhA (يعتبروها) (ils l'ont considéré)*. Les clitiques du verbe expriment l'Arg0 et l'Arg1. Donc, une telle construction peut être annotée de la façon suivante :

*(S (VP (yaEotabiruwh | يعتبرو | ils considèrent)  
 [(NP-SBJ (-NONE-)) Arg0]  
 [(S (NP-SBJ (hA | ها | le)*

Dans l'exemple, le sujet (ils) est compris dans le verbe, donc c'est un syntagme nominal non exprimé linguistiquement (NONE). Quant à l'objet, c'est un syntagme nominal morphologiquement attaché au verbe au moyen d'un pronom (ce qui constitue une particularité de la langue arabe par rapport aux autres langues, comme le français et l'anglais).

Finalement, nous donnons dans la figure 23 un exemple d'une entrée dans APB. C'est le verbe *qabila (قبل) (accepter)*. La phrase analysée est :

*Alr}ys AlErAqy SdAm hsyn yqbl bEwdp Almft\$yn  
 الرئيس العراقي صدام حسين يقبل بعودة المفتشين  
 Le président irakien Saddam Hussein accepte le retour des inspecteurs*

**Predicate: *qabil-a***

**Roleset id: 01 , to accept**

**Arg0:** *acceptor*  
**Arg1:** *thing accepted - marked by ب*  
**Arg2:** *from*

**Frame:**

. " وأعرب \* في مؤتمر صحفي مشترك مع نائب الرئيس الأميركي ديك تشيني الأربعاء عن اعتقاده ان الرئيس العراقي صدام حسين " سيقبل 1-\*\* بعودة المفتشين

**Arg0:** *\*\*-1 < الرئيس العراقي صدام حسين >*  
**Gloss:** *-NONE- < the Iraqi president Saddaam Husein*  
**Arg1:** *بعودة المفتشين*  
**Gloss:** *the return of the inspectors*

Figure 23 : Frameset du verbe *qabila (قبل) (accepter)* dans APB (Arabic Propbank 2015)

### 3.2.1.3 WordNet

WordNet (Fellbaum 1998, 2006 et Miller et al. 1990) est une ressource lexicale qui se veut une combinaison d'une pratique lexicographique traditionnelle et d'une approche nouvelle basée sur l'informatique dans le traitement des informations lexicales. Elle est conçue pour la langue générale anglaise et est disponible pour plusieurs langues, dont l'ASM et le français.

WordNet organise les informations lexicales en fonction des sens des mots, plutôt que leurs formes. De ce point de vue, WordNet peut être considéré comme un thésaurus (Miller et al. 1990 et Fellbaum 2006). Cependant, il se distingue des thésaurus traditionnels par le nombre fini de relations entre les mots qu'il comprend. WordNet regroupe les parties du discours (noms, verbes, adjectifs et adverbes) dans des ensembles de synonymes cognitifs appelés *Synsets*. La synonymie est la relation dominante dans WordNet. Une autre relation est l'antonymie. Les mots sont également reliés au moyen de relations morphologiques et sémantiques. En outre, dans la conception de WordNet, les deux composants importants sont *l'héritage* et *la réversibilité* (Fellbaum 2006 : 667). L'héritage concerne les relations de type hiérarchique. Par exemple, si *mailbox* et *letter box* sont encodés comme hyponymes de *box*, et *box* est un hyponyme de *container*, donc *mailbox* et *letterbox* sont automatiquement encodés comme hyponymes de *container*. En ce qui concerne la réversibilité, les relations sont encodées dans WordNet seulement une fois entre une paire de Synset ou une paire de mots. Par exemple, si *tree* est encodé comme un méronyme de *forest*, alors *forest* devient automatiquement un holonyme de *tree* (ibid. : 667).

WordNet est un dictionnaire basé sur des principes psycholinguistiques (Miller et al. 1990), dans la mesure où il tente de combler les lacunes dont souffrent les dictionnaires traditionnels de la langue anglaise. Miller et al. (ibid. : 235) considèrent que ces dictionnaires ne mettent pas l'accent sur l'aspect synchronique des lexiques. Cet aspect, issu des développements dans les domaines de la psychologie et de la linguistique, concerne les bases cognitives de la recherche linguistique. De plus, ces bases ont abouti à des découvertes de plusieurs propriétés synchroniques en lexique mental qui peuvent être exploitées en lexicologie (ibid. : 236).



Selon le site officiel de WordNet, la dernière version porte le numéro 3.1 (2015). Cette version comprend les informations suivantes (figure 24).

<b>Number of words, synsets, and senses</b>				
<b>POS</b>	<b>Unique Synsets</b>		<b>Total</b>	
	<b>Strings</b>		<b>Word-Sense Pairs</b>	
Noun	117798	82115	146312	
Verb	11529	13767	25047	
Adjective	21479	18156	30002	
Adverb	4481	3621	5580	
Totals	155287	117659	206941	

<b>Polysemy information</b>			
<b>POS</b>	<b>Monosemous</b>	<b>Polysemous</b>	<b>Polysemous</b>
	<b>Words and Senses</b>	<b>Words</b>	<b>Senses</b>
Noun	101863	15935	44449
Verb	6277	5252	18770
Adjective	16503	4976	14399
Adverb	3748	733	1832
Totals	128391	26896	79450

Figure 24 : Statistiques de la version 3.1 de WordNet (WordNet 2015)

Actuellement, WordNet comprend 117 798 noms, 11 529 verbes, 21 479 adjectifs et 4 481 adverbes. Pour toutes les parties du discours, il y a un total de 117 659 Synsets répartis comme suit : 82 115 Synsets pour les noms, 13 767 Synset pour les verbes, 18 156 Synsets pour les adjectifs et, finalement, 3621 Synsets pour les adverbes.

Dans WordNet, les verbes sont organisés en fonction des relations d'implication (*entailment relations*) (Fellbaum et Miller 1990, Fellbaum 1998, Fellbaum 1990 et Fellbaum 2006). Parmi ces relations, une relation dite *troponymie* a été introduite. Cette relation exprime le fait que  $V_1$  implique  $V_2$ , comme dans la paire *limp-walk*. Le verbe *limp* désigne une manière particulière de marcher (*walk*), par conséquent, *limp* est un troponyme de *walk* (Fellbaum 1990 : 285).

D'autres relations existent aussi, telle que la relation qui relie des paires comme *buy-pay* et *snore-sleep*. Ces relations ne sont pas de type troponymique et sont marquées par une inclusion temporelle exacte, dans le sens où la personne peut dormir avant ou après le ronflement et l'action d'acheter inclut d'autres activités que payer. C'est pour cette raison que la relation de type troponymique est de nature temporelle co-étendue (*co-extensive*) (Fellbaum 1990 : 286), dans le sens où dans le cas des verbes *walk* et *limp*, l'action de marcher est accompagnée de boiter.

Dans WordNet, les verbes sont reliés par une relation d'antonymie. Ces verbes sont aussi reliés par une relation d'implication, sauf que cette relation ne dénote pas une inclusion temporelle. Par exemple, dans le cas des verbes comme *hit*, *miss* et *aim*, les actions sont reliées par un ordre séquentiel, c'est-à-dire que l'action de viser est une condition préalable à celle de rater ou de frapper (ou atteindre). Cette relation d'implication s'appelle *backward presupposition* (Fellbaum 1990 : 289).

La dernière relation est celle de *cause*. Dans cette relation les verbes concernés sont de type *causatif* et *résultatif*, comme dans la paire *give-have*. Finalement, la figure 25 résume les relations discutées.

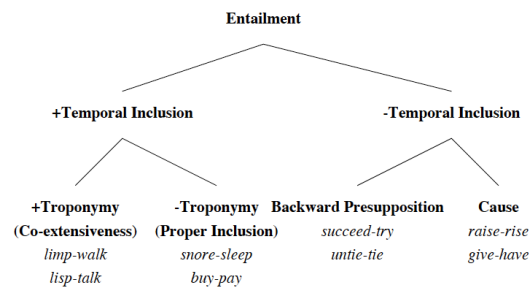


Figure 25 : Quatre types de relations d'implication selon Fellbaum (Fellbaum 1990 : 292)

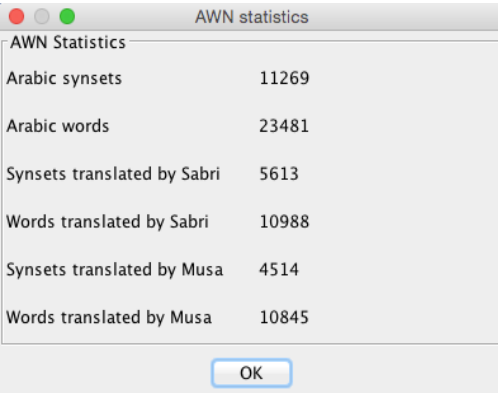
En outre, les verbes dans WordNet sont divisés en fonction de 15 critères sémantiques (Fellbaum 1990) (ou *domaines lexicaux* discutés dans la section 3.1.4). Ces critères sont : les verbes de soins et de fonctions corporels (*verbs of bodily care and functions*), de changement, de cognition, de communication, de compétition, de consommation, de contact, de création, d'émotion, de mouvement, de perception, de possession, d'interaction sociale, et verbes de climat (*weather verbs*).

### 3.2.1.4 The Arabic WordNet (AWN)

Arabic WordNet, ou AWN, (Rodríguez et al. 2008, Black et al. 2006, Elkateb et al. 2006a et Elkateb et al. 2006b) est la version en ASM de WordNet. AWN est basé sur WordNet et est construit en fonction des méthodes suivies dans EuroWorNet (EWN) (Vossen

1998). EWN maximise la compatibilité entre les différents WN et se concentre sur l'encodage manuel d'un ensemble des concepts de base (Elkateb 2006b). AWN est lié directement aux EWN et WordNet, ce qui permet la traduction, au niveau lexical, de l'anglais et d'un certain nombre d'autres langues vers l'ASM. La base sémantique formelle d'AWN est le SUMO (*Suggested Upper Merged Ontology*).

AWN est accessible au moyen d'un navigateur. En téléchargeant ce navigateur, nous avons pu avoir accès aux statistiques de la version la plus récente d'AWN (figure 26)<sup>19</sup>.



AWN Statistics	
Arabic synsets	11269
Arabic words	23481
Synsets translated by Sabri	5613
Words translated by Sabri	10988
Synsets translated by Musa	4514
Words translated by Musa	10845

Figure 26 : Statistiques de la version 2.0 d'AWN

D'après la figure, la dernière version d'AWN porte le numéro 2.0. Cette version comprend 11 269 Synsets et 23 481 mots. De plus, elle comprend des mots et des Synsets traduits. Les Synsets traduits font un total de 10 127 Synsets, tandis que les mots traduits sont au nombre de 21 833.

### *Méthodologie de travail*

Dans les paragraphes qui suivent, nous discutons de la méthodologie de travail dans AWN telle que décrite dans Elkateb et al. (2006a) et Black et al. (2006b). Pour commencer,

---

<sup>19</sup> Il est à noter qu'AWN n'a pas d'interface web à consulter. L'utilisateur doit télécharger la base de données, ou le navigateur AWN, afin de consulter le contenu. Dans les articles que nous avons consultés, les auteurs parlent d'une interface web, mais les liens vers ces interfaces ne sont pas fonctionnels.

nous avons remarqué que la question qui revient toujours lorsqu'il s'agit de construire une ressource lexicale de l'ASM est celle de la complexité du système morphologique de cette langue. Nous avons montré dans la section 3.2.1.2 le genre de problèmes soulevés par la construction d'un Propbank de l'ASM. Pour AWN les mêmes contraintes se présentent.

AWN se base sur deux composants essentiels : le WordNet noyau (*core wordnet*) et l'extension de ce dernier à des concepts plus spécifiques en utilisant d'autres critères (Elkateb et al. 2006b). En ce qui concerne le premier composant, le WordNet noyau, il s'agit d'encoder les *Common Base Concepts* (CBCs) des 12 langues dans EWN et BalkaNet (un WN pour les langues balkaniques) comme Synsets dans AWN. Pour les concepts relatifs à l'ASM, ils sont traduits et ajoutés manuellement. Les Synsets en ASM sont étendus au moyen des relations de type hyperonyme afin de construire une hiérarchie sémantique. Tout comme dans EWN, pour chaque verbe en ASM ajouté, les extensions sont faites à partir des entrées verbales, ce qui inclut les dérivés verbaux, les nominalisations, les noms déverbaux, etc.

En ce qui a trait à l'extension du WordNet noyau, le travail consiste à choisir une couche d'hyponymes. Ce choix est basé sur trois éléments : la connectivité, la pertinence et la généralité. La connectivité concerne le fait qu'AWN doit être aussi connecté que possible par des relations d'hyperonymie et d'hyponymie. La plupart des Synsets d'AWN doivent avoir une correspondance avec ceux de WordNet. La pertinence relève des concepts les plus fréquents. Ces concepts sont prioritaires dans les deux langues (anglais et ASM). Finalement, la généralité concerne les Synsets qui se trouvent au niveau le plus haut dans WordNet. À ce stade, deux procédures sont nécessaires, à savoir la préparation des règles lexicales et morphologiques et le traitement des ressources bilingues disponibles à partir desquelles les lexicographes d'AWN construisent une ressource bilingue qui comprend des informations sur des paires des mots anglais/ASM. Ces informations sont les paires anglais/ASM, la partie du discours, la racine arabe et la fréquence.

Après cette étape, 17 procédures heuristiques sont appliquées (Farreres 2005). Ces procédures sont utilisées dans EWN. L'objectif est d'obtenir une mise en correspondance entre les mots ASM et les Synsets anglais. La dernière étape dans la construction d'AWN est la validation manuelle des données obtenues et de combler les lacunes dans la structure générale d'AWN, comme ajouter des noms d'entités.

### Base de données

La base de données d'AWN comprend les informations suivantes : les items, les mots, les formes et les liens.

- Les items sont les entités conceptuelles, ce qui inclut les Synsets et les classes de l'ontologie (voir paragraphe suivant). Chaque item a son propre identificateur;
- Les mots sont des entités de sens. Chaque mot est associé à un item via son identificateur;
- Les formes sont des entités qui comprennent les informations lexicales;
- Les liens ont pour fonction de relier deux items et donner des spécifications sur le type de ceux-ci. Par exemple, dans AWN, on trouve les liens entre les Synsets d'AWN et de WordNet, ou les liens entre les Synsets et les concepts du SUMO.

### Ontologie

Finalement, les concepts dans AWN sont basés sémantiquement sur une ontologie appelée SUMO (*Suggested Upper Merged Ontology*). Le SUMO est une ontologie formelle qui comprend 1000 termes et 4000 énoncés définitoires. Les Synsets sont mis en correspondance avec les termes compris dans SUMO (figure 27).

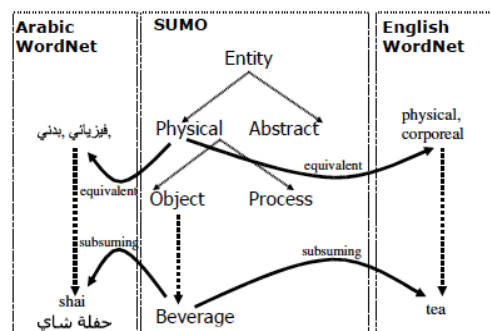


Figure 27 : Correspondance entre AWN et WordNet via SUMO (Elkateb et al. 2006a et Black et al. 2006b).

Dans la figure, nous remarquons que des termes comme *fyzya* (فيزيائي، بدني) et *badany* (فيزيائي، بدني) (physique et corporel) sont liés par une relation de type *équivalence* aux termes anglais *physical*, *corporeal* via l'ontologie SUMO. De plus, un terme comme *Haflap \$Ay* (حفلة شاي) (thé) est lié au terme *tea* via une relation d'inclusion. Cette relation est déterminée dans l'ontologie au moyen des concepts : Physical, Object et Beverage.

### Le navigateur d'AWN (AWN browser frame)

Comme signalé dans cette section, nous n'avons pas trouvé une interface web pour AWN. Toutefois, AWN est téléchargeable en fichier Java ce qui le rend exécutable sur plusieurs plateformes informatiques (figure 28).

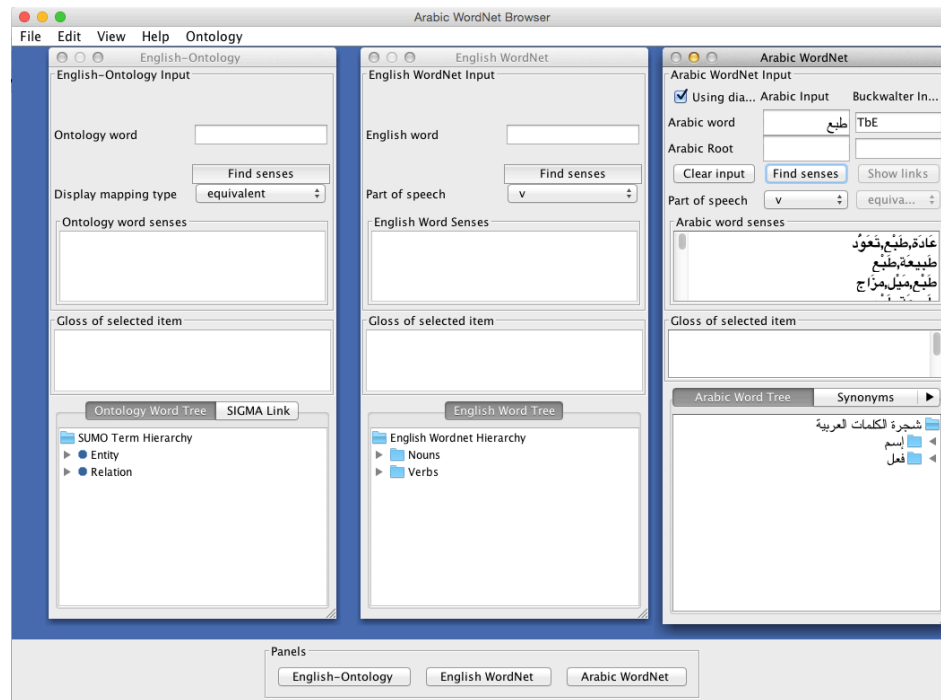


Figure 28 : Arabic WordNet Browser

Le navigateur comprend trois parties : AWN, le WordNet anglais et l'ontologie SUMO. L'utilisateur a la possibilité d'afficher le WN de son choix. La partie consacrée à AWN a ses propres particularités. Par exemple, l'utilisateur peut chercher soit par mot ou par

racine. De plus, il peut interroger un mot soit par caractères arabes ou par forme translittérée. La translittération utilisée est celle de Buckwalter. Enfin, la recherche peut être raffinée en choisissant la partie du discours désirée.

Une requête dans AWN donne le résultat suivant (figure 29).



Figure 29 : Requête dans AWN

En tapant le mot *barmaja* (برمج) (programmer) dans le navigateur, le système donne automatiquement la forme translittérée. Dans la partie « Arabic Word Senses », le système fournit les sens que le mot peut avoir. Dans le cas du mot recherché dans la figure, les deux sens obtenus sont >Ed~a barnAmajaN (أعد برنامجًا) (faire un programme) et *barmaja* (برمج) (programmer). Plus bas, une barre donne trois types d'informations réparties sur trois onglets : l'arbre des mots ASM, les synonymes et l'analyse morphologique. Dans ce genre de requête, les informations qui nous intéressent sont les deux dernières (nous reviendrons sur le premier onglet plus bas). Pour le mot recherché, le système ne donne pas de synonymes.

En ce qui concerne l'analyse morphologique, le système d'analyse utilisé est Aramorph (voir chapitre 5). L'analyse morphologique de l'ASM consiste à montrer ce qui suit : la forme, la translittération, la forme lemmatisée, la forme vocalisée (en ajoutant les voyelles courtes ou les diacritiques), les suffixes et préfixes, la racine et l'équivalent anglais. En outre, si le mot est polysémique ou ambigu, on trouve une liste des analyses morphologiques le plus

probables. Dans le cas de notre exemple, les résultats de l'analyse sont montrés dans la figure 30.

```
SOLUTION #1
Lemma : baromaj
Vocalized as : بَرْمَجْ
Morphology :
  prefix : Pref-0
  stem : PV
  suffix : PVSuff-a
Grammatical category :
  prefix : ʾ PVSUFF_SUBJ:3MS
Glossed as :
  stem : program
  suffix : he/it <verb>|
```

Figure 30 : Analyse d'Aramorph dans le navigateur d'AWN

Le premier onglet dans la barre discuté ci-haut est dédié au glossaire des mots ASM. Ce glossaire comprend deux volets : les noms et les verbes<sup>20</sup>. Cet onglet sert à afficher les correspondances entre AWN et WordNet anglais via l'ontologie. Afin de donner un exemple de son mode de fonctionnement, nous avons choisi le verbe >*qlaEa* (أقلع) (dans le sens de *arrêter, éviter et quitter*). AWN ne donne pas de sens spécialisé comme dans l'exemple de *programmer*. Les synonymes qu'il énumère relèvent tous de la langue générale.

Dans la partie du WordNet anglais, nous trouvons l'équivalent anglais, *forgo*, avec les synonymes disponibles et une définition. Dans la partie ontologique, nous trouvons les mêmes informations, sauf que cette fois-ci l'ontologie montre la correspondance du verbe recherché avec les concepts compris dans l'ontologie et leur définition et le type de relation. Donc, dans le contexte du verbe >*qlaEa* (أقلع) et son équivalent anglais, *forgo*, le concept qui leur correspond est *Intentional Process*. Dans la hiérarchie de l'ontologie, ce concept appartient à l'arbre suivant le chemin : Entity-> Physical-> Process-> Intentional Process. Le type de relation reliant le concept avec le verbe est celle d'inclusion (*Subsuming*).

---

<sup>20</sup> Nous n'avons pas trouvé une façon qui nous permet de rechercher un mot en particulier. Donc, nous devons faire défiler la liste pour en découvrir le contenu.



### 3.2.1.5 VerbNet

La dernière ressource que nous discutons dans cette section est VerbNet. Initialement, VerbNet (Kipper et al. 2000a et Kipper et al. 2000b) est conçu comme une ressource lexicale pour les verbes anglais. Il fournit une description sémantique et syntaxique des classes des verbes telle qu'établie par Levin (1993). Or, ces classes ont été augmentées et enrichies afin de rendre compte de la classification sémantique des verbes. Korhonen et Briscoe (2004) ont proposé une extension aux classes initiales proposées par Levin. Ils ont incorporé 57 nouvelles classes (53 autres classes ont été ajoutées plus tard) qui ont permis d'élargir la couverture de la classification de Levin. La version actuelle de VerbNet porte le numéro 3.2 et comprend 274 classes, 3769 lemmes et 5257 sens des verbes. Les verbes de VerbNet sont mis en correspondance avec les données de certaines ressources lexicales comme Propbank, WordNet, FrameNet et Xtag (Doran et al. 1994). Nous signalons que VerbNet existe pour d'autres langues, notamment la langue portugaise (Brésil). En outre, une version française de VerbNet, appelée « Verb3net », est en développement (Pradet et al. 2014).

Chaque verbe dans VerbNet est décrit au moyen des éléments suivants :

1. Les membres : ce qui comprend la liste des verbes appartenant à une classe ou à une sous-classe. La plupart des membres sont mis en correspondance avec des entrées dans les ressources lexicales déjà mentionnées;
2. Les rôles : ces sont des rôles thématiques attribués à la structure argument-prédicat des membres. VerbNet utilise des rôles thématiques hiérarchiques. Pour chaque classe, les rôles de chaque membre d'un verbe sont spécifiés;
3. Les restrictions : ces sont des restrictions qui donnent davantage d'informations sur la nature des rôles; et
4. Les cadres (*frames*)<sup>21</sup> : ce sont des cadres syntaxiques qui comportent une description syntaxique, les prédicats, des exemples et la correspondance entre les rôles sémantiques et les arguments syntaxiques.

---

<sup>21</sup> Il importe de signaler que les cadres de VerbNet se distinguent de ceux de FrameNet et de la Sémantique des cadres (voir section 4.3).

Dans la figure 31, nous donnons un exemple d'une classe dans VerbNet.

No Comments		<b>create-26.4</b> <small>Members: 28, Frames: 4</small>	<a href="#">POST COMMENT</a>	CLASS HIERARCHY <a href="#">CREATE-26.4</a> <a href="#">CREATE-26.4-1</a>
<b>MEMBERS</b> <span style="float: right;">REF KEY</span>				
AUTHOR (WN 1)	COWRITE	IMPROVISE (WN 1)	PIECE_TOGETHER (G 1)	
COIN (FN 1, 2; WN 1, 2; G 1, 2)	CREATE (FN 1, 2, 3; WN 5, 6; G 1)	INVENT (FN 1, 2; WN 1; G 1)	PRODUCE (FN 1, 2, 3, 4; WN 2; G 1)	
COMPOSE (FN 1, 2; WN 2, 3, 4, 6; G 1, 3)	DERIVE (WN 1; G 2)	LAY (WN 3; G 2)	REBUILD (WN 1; G 1)	
COMPUTE (WN 1; G 1)	DRAFT (FN 1; WN 1; G 1)	MANUFACTURE (FN 1; WN 1; G 1)	RECREATE	
CONCOCT (FN 1, 2; WN 1, 2, 4; G 1, 2)	FABRICATE (FN 1; WN 1; G 1)	MASS-PRODUCE (WN 1)	STYLE (WN 2; G 1)	
CONSTRUCT (FN 1; WN 1, 3; G 1, 2)	FORM (FN 1, 2; WN 4, 5; G 1, 3)	MODEL (WN 1, 2; G 1)	SYNTHESIZE (WN 1; G 1)	
CONTRIVE (FN 1; G 2)	FORMULATE (WN 1, 2; G 1)	ORGANIZE (WN 3; G 2)	WRITE (WN 1, 6, 10, 3; G 1, 4)	
<b>ROLES</b> <span style="float: right;">REF</span>				
<ul style="list-style-type: none"> <li>• AGENT [+HUMAN] [+MACHINE]</li> <li>• RESULT</li> <li>• MATERIAL</li> <li>• BENEFICIARY [+HUMAN]</li> <li>• ATTRIBUTE</li> </ul>				
<b>FRAMES</b> <span style="float: right;">REF KEY</span>				
<b>NP V NP</b>				
EXAMPLE	"David constructed a house."			
SYNTAX	AGENT V RESULT			
SEMANTICS	NOT(EXIST(START(E), RESULT)) EXIST(RESET(E), RESULT) CAUSE(AGENT, E)			
<b>NP V NP PP MATERIAL</b>				
EXAMPLE	"David constructed a house out of sticks."			
SYNTAX	AGENT V RESULT {FROM OUT_OF} MATERIAL			
SEMANTICS	NOT(EXIST(START(E), RESULT)) EXIST(RESET(E), RESULT) MADE_OF(RESET(E), RESULT, MATERIAL) CAUSE(AGENT, E)			

Figure 31 : Classe *create* dans VerbNet (VerbNet 2015)

La classe Create a 28 membres et 4 cadres. Dans la section « Members », on trouve les verbes qui appartiennent à cette classe et, pour chaque verbe, on mentionne la présence du verbe dans d'autres ressources lexicales (comme WordNet, Propbank et FrameNet). La section « Roles » comprend les rôles sémantiques pour les membres. Finalement, la section « Frames » donne les structures syntaxiques. Dans le cas de la classe Create, il existe quatre cadres possibles. Par exemple, on y trouve la structure NP V NP comme dans l'exemple : *David<sub>[NP]</sub> constructed<sub>[V]</sub> a house<sub>[NP]</sub>*. Cette structure a les rôles sémantiques suivants : Agent ~ Result.

### 3.2.1.6 Arabic VerbNet (AVN)

Arabic VerbNet, ou AVN, (Mousser 2010 et 2011) a été développé suivant la méthodologie du VerbNet anglais. L'auteur constate que construire une ressource lexicale des verbes en ASM représente un vrai défi, puisque les travaux portant sur les verbes en ASM n'ont pas pour objectif la construction d'une ressource lexicale pour ces unités (Mousser

2010 : 2675). C'est pour cette raison qu'il estime que l'élaboration d'une base de données pour les verbes permet de mieux cerner leurs propriétés sémantique et syntaxique.

L'auteur expose deux approches dans la construction d'AVN (Mousser 2010 : 2676). La première approche consiste à se baser sur un ensemble de classes lisibles par ordinateur afin de leur associer des membres qui partagent les mêmes propriétés sémantiques et syntaxiques, comme on le fait dans VerbNet anglais. La deuxième approche consiste à se baser sur un corpus afin d'y extraire des classes à l'aide d'algorithmes. Cette extraction est fondée sur les propriétés syntaxiques partagées. AVN fait appel à la première approche. L'auteur a utilisé les classes de verbes telles qu'établies par Levin (1993) et leur extension dans les travaux de Korhonen et Briscoe (2004).

Chaque membre d'une classe est traduit en ASM en ajoutant des relations comme la synonymie et l'hyponymie. La traduction se fait à l'aide de dictionnaires de l'arabe classique. L'utilité de ces dictionnaires réside dans le fait qu'ils donnent des renseignements utiles sur l'étymologie des verbes. De plus, ces renseignements permettent de déterminer la racine et la forme du verbe à l'infinitif. Après cette étape, un verbe prototypique est choisi à partir d'un ensemble de verbes traduits en lui associant tous les cadres correspondants possibles. Tout comme VerbNet, les cadres d'AVN sont décrits au moyen d'un exemple, une structure syntaxique et sémantique montrant les aspects des verbes. Après avoir choisi le verbe prototypique, le reste des verbes est ajouté à la classe en fonction des propriétés partagées avec ce verbe. Dans la figure 32, nous donnons un exemple des étapes expliquées.

Group 1		Groupe 2	
<i>awḥāā</i>	'reveal'	<i>tamtama</i>	'mumble'
<i>hashasa</i>	'swish'	<i>walwala</i>	'make a howl'
<i>hassa</i>	'murmur'	<i>ḡaḡā</i>	'growl'
<i>waswasa</i>	'murmur'	<i>damdama</i>	'burr'
<i>wašwaš</i>	'whisper'	etc.	
<i>hamasa</i>	'whisper'		
etc.			

Figure 32 : Deux groupes obtenus à la suite de la traduction de la classe Manner\_speaking dans VerbNet (Mousser 2010)

La figure 32 présente le résultat de la traduction de la classe Manner\_speaking dans VerbNet en ASM. Cette classe a la structure Agent-Topic-Recipient. Nous remarquons deux

groupes séparés. Chaque groupe comprend les verbes qui partagent les mêmes propriétés sémantico-syntaxiques. Dans le cas du premier groupe, le verbe typique est *waswasa* (وسوس) (*murmurer*). Les verbes qui appartiennent à ce groupe ont tous la même structure sémantique. Par contre, le deuxième groupe atteste d'une différence. Ce groupe ne partage pas les mêmes propriétés, puisque Topic et Recipient n'apparaissent pas.

Dans son travail, Mousser aborde le problème de l'alternance<sup>22</sup> qui peut se produire à la suite de la traduction des verbes vers l'ASM ou après l'application des classes de Levin, puisqu'il risque d'y avoir certaines contraintes liées à la langue même. Dans cette perspective, l'auteur (Mousser 2010 : 2677-2678) trouve que 65 % des alternances trouvées en anglais s'observent aussi en ASM. Par exemple, l'auteur évoque des cas comme l'alternance causative/inchoative et l'alternance dative :

- [1] *qaTaEa Alfal~aHu Al\$ajarap*  
 قطع الفلاح الشجرة  
 (*L'agriculteur a coupé l'arbre*)
- [2] *AinqaTaEat Al\$ajarap*  
 إنقطعت الشجرة  
 (*L'arbre s'est coupé facilement*)

Ces phrases correspondent à :

- [3] *The farmer cuts the tree*  
 [4] *The tree cuts easily*

Dans ces exemples, l'alternance est celle de causative/inchoative. Le verbe ASM *qaTaEa* (قطع) (*couper*) est transitif dans la première phrase. Dans la deuxième, le verbe admet dans sa structure morphologique un préfixe, *A* (!), qui le rend intransitif<sup>23</sup>.

En ce qui concerne l'alternance dative, l'auteur donne les exemples suivants :

<sup>22</sup> Les alternances affectent l'expression syntaxique des arguments d'un verbe. Les alternances sont parfois accompagnées par un changement de sens.

<sup>23</sup> D'après la morphologie des verbes, certains verbes transitifs peuvent devenir intransitifs en ajoutant la lettre *A* (!) au début du verbe.

- [1] *bAEa jamylyN say~Aratuhu liHamad*  
 باع جمال سيارته لحمد  
 (Jamil a vendu sa voiture à Hamad)
- [2] *bAEa jamylyN Hamada say~Aratuhu*  
 باع جمال حمدا سيارته  
 (Jamil a vendu à Hamad sa voiture)

Les phrases correspondent à :

- [3] *Jamil sold a car to Hamad*  
 [4] *Jamil sold Hamad a car*

Si, dans une classe donnée, certains verbes n'ont pas les mêmes alternances que les verbes anglais, ou s'ils n'ont pas d'alternance, une nouvelle classe doit être créée pour ces verbes. Ce genre de classe est appelé *sibling class* (Mousser 2011). La *sibling class* est « a way to resolve the problem arising from the discrepancy between two derivationally related morphological verb forms which participate in the same set of alternations and therefore share the same semantic meaning » (ibid. : 356). Par exemple, la classe *Change of State* a l'alternances causative/inchoative (figure 33 et figure 34).

Class: Change of State			
Members: <i>aşrana</i> 'modernize', <i>haşhaşa</i> 'privatize', <i>awolama</i> 'globalize', <i>arraba</i> 'arabize', etc.			
Roles and Restrictions: Agent [+int.control] Patient Instrument			
Descriptions	Examples	Syntax	Semantics
Basic Intransitive	<i>naşşafa saliym malābisahu.</i> (Salim dried his clothes)	V Agent Patient	<i>cause(Agent, E), state(result(E), Endstate, Patient)</i>
NP-PP	<i>naşşafa saliym malaābisahu biālbūhaār.</i> (Salim dried his clothes with the vapour)	V Agent Patient {bi} Instrument	<i>cause(Agent, E), state(result(E), Endstate, Patient), use(during(E), Agent, Instrument)</i>
Instrument Subject	<i>naşşafa ālbūhaāru ālmālābisa.</i> (The vapour dried the clothes.)	V Instrument Patient	<i>use(during(E), ?Agent, Instrument), state(result(E), Endstate, Patient)</i>
Subclass			

Figure 33 : Class *Change of State* dans AVN (Mousser 2011: 357)

Sibling Class: Change of State			
Members: <i>taaşrana</i> 'INC-modernize', <i>tahaşhaşa</i> 'INC-privatize', <i>taawolama</i> 'INC-globalize', <i>taarraba</i> 'INC-arabize', etc.			
Roles and Restrictions: Agent [+int.control] Patient Instrument			
Descriptions	Examples	Syntax	Semantics
V NPpatient	<i>naşafati ālmālābisahu.</i> (The clothes dried)	V Patient	<i>state(result(E), Endstate, Patient)</i>
PP	<i>naşafati ālmālābisahu biālbūhaār.</i> (The clothes dried with the vapour.)	V Patient Instrument	<i>use(during(E), ?Agent, Instrument), state(result(E), Endstate, Patient)</i>
Subclass			

Figure 34 : Sibling class de *Change of State* dans AVN (Mousser 2011 : 357)

Dans la figure 33, la phrase qui représente l'alternance causative est la suivante :

na\$~afa salymuN malAbisahu  
 نشف سليم ملايسه  
 (Salim a séché ses vêtements)

Dans le cas d'une alternance inchoative, une classe (*sibling*) est créée à part. Dans la figure 34, la phrase ayant cette alternation est la suivante :

na\$afati AlmalAbisu  
 نشفت الملايس  
 (Les vêtements se sont séchés)

Finalement, AVN a une interface web accessible gratuitement. La version la plus récente d'AVN comprend 336 classes, 7748 verbes, 7819 déverbaux, 1399 cadres et 231 sous-classes. Dans la figure 35, nous montrons les résultats obtenus pour la classe *nasaxa* (نسخ) (*copier*).

The screenshot displays the AVN web interface for the class 'nasaxa' (نسخ). The interface is organized into several sections:

- CLASS:** nasaxa-1
- MEMBERS:** A list of 10 members, each with its name, root, deverbale form, and participial form. A 'Get all members' button is located below the list.
- THEMROLES:** A list of roles: AGENT [+ANIMATE], THEME [], and DESTINATION [+CONCRETE].
- FRAMES:** Two frames are shown:
  - BASIC TRANSITIVE:**
    - EXAMPLE: "دوتت الكائبة المرأسلة."
    - SYNTAX: V AGENT THEME
    - SEMANTIC: CREATED\_IMAGE(RERESULT(E), THEME), LOCATION(END(E), THEME, ?DESTINATION), CAUSE(AGENT, E)
  - NP-PP:**
    - EXAMPLE: "دوتت الكائبة المرأسلة في سيجل الشركة."
    - SYNTAX: V AGENT THEME PREP <+Loc|+dest\_conf> DESTINATION
    - SEMANTIC: CREATED\_IMAGE(RERESULT(E), THEME), PREP(END(E), THEME, DESTINATION), CAUSE(AGENT, E)
- SUBCLASSES:** (Empty)
- SIBLING\_CLASSES:** (Empty)

Figure 35 : Cadre *nasaxa* (نسخ) (*copier*) dans AVN (AVN 2015)

La classe *nasaxa* (نسخ) (*copier*) comporte dix membres et deux cadres. Elle a la structure Agent-Theme-Destination. Nous remarquons que les verbes sont présentés sous différentes formes :

- L'infinifit : *nasaxa* (نسخ) (*copier*);
- La racine : *nsx* (نسخ) (l'idée de copier);
- Le nom déverbal : *nasx* (نسخ) (*copie*); et
- Le participe : *nAsix* (ناسخ), (*copié*).

### **3.2.1.7 Les contraintes liées à la constitution d'une ressource lexicale de l'arabe standard moderne**

Après avoir discuté des trois ressources que nous estimons être les plus importantes pour la langue générale, certaines questions se posent en ce qui concerne les défis rencontrés lors de leur conception. Il s'agit principalement des enjeux et des contraintes soulevées par une langue sémitique lorsqu'il est temps de la traiter de la même manière que les langues européennes. Parmi ces questions se trouve la complexité du système morphologique de la langue (voir section 2.3). Dans les trois travaux que nous venons de présenter, les auteurs attestent des difficultés soulevées par une telle complexité. Par exemple, Mousser (2010 : 2675) et Diab et al. (2008 : 3469) évoquent la question des entrées dans un dictionnaire de l'ASM. Dans les ressources classiques et historiques de cette langue, les entrées correspondent aux racines des lexèmes, c'est-à-dire qu'il n'existe pas d'organisation alphabétique du lexique, surtout lorsqu'il s'agit des verbes. Or, cette approche, pourtant utile pour regrouper toutes les entrées de la même racine, n'est pas toujours utile pour les relations sémantiques plus complexes. Bien entendu, le choix a porté sur la forme lemmatisée des lexèmes pour les deux ressources (AVN et APB).

Pour leur part, Elkateb et al. (2006b : 30-31) dressent une liste de six facteurs à prendre en compte afin de produire un système sur la base de l'analyse morphologique qui est efficace sur les plans linguistique et computationnel. Les six facteurs sont :

1. les schèmes, les racines et les radicaux sont pris en compte;
2. l'existence des formes morphologiques dépend des autres formes morphologiques;
3. la forme peut avoir plusieurs fonctions morphologiques;
4. un mot est la suite d'une racine, manuellement encodée, et un schème diacritisé;
5. l'extraction des racines s'effectue par l'élimination des affixes, afin d'identifier la forme de base d'un mot diacritisé; et
6. l'utilisation du système de translitération.

Cependant, les auteurs ne précisent pas les détails les plus complexes de la question. Parmi les six facteurs, nous n'abordons que le sixième, car il nous semble important. Certains outils de traitement automatique n'acceptent pas de textes écrits en caractères arabes. Ces systèmes sont basés sur une technique développée par Buckwalter (Habash et al. 2007) qui consiste à simplifier le traitement morphologique de l'ASM en convertissant les textes en caractères latins selon des normes spécifiques. Cette technique a permis d'obtenir des résultats considérables en ce qui concerne l'analyse morphologique, sauf qu'elle est incapable de traiter adéquatement la question des formes ambiguës (voir section 5.2.4.4).

Il est à noter qu'AWN et APB sont tous deux basés sur la translitération Buckwalter et le système d'analyse morphologique est Aramorph. AVN est une exception. Mousser (2010 et 2011) ne fait pas état d'une analyse morphologique basée sur Aramorph et le système de translitération n'est pas basé sur celui de Buckwalter. Toutefois, en consultant la ressource en ligne, nous avons remarqué que toutes les entrées verbales sont translittérées, et l'utilisateur peut rechercher un verbe par sa forme translittérée. Dans le chapitre 5, nous nous attardons sur cette question et nous donnons des détails sur des résultats obtenus en utilisant ce système.

Finalement, AVN et AWN ne sont pas basés sur un corpus, mais sur la traduction des données de VerbNet et de WordNet. En ce qui concerne APB, étant donné qu'il est basé sur le Penn Arabic Treebank, l'encodage et le traitement des données se font à partir d'un large corpus.



## 3.2.2 Les ressources spécialisées

Dans cette section, nous nous attardons sur les ressources spécialisées de l'ASM. Comme nous l'avons mentionné au début de la section 3.2, nous avons choisi de présenter les ressources suivantes : WIPO Pearl, ARABTERM et Unified Medical Dictionary (UMD).

### 3.2.2.1 WIPO Pearl

WIPO Pearl est une ressource terminologique multilingue créée par la *World Intellectual Property Organization* (WIPO). WIPO est une organisation internationale, basée à Genève, en Suisse, qui se charge des droits de propriétés intellectuelles. La base de données décrit les termes de dix langues, à savoir l'allemand, l'anglais, l'ASM, le chinois, le français, le coréen, le japonais, le portugais, le russe et l'espagnol. Nous n'avons pas trouvé de documents qui décrivent les méthodes de travail utilisées pour compiler cette ressource, mais sur le site officiel de l'organisation il existe suffisamment d'informations qui décrivent la façon dont la terminologie est traitée et gérée.

Actuellement, la base comprend 16 000 concepts et 105 000 termes sur une variété de domaines scientifiques et techniques. Elle est destinée aux professionnels de l'industrie langagière, aux professionnels des brevets et aux spécialistes des domaines.

WIPO Pearl est basé sur un corpus composé de documents de brevets. Ce corpus s'appelle PATENTSCOPE. Le corpus donne accès aux demandes internationales selon le traité de coopération en matière de brevets (PCT) ainsi qu'aux documents de brevet des offices de brevets nationaux ou régionaux participants. Ce corpus est actualisé et enrichi tous les jours. En outre, intégrés à cette base de données, se trouvent plusieurs outils de traduction automatique et des moteurs de recherche basés sur corpus. Par exemple, on y trouve une fonction de recherche qui permet aux utilisateurs de rechercher des documents relatifs aux brevets dans différentes langues. Cette fonction s'appelle *Cross-Lingual Information Retrieval* (CLIR). Dans la recherche d'information, la recherche fait appel à un corpus parallèle via un système d'analyse statistique. Cette fonction de recherche n'est pas encore compatible avec l'ASM.

Le site de WIPO ne fait pas état de la façon dont les termes sont extraits (des critères ou des démarches). Toutefois, des informations existent quant à la validation des termes. La validation terminologique des termes passe par un indice de fiabilité (*term reliability score*). Il s'agit d'un processus qui attribue une évaluation à un terme donné en fonction des critères suivants :

- À défaut d'une source fiable, la validation d'un terme passe par le service de traduction WIPO;
- À défaut d'une source fiable, la validation d'un terme passe par le service de traduction WIPO ainsi qu'un expert du domaine;
- En présence d'une source fiable, le terme est validé par le service de traduction WIPO;
- En présence d'une source fiable, le terme est validé par le service de traduction WIPO ainsi qu'un expert du domaine.

Cependant, nous n'avons pas trouvé plus de renseignements en ce qui concerne la façon exacte dont les termes sont validés. Le seul renseignement que nous avons trouvé concerne l'idée d'une « source fiable » : « any text in which a term occurs that was originally drafted by a subject field expert in the language in question; reliable sources include, for example, patents, scientific and technical articles and textbooks, academic papers »<sup>24</sup>.

WIPO propose un certain nombre de renseignements qui aident l'utilisateur à mieux comprendre certains concepts utilisés dans le dictionnaire. Parmi ces renseignements, nous citons :

- *Common name* : un synonyme d'un terme scientifique utilisé dans la langue générale;
- *Geographical variant* : une variation orthographique d'un terme qui appartient à une zone géographique dans laquelle une variante d'une langue est parlée;

---

<sup>24</sup> <http://www.wipo.int/reference/en/wipopearl/guide.html#rs>

- *Obsolete* : un terme qui n'est plus utilisé;
- *Proposed term* : un terme associé à un concept pour lequel on n'a pas trouvé une désignation dans une source fiable. Ou il peut s'agir d'un nouveau terme créé par le service de traduction de WIPO;
- *Term note* : des informations étymologiques et phraséologiques sur un terme, son usage et ses collocations.

L'interface permet à l'utilisateur d'effectuer une recherche sur des termes ou sur des concepts. De plus, plusieurs filtres sont mis à la disposition des utilisateurs afin de raffiner leur recherche. Par exemple, en tapant le terme *barnAmij* (برنامج) (programme), nous avons obtenu les résultats suivants (figure 36).

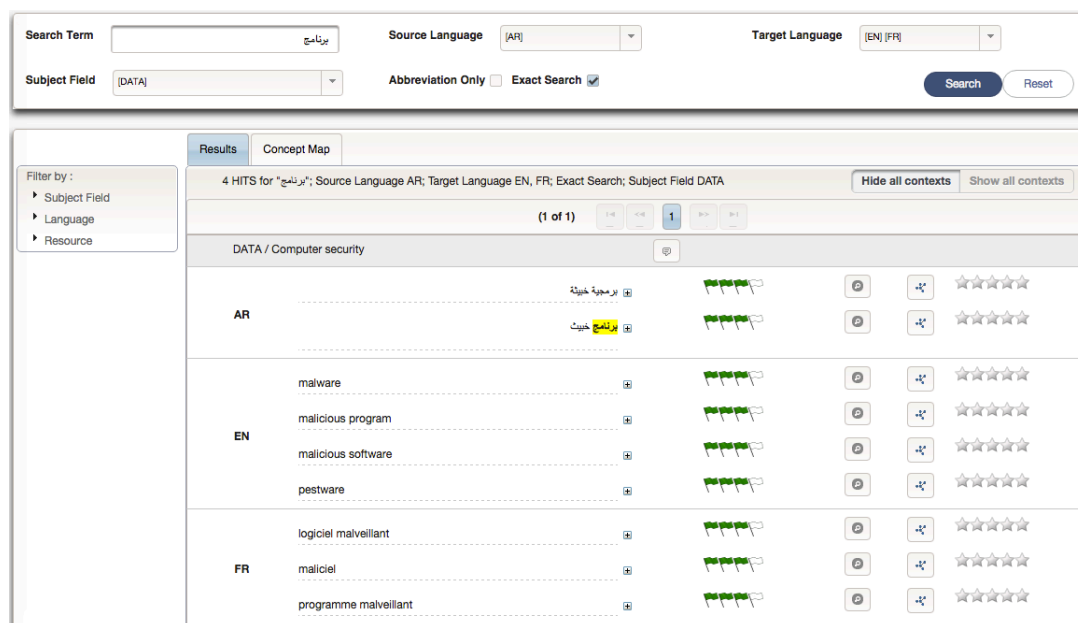


Figure 36 : Recherche d'un terme arabe dans WIPO (WIPO 2015)

En spécifiant la langue arabe comme langue source et le français et l'anglais comme langues cibles, le système a proposé quatre entrées comprenant le terme *programme* (français). Dans la figure, chaque terme est accompagné d'un indicateur de fiabilité. Les termes obtenus à la suite de la requête sont tous des termes complexes. Nous remarquons que même si nous

avons raffiné la recherche pour avoir le terme exact, *barnAmij* (برنامج) (programme), le système a donné des termes complexes contenant le terme *programme*, *barmajy~ap xabyv* (برمجية خبيثة) (*logiciel malveillant*) et *barnAmij xabyv* (برنامج خبيث) (*programme malveillant*). Nous remarquons aussi que ces termes sont sémantiquement liés, c'est-à-dire qu'il concerne les programmes malveillants. Nous avons essayé à plusieurs reprises d'obtenir des résultats pour des verbes, mais le système n'a donné que des termes complexes nominaux. Nous trouvons logique que la base de données ne contienne pas ce type d'unités, puisqu'il est basé sur un corpus comprenant des documents des brevets. Dans ce genre de système, l'intérêt (vu l'organisation) porte principalement sur les noms d'entités.

En cliquant sur le signe « + », nous obtenons les contextes associés à chaque terme (figure 37).



Figure 37 : Contexte pour un terme dans WIPO (WIPO 2015)

Nous avons évoqué plus tôt dans cette section le genre de contextes qui intéressent les terminologues de cette ressource. Dans la figure, le contexte est extrait d'une revue scientifique (Iraqi Academic Scientific Journals). La date et le nom de l'auteur sont également donnés. Dans le cas où le contexte est extrait d'un document de brevets stocké dans la base de données, ceci sera indiqué via un hyperlien<sup>25</sup>.

Finalement, en cliquant sur le pictogramme, l'utilisateur est dirigé vers une représentation graphique qui montre le concept et les relations qu'il entretient avec d'autres concepts dans le domaine (figure 38).

<sup>25</sup> Par exemple, pour l'équivalent anglais, le lien dirige vers un document de brevet pour la compagnie BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY.



Figure 38 : Représentation graphique des concepts dans WIPO (WIPO 2015)

Les courbes servent d'identificateurs de relations entre les concepts. Dans la figure, une relation de type *associatif* est établie entre *barnAmij xabyv* (برنامج خبيث) (*programme malveillant*) et *barnAmij Did AlfayrwsAt* (برنامج ضد الفيروسات) (*programme antivirus*). En outre, la couleur des points à côté de chaque concept désigne le type de concept même. Dans la figure, la couleur rouge associé à *barnAmij Did AlfayrwsAt* (برنامج ضد الفيروسات) (*programme antivirus*) désigne un concept qui appartient à un autre domaine.

### 3.2.2.2 ARABTERM (ArT)

L'autre ressource dont nous discutons dans cette section est ARABTERM (ArT). ArT est un dictionnaire technique multilingue, allemand, anglais, ASM et français, issu d'une collaboration entre ALESCO (*Arab League Educational, Cultural and Scientific Organization*) et BMZ (*German Federal Ministry for Economic Cooperation and Development*). La particularité de l'ArT est le fait qu'il est basé sur les résolutions prises par l'ALESCO et le Bureau de coordination de l'arabisation (BCA) en ce qui concerne la standardisation terminologique. De ce fait, ArT est un projet indépendant visant à rendre compte de mesures afin de standardiser la terminologie technique de l'ASM. Ces mesures de standardisation sont issues des résolutions prises aux conférences générales de l'arabisation

tenues<sup>26</sup> régulièrement par l'ALESCO. D'après ArT, ces mesures ont débouché sur la publication par l'ALESCO d'un dictionnaire unifié<sup>27</sup>.

Puisqu'il s'agit d'un nouveau projet, ArT n'a pas encore fait l'objet d'articles scientifiques. Pour cette raison, pour décrire cette ressource, nous nous appuyons sur les informations disponibles dans le site du dictionnaire.

L'approche dans la création d'ArT comprend quinze étapes. Ces étapes constituent la base du travail terminologique menant à la création de ce qu'ArT appelle les *Dictionnaires unifiés*<sup>28</sup>. Ces étapes comprennent, entre autres, la validation des termes par des experts qui font partie d'un comité inter-arabophone (ce comité est constitué d'experts venus de 22 pays où l'ASM est une langue officielle). En outre, une étape comprend une validation linguistique de la part du BCA. Cependant, il n'existe pas de mention en ce qui concerne la source des termes compilés. Le site d'ArT ne fait pas état d'un corpus ou d'une traduction des termes existants dans d'autres ressources terminologiques.

Il convient de dire qu'ArT donne beaucoup d'informations sur un travail que nous estimons être très important. Ce travail concerne l'arabisation et la création terminologique. Par exemple, pour ce qui est de la création terminologique, la méthodologie repose sur les quatre points suivants :

1. régularité et usage commun;
2. facilité d'usage;
3. commodité et compatibilité;
4. possibilité de dérivation.

---

<sup>26</sup> Le site d'ArabTerm ne fait pas état des lieux où sont tenues ces conférences.

<sup>27</sup> Il est à noter que la mention de *dictionnaire unifié* porte à confusion, étant donné qu'ArT ne précise pas le type de terminologie décrite dans ce dictionnaire. Le seul dictionnaire que nous avons trouvé portant la mention *unifié* est *Unified Medical Dictionary*, pour la terminologie médicale. Donc, nous ne sommes pas certain que c'est ce dictionnaire qui est visé par leur mention.

<sup>28</sup> <http://www.arabterm.org/index.php?id=107&L=2>

En ce qui a trait à la méthodologie d'arabisation, ArT<sup>29</sup> donne dix-huit principes dont les suivants :

1. Établir un lien approprié entre le sens linguistique du terme et sa signification idiomatique, mais sans qu'il contienne tout son sens scientifique.
2. Extrapoler et relancer l'héritage linguistique arabe, notamment les anciens termes arabisés qui restent utilisables à ce jour.
3. Utiliser les outils linguistiques pour générer une nouvelle terminologie scientifique tels que la dérivation et la localisation.
4. Préférer les termes en arabe classique aux termes arabisés.
5. Éviter l'usage de la terminologie familière et informelle, sauf lorsqu'elle est partagée entre plusieurs dialectes arabophones. Dans ce cas, elle peut être mise entre parenthèses par exemple.
6. Préférer les termes qui acceptent la dérivation à ceux qui ne l'acceptent pas.
7. Préférer la forme singulière du terme, parce qu'elle aide à faciliter la dérivation, la combinaison, la localisation, etc.
8. Localiser ou arabiser, en cas de besoin, les termes de nature globale, les termes d'origine grecque ou latine, les noms des scientifiques, et les éléments ou composés chimiques.

Un dernier point qui s'ajoute à la liste concerne l'arabisation des termes étrangers. ArT préconise ce qui suit<sup>30</sup> :

1. Privilégier et adopter le terme facile à prononcer lorsque ces prononciations dans les langues étrangères sont différentes.
2. Changer la forme du terme afin qu'il soit conforme aux règles de la langue arabe.
3. Considérer le terme arabisé un terme arabe qui soumet les règles de la langue.
4. Redresser et corriger les mots arabes qui ont été altérés par les langues étrangères, en adoptant leurs formes originales.
5. Ajuster les termes en général, et les termes arabisés en particulier, afin d'assurer qu'ils soient précis et prononcés correctement.

ArT a une interface utilisateur conviviale. D'abord, le système peut détecter automatiquement la langue de la requête. De plus, habituellement, dans chaque ressource lexicale de l'ASM, un clavier arabe virtuel est ajouté à la barre de recherche afin de faciliter la saisie pour les utilisateurs qui n'en disposent pas. Dans ArT, une extension, qui s'appelle *Yamli* (le verbe arabe pour *dicter*), est intégrée afin de convertir les caractères latins en caractères arabes (figure 39).

---

<sup>29</sup> <http://www.arabterm.org/index.php?id=88&L=2>

<sup>30</sup> <http://www.arabterm.org/index.php?id=88&L=2>



Figure 39 : Extension *Yamli* dans ArT (ArabTerm 2015)

Dans la figure 39, nous avons saisi le terme *barmajap* (برمجة) (*programmation*), l'extension affiche des propositions pour des mots en caractères arabes qui correspondent au terme saisi en caractère latins. Nous signalons qu'il ne s'agit pas d'une extension pour la translittération Buckwalter. L'extension comprend un autre système de translittération au nom de *Arabizi*.

Comme nous l'avons mentionné dans cette section, ArT est un dictionnaire multilingue. Les résultats d'une recherche s'affichent dans les quatre langues du dictionnaire. L'utilisateur a la possibilité de raffiner sa recherche en choisissant le domaine (ou le *volume*, selon l'appellation d'ArT) et la catégorie. Par exemple, dans le volume Génie Électrique, on peut choisir parmi trois catégories différentes. Ces catégories se divisent en sous-catégories. Par exemple, la catégorie « régulation » du volume Génie Électrique a des sous-catégories comme algorithmes de régulation, comportement dérivateur, comportement intégrateur et comportement proportionnel. Actuellement, le dictionnaire comprend un volume pour l'informatique, mais il est en cours de développement.

Une recherche dans le dictionnaire peut donner des résultats comme ceux qui ont été reproduits à la figure 40.

★ ○	<a href="#">processus d'apprentissage [m.]</a>	training process	Einlernvorgang [m.]	برمجة تعليمية
★	<a href="#">programmation en atelier de machines CN (f.)</a>	shop floor programming	Werkstattprogrammierung von NC-Maschinen (f.)	برمجة الورشة لماكينات التحكم الرقمي
★	<a href="#">programmation (f.)</a>	programming	Programmierung (f.)	برمجة
★	<a href="#">programmation (f.) absolue</a>	absolute programming	Absolutprogrammierung (f.)	برمجة مُطلقة

Figure 40 : Partie des résultats pour le terme *barmajap* (برمجة) (*programmation*) dans ArT (ArabTerm 2015)



Pour le terme recherché, *barmajap* (برمجة) (*programmation*), nous avons obtenu 82 résultats, tous étant des termes complexes (le terme recherché fait partie des résultats obtenus). À côté de chaque terme, nous distinguons deux formes : un cercle et un astérisque. Le cercle indique qu'il s'agit d'un terme en ASM « unifié et validé lors du Congrès d'Arabisation organisé par l'Organisation Arabe pour l'Education, la Culture et les Sciences (ALECSO) affiliée à la Ligue Arabe ». Quant à l'astérisque, il indique que le terme est validé.

En cliquant sur le terme, l'utilisateur est redirigé vers une page, où l'on peut avoir accès à une définition (figure 41).



Figure 41 : Définition d'un terme dans ArT (ArabTerm 2015)

Les mots en caractère gras sont des termes (**programmes** et **code source**). Par ailleurs, des liens hypertextes dirigent vers des entrées dans le dictionnaire.

### 3.2.2.3 Unified Medical Dictionary (UMD)

*Unified Medical Dictionary* (UMD) est un dictionnaire médical multilingue, allemand, anglais, ASM, français, espagnol et farsi. La première édition de ce dictionnaire date des années 1960 et elle a été produite par l'Union des médecins arabes à Bagdad, en Iraq. L'un des objectifs principaux de ce dictionnaire était la normalisation de la terminologie médicale entre les pays arabophones. En 1973, l'Organisation mondiale de la santé (OMS) a pris en charge la troisième édition du dictionnaire ainsi que son enrichissement et son développement. ALESCO, Union des médecins arabes et le conseil des ministres de santé arabes ont participé au développement de ce dictionnaire. Pour cette raison, un comité de spécialistes a été créé au bureau régional de l'OMS au Moyen-Orient pour la collecte et la validation des termes médicaux en ASM. La quatrième édition du dictionnaire comprend 150000 termes en anglais

et en ASM. La plupart de ces termes sont disponibles pour l'allemand, le farsi, le français et l'espagnol.

Nous n'avons pas trouvé d'informations en qui concerne les détails de la méthodologie de travail dans UMD et il n'existe pas d'information autre que celle qui donne un aperçu de son histoire. Il est à noter que ce dictionnaire existe en trois formats : papier, informatisé (CD-ROM) et en ligne. La version web du dictionnaire est reproduite à la figure 42.



Figure 42 : Interface web d'UMD (UMD 2015)

L'interface utilisateur sur le web propose un certain nombre d'options. Par exemple, dans la liste de gauche, l'utilisateur peut choisir la partie du discours, la traduction, la racine (dans le cas des termes ASM), les hyponymes, l'abréviation, etc. Dans la liste de droite, l'utilisateur peut choisir la langue source et la langue cible. Une particularité de ce dictionnaire par rapport aux autres ayant fait l'objet des sections précédentes est qu'il propose des verbes spécialisés dans le domaine médical. Dans la figure 43, nous donnons l'exemple du verbe *yuEAlij* (يعالج) (*traiter*).

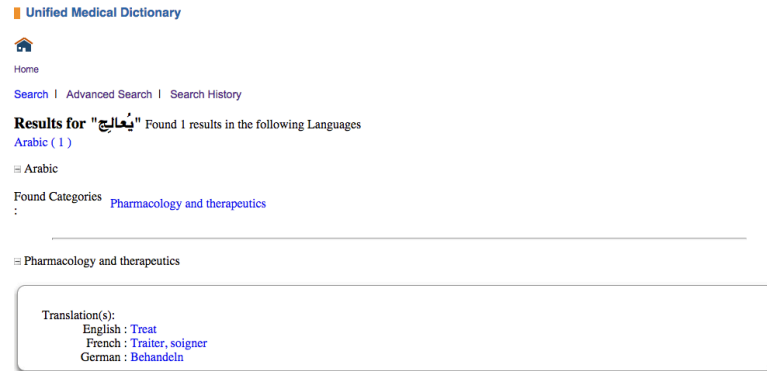


Figure 43 : Verbe *yualij* (يعالج) (*traiter*) dans UMD (UMD 2015)

Dans les résultats de la recherche, nous remarquons qu'UMD définit la catégorie à laquelle appartient le terme. Dans la figure, on voit la catégorie de Pharmacology and Therapeutics. Ensuite, selon les langues sources choisies, UMD donne les équivalents du terme, *treat* (anglais), *traiter* et *soigner* (français) et *behandeln* (allemand). Toutefois, le dictionnaire ne propose pas de collocations ou d'autres informations relatives au verbe.

Une autre particularité de cette ressource est l'entrée du verbe *yujry Eamaliy~ap* (يجري عملية) (*opérer* ou littéralement *faire une opération*). Ce verbe spécialisé est composé de deux éléments *yujri* (يجري) (*faire*) et *Eamaliy~ap* (عملية) (*une opération*). Il est considéré comme complexe, puisqu'en ASM l'on ne peut pas dériver un verbe de *opération*.

Finalement, UMD se distingue des autres ressources spécialisées par le fait qu'il vocalise les termes (les termes sont tous diacritisés).

### 3.2.3 Les travaux ontologiques sur la langue arabe

Après avoir discuté des ressources générales et spécialisées de l'ASM, nous trouvons important de mentionner quelques ontologies. La construction d'ontologies est un domaine qui retient de plus en plus l'attention des spécialistes, qu'ils soient terminologues, informaticiens ou experts désirant construire une infrastructure pour le contenu textuel arabe sur le web. Dans ce qui suit, nous discutons de quelques travaux que nous jugeons importants pour nous. Nous mettons l'accent sur les travaux qui portent sur le verbe et sur notre domaine de recherche

(l'informatique) dans la construction des ontologies. De plus, nous présentons un dernier travail qui se distingue par une méthodologie pour une modélisation du lexique.

### **3.2.3.1 Ontologie basée sur les verbes**

Dans cette sous-section, nous exposons les travaux de Belkredim et Meziane (2008) et Belkredim et El Sebai (2009). Les auteurs proposent une ontologie basée sur les verbes, ou *DERivational ARabic Ontology Based on Verbs* (DEAR-ONTO). Leur travail est original par rapport aux autres travaux qui se basent sur des ontologies existantes et conçues initialement pour les langues européennes, comme l'anglais et le français. L'ontologie en question tente de répondre aux besoins en ce qui concerne la création d'applications pour le traitement automatique de l'ASM. Les auteurs croient qu'afin de développer un système computationnel efficace pour cette langue, il est recommandé d'avoir recours à de bonnes descriptions des verbes (Belkredim et El Sebai 2009 : 46).

Dans cette approche, le verbe est traité en fonction de ses trois composants donnant des informations morphologiques (voir section 2.3) : la racine, le schème et le radical. L'objectif est de voir de quelle façon l'on peut se servir de ces informations afin d'obtenir des connaissances sémantiques et contextuelles. Par conséquent, à chaque dérivé d'un verbe est assigné une classe distincte, ensuite cette classe fera partie d'une ontologie.

L'ontologie a plusieurs niveaux. Le premier niveau comprend la classe des dérivés nominaux. Cette classe a un certain nombre de sous-classes : les noms déverbaux, le participe actif (ou le nom de l'agent), le participe passif (ou le nom du patient), les noms d'espace, les noms de temps, les noms d'instruments et les adjectifs. Les adjectifs constituent un cas particulier. Ils se subdivisent en quatre sous-classes.

Le deuxième niveau dans l'ontologie comprend les dérivés verbaux. Dans ce niveau, deux types de verbes sont pris en compte, à savoir les verbes à racine trilitère et les verbes à racine quadrilitère (voir section 2.3 pour la notion de « racine »). En se basant sur les schèmes générés par ces racines, on peut avoir un certain nombre de dérivés. À partir des verbes à racine trilitère, on peut avoir douze dérivations primaires. Quant aux verbes à racine quadrilitère, on peut avoir jusqu'à trois dérivations de ces verbes. Finalement, le troisième

niveau comprend les formes fléchies des verbes. Ces formes sont le temps des verbes, passé, présent et impératif, la voix active et la voix passive.

Cette ontologie, étant destinée à donner une structure formelle de l'ASM, peut servir d'infrastructure pour plusieurs applications (Belkredim et Meziane 2008) que nous résumons dans ce qui suit :

1. la construction automatique d'ontologies;
2. l'analyse morphologique de l'ASM;
3. la compréhension de l'ASM;
4. l'enseignement de l'ASM.

Les auteurs ne font pas état d'une application pratique de leur modèle. Ils ne présentent qu'une réflexion générale qui ne semble pas basée sur des données réelles. Pour cette raison, dans les articles consultés sur le modèle proposé, nous n'avons pas trouvé une évaluation de l'ontologie en question. De plus, une ontologie basée sur les racines de mots peut s'avérer inexacte dans certains cas, puisqu'on peut avoir deux mots ayant la même racine, mais qui ne partagent pas un lien sémantique.

### **3.2.3.2 Ontologie pour l'extraction du contenu des blogues**

Al-Safadi et al. (2011) proposent une ontologie qui exploite le contenu spécialisé (domaine de l'informatique) des blogues en ASM. Les auteurs partent du constat que les moteurs de recherche actuels ne supportent pas l'ASM d'une manière adéquate. Ils estiment que ces moteurs sont limités aux mots clés et ils ne prennent pas en compte la sémantique du contenu. De plus, les ontologies disponibles pour l'ASM ne sont pas construites sur des données extraites du web pour tester la faisabilité de leurs modèles (ibid. : 42). Pour cette raison, l'objectif est de construire une ontologie qui est capable de représenter la connaissance spécialisée du domaine de l'informatique à partir des blogues. Cette ontologie a pour objectif de fournir l'infrastructure pour des applications futures pour l'ASM, puisque les ontologies actuelles sont toutes conçues pour la langue anglaise.

### *Une étape de prétraitement*

Étant donné que l'ontologie en question s'appuie sur les blogues, un prétraitement doit être fait. Donc, les données collectées sont filtrées en éliminant les chiffres et les mots vides. Ensuite, les termes sont choisis en fonction de leur occurrence dans les textes. Le terme est retenu s'il a au moins trois occurrences dans un texte. Les résultats obtenus prennent en compte la présence de trois types de termes :

- des termes qui relèvent de l'ASM, comme *mudaw~anap* (مدونة) (blogue);
- des noms des marques, comme *blAkbyry* (بلاكبيرى) (BlackBerry); et
- des noms étrangers qui n'ont pas d'équivalent en ASM, comme *jafaskrybt* (جافاسكريبت) (*Javascript*).

L'importance de la prise en compte de ces types des termes vient du fait que dans une ontologie basée sur le web, il est important d'englober la diversité du lexique utilisé pour que cette ontologie puisse refléter l'usage des termes dans le domaine.

### *Construction de l'ontologie*

La conception de l'ontologie s'appuie sur WESMO (Fensel 2001). La particularité de cette ontologie est l'approche qu'elle adopte. Étant donné qu'elle est basée sur les blogues, elle met l'accent sur les questions auxquelles l'ontologie est censée répondre et les besoins des utilisateurs. La construction de l'ontologie passe par les étapes suivantes : l'extraction des termes, la découverte des synonymes, l'obtention des concepts, l'extraction des hiérarchies des concepts, la définition des relations entre les concepts et la déduction des règles et des axiomes. À la suite de ces étapes, 110 classes ont été obtenues (figure 44).

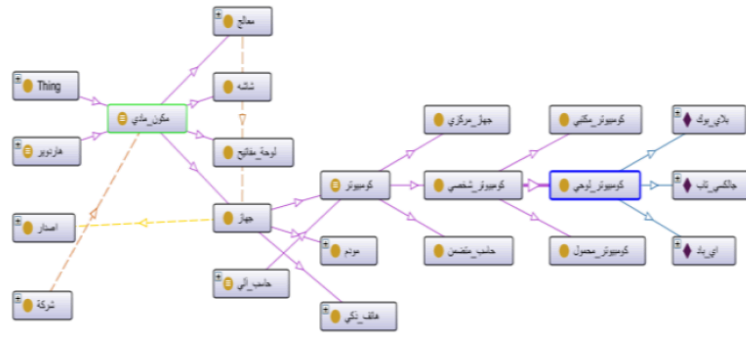


Figure 44 : Sous-graphes représentant des classes dans une ontologie de l'informatique (Al-Safadi et al. 2011)

Afin d'établir les relations entre ces classes, une hiérarchie avec des relations ontologiques est constituée. Au total, 48 relations de différents types ont été obtenues. Parmi ces relations, on peut trouver :

- Une relation d'héritage, comme *jihAz* (جهاز) (*appareil*) est une sous-classe de *mukaw~in mad~iy* (مكون مادي) (*matériel*);
- Une relation de type *is-a*, comme *hArdwyr* (هاردوير) (*hardware*) est une classe équivalente de *mukaw~in mad~iy* (مكون مادي) (*matériel*);
- Une relation de méronymie, comme *Antirnit* (إنترنت) (*Internet*) est une partie de *wyb* (وب) (*web*);
- Une relation reliant deux termes coordonnés, les relations entre les nœuds est bidirectionnelle (X et Y partagent un hyperonyme), comme *kumbywtar maHmwl* (كومبيوتر محمول) (*ordinateur portatif*) et *kumbywtar lawHiy* (كمبيوتر لوحي) (*tablette*).

Enfin, cette ontologie a été encodée au moyen du Web Ontology Language (OWL). Pour ce faire, un éditeur d'ontologie a été utilisé (Protégé).

### 3.2.3.3 Arabic Ontology Project

L'Arabic Ontology Project<sup>31</sup>, ArabicOntology, (Jarrar 2006 et 2011) est un projet initié en 2010 à l'Université de Birzeit, en Palestine. Le projet a pour but de représenter d'une manière formelle les concepts des termes en ASM. L'Arabic Ontology peut être considérée comme un WordNet en ASM, mais il se distingue de ce dernier par les relations bien formalisées qu'il établit entre les concepts et les principes strictement ontologiques sur lesquels il se base.

#### *La construction d'ArabicOntology*

La construction d'ArabicOntology se fait en cinq étapes (Jarrar 2011).

#### *Étape 1*

Au cours de cette étape, l'ontologue procède à la collecte des concepts et des définitions à partir de dictionnaires généraux et spécialisés. La collecte se fait manuellement. Cependant, une étape de filtrage automatique succède à la collecte. Le problème rencontré lors de la collecte des données des dictionnaires arabes est la structure de ces derniers. La plupart des dictionnaires de la langue arabe se concentrent sur la morphologie de mots (la recherche dans ces dictionnaires se fait selon la racine). Donc, les dictionnaires recherchés sont ceux qui donnent des informations relatives à la sémantique des unités lexicales. Pour cette raison, trois conditions président au choix d'un dictionnaire :

1. Le dictionnaire doit être basé sur la sémantique des unités lexicales et non pas sur leur morphologie;
2. Les sens des unités lexicales doivent être clairement exprimés; et
3. Les définitions doivent être de bonne qualité et bien construites.

---

<sup>31</sup> <http://sina.birzeit.edu/ArabicOntology/>



### *Étape 2*

À partir des données obtenues à l'étape 1, l'ontologue restructure manuellement les définitions. Cette restructuration se fait selon les cinq critères ontologiques suivants (Jarrar 2006 : 499) :

1. La définition doit commencer avec le type supérieur/principal du concept. Par exemple, « Search engine: A computer program that... »;
2. La définition doit avoir la forme d'une proposition, dans le sens où elle propose au lecteur des connaissances déductives pour qu'il puisse construire l'image du concept. Par exemple, *search engine* peut être défini comme « A computer program that enables users to search and retrieve documents or data from a database or from a computer network... » au lieu de « A computer program for searching the internet »;
3. La définition doit se concentrer sur les propriétés intrinsèques qui distinguent un concept d'un autre;
4. La définition doit comporter des exemples;
5. La définition doit être cohérente avec les axiomes formels dans l'ontologie, c'est-à-dire que le contenu de la définition ne doit pas contredire les axiomes formels et vice versa; et
6. La définition doit être claire, adéquate et facile à comprendre.

### *Étape 3*

Cette étape consiste à mettre en correspondance (*mapping*) les définitions obtenues avec leurs équivalents dans WordNet (anglais). Pour ce faire, un logiciel a été développé. Ce logiciel recherche, pour chacune des définitions obtenues, une définition correspondante dans WordNet. De plus, ce logiciel a pour tâche de repérer les définitions répétitives dans les ressources arabes.

#### Étape 4

À la suite de l'étape 3, on procède à l'établissement des relations conceptuelles entre les définitions afin d'aboutir à un arbre conceptuel. Cette étape dépend des étapes 1 et 2, dans le sens où chaque concept a un père (SuperType) et chaque père a plusieurs enfants (SubType). De plus, l'établissement des relations conceptuelles dépend aussi de l'étape 3. En effet, la mise en correspondance entre ArabicOntology et WordNet a pour but d'extraire, et par la suite déduire, des relations conceptuelles. Jarrar (2011) signale que le fait d'extraire les relations conceptuelles de WordNet ne constitue pas une traduction des concepts, mais une déduction logique. Finalement, afin de vérifier l'exactitude des résultats obtenus, une méthodologie dite OntoClean est utilisée.

#### Étape 5

Cette dernière étape relie les concepts et les définitions obtenues à l'étape 2 et les liens obtenus à l'étape 4 avec les concepts de niveau élevé (*Top-Level Concepts*) de l'ASM. Ces concepts sont un arbre conceptuel constitué de dix niveaux, environ 400 concepts (figure 45).

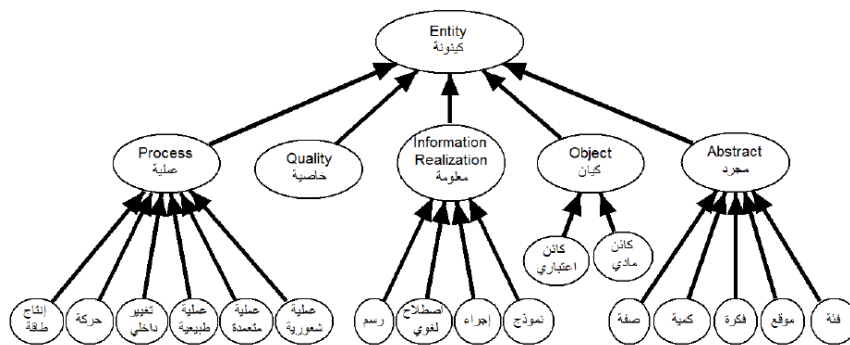


Figure 45 : Partie de niveau élevé de ArabicOntology (Jarrar 2011)

L'ArabicOntology est un projet à long terme. Jarrar (2011) souligne que le projet est en cours et il est toujours enrichi par de nouveaux concepts. De plus, le projet se distingue des autres ressources ontologiques de l'ASM en ce qui concerne la non-dépendance sur la traduction des autres ontologies (par exemple AWN).

### **3.3 Vers une nouvelle méthodologie**

Nous avons présenté dans cette section un ensemble de ressources spécialisées et ontologiques pour l'ASM. Nous avons mis l'accent sur leurs principales caractéristiques, tout en montrant les méthodologies adoptées pour leur construction. Bien que notre travail s'inscrive dans une démarche qui partage un certain nombre de points communs avec ces ressources, nous nous distinguons des ressources discutées par ce qui suit.

Dans un premier temps, notre projet est entièrement basé sur un corpus moderne et diversifié de l'ASM de l'informatique. Nous avons montré que la seule ressource spécialisée basée sur corpus est celle de WIPO. Cependant, la compilation d'un tel corpus représente une tâche difficile, notamment pour déterminer ses frontières et ses limites, puisqu'il s'agit d'un corpus ouvert (de nouveaux textes y sont ajoutés constamment). Il s'appuie sur des documents de brevets et des articles scientifiques. Néanmoins, dans le cas de l'ASM, où la standardisation des termes est une question encore épineuse, le choix des termes reste une question capitale.

Dans un second temps, notre projet se distingue par le fait qu'il fait appel à une méthodologie faisant état de critères définis pour l'extraction des termes, leur validation et l'encodage des données. Les trois ressources spécialisées discutées (WIPO, ArT et UMD) ne font pas état de la démarche suivie dans l'extraction des termes. Or, cette étape constitue un grand défi pour une langue complexe comme la langue arabe (voir chapitre 5). Dans notre projet, nous avons opté pour un extracteur automatique existant auquel nous avons intégré l'ASM.

Dans un troisième temps, notre projet s'appuie sur un cadre théorique (chapitre 4) qui permet d'inclure des informations sémantiques et syntaxiques sur les termes. Nous avons remarqué une absence quasi totale de telles informations dans les ressources spécialisées. À notre avis, un travail terminologique (et terminographique) doit mettre l'accent sur les propriétés sémantico-syntaxiques des termes : leur intégration à une ressource spécialisée permet de mieux gérer diverses propriétés des termes, notamment de mieux décrire la polysémie.

Enfin, notre projet a la particularité de traiter les unités prédicatives dans un domaine de spécialité. Cela constitue une distinction majeure, puisque cette catégorie lexicale est encore très peu décrite dans les ressources terminologiques.

## 4. Cadre théorique

Dans ce chapitre, nous présentons le cadre théorique de notre travail de recherche. Nous commençons d'abord par donner un aperçu historique où nous abordons un modèle qui a précédé la théorie de la Sémantique des cadres dans la littérature linguistique, puis, nous expliquons les concepts théoriques de base de cette théorie. Les dernières parties de ce chapitre concernent les applications de la Sémantique des cadres en langue générale et en langue de spécialité. Nous concluons en justifiant notre choix.

### 4.1 La révolution cognitive et les travaux sur les verbes

La Sémantique des cadres est une théorie qui relève d'un domaine plus large, à savoir la linguistique cognitive (Talmy 2011). L'introduction de la cognition en linguistique a commencé dans les années 1950 avec la parution de *Syntactic Structures* (Chomsky 2002) de Noam Chomsky, paru pour la première fois en 1957. Cette œuvre est à l'origine de nombreux travaux sur le rôle de la cognition en linguistique (Miller 2003 : 142-143). Les idées de Chomsky à cette époque apparaissent dans une période où un certain nombre de disciplines étaient en cours de développement, à savoir l'informatique, la psychologie et l'intelligence artificielle. Bien que *Syntactic Structures* ne soit pas une étude de la linguistique cognitive telle qu'on la connaît aujourd'hui, l'œuvre a ouvert une perspective sur le rôle joué par la cognition dans l'acquisition de la langue et la construction de la grammaire (Lightfoot 2002 : vi). Suivant la révolution cognitive des années 1950, les années 1970, 1980 et 1990 ont vu apparaître des travaux pionniers sur la linguistique cognitive, notamment ceux de Fauconnier (1985), Fillmore (1975, 1976), Lakoff (1987, 1992), Langacker (1987, 1991), et Talmy (2000). À cette époque, l'idée centrale de ces travaux était la représentation linguistique des structures conceptuelles (Talmy 2011 : 622).

En ce qui concerne Fillmore, partant de ses travaux sur la syntaxe transformationnelle, l'auteur s'est intéressé aux propriétés distributionnelles des verbes (Fillmore 1965), surtout en ce qui concerne la substituabilité des mots tout en préservant le sens des énoncés dans des structures semblables aux cadres syntaxiques (Cienki 2007 : 171). Ces recherches ont permis à

Fillmore d'intégrer la sémantique à l'étude des verbes. À ce propos, Fillmore (1982 : 114) constate que :

I began to believe that certain kinds of groupings of verbs and classifications of clause types could be stated more meaningfully if the structures with which verbs were initially associated were described in terms of the semantic roles of their associated arguments.

Ce constat de Fillmore résume le rôle de la sémantique dans la description des verbes, celui des arguments des verbes comme outils de classification, et l'intérêt des rôles sémantiques auxquels on peut associer les arguments. Dans ce qui suit, nous nous attardons sur ces questions tout en évoquant les bases historiques de la Sémantique des cadres.

## 4.2 De la syntaxe structurale à la Sémantique des cadres

Le premier emploi de notions importantes comme « actant » et « circonstant » peut être retracé dans l'ouvrage du linguiste français Tesnière, *Éléments de syntaxe structurale* (1959). D'après Tesnière, l'unité de base de la syntaxe est la phrase qui présente le déroulement d'un procès dont le verbe est le nœud : « l'objet de la syntaxe structurale est l'étude de la phrase » (Tesnière 1959 : 11). Ainsi, la syntaxe structurale s'appuie sur des relations asymétriques et binaires entre les unités lexicales. Ces relations prennent une forme de « dépendance ». Dans son œuvre, Tesnière (ibid. : 11-13) fait observer que :

La phrase est un *ensemble organisé* dont les éléments constituants sont les *mots*. [1.2] Tout mot qui fait partie d'une phrase cesse par lui-même d'être isolé comme dans le dictionnaire. Entre lui et ses voisins, l'esprit aperçoit des *connexions*, dont l'ensemble forme la charpente de la phrase. [1.3] Les connexions structurales établissent entre les mots des rapports de *dépendance*. Chaque connexion unit en principe un terme *supérieur* à un terme *inférieur*. [2.1] Le terme supérieur reçoit le nom de *régissant*. Le terme inférieur reçoit le nom de *subordonné*.

Par exemple, une phrase comme *les petits ruisseaux font les grandes rivières* peut être représentée comme suit (figure 46).

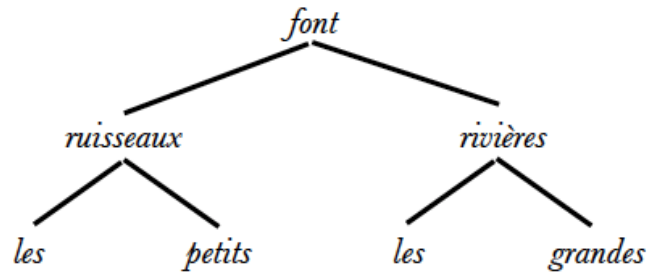


Figure 46 : Structuration de la phrase selon Tesnière (Tesnière 1959 : 11)

Cette représentation va à l'encontre de la grammaire traditionnelle qui représente la phrase de manière linéaire (figure 47).

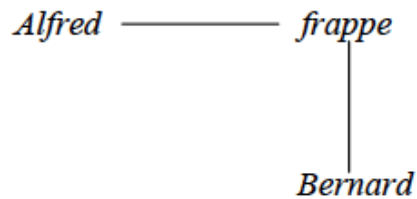


Figure 47 : Structuration linéaire de la phrase d'après la grammaire traditionnelle (ibid. : 3)

D'après la grammaire traditionnelle, la phrase *Alfred parle* est composée de deux éléments, à savoir *Alfred* et *Parle*. Toutefois, d'après Tesnière, la phrase comporte trois éléments :

[...] une phrase du type *Alfred parle* n'est pas composée de deux éléments 1° *Alfred*, 2° *parle*, mais bien de trois éléments, 1° *Alfred*, 2° *parle* et 3° la connexion qui les unit et sans laquelle il n'y aurait pas de phrase. Dire qu'une phrase du type *Alfred parle* ne comporte que deux éléments, c'est l'analyser d'une façon superficielle, purement morphologique, et en négliger l'essentiel, qui est le lien syntaxique (ibid. : 11).

Les principes de la syntaxe structurale de Tesnière se sont avérés utiles pour d'autres linguistes, notamment en langue de spécialité. Par exemple, nous avons montré dans la section 3.1.1 comment Gross (1994) et Gross et Vivès (2001) considèrent que l'unité minimale de

sens est la phrase et non pas le mot. Les auteurs sont d'avis qu'une telle considération garantit une description linguistique qui rend compte de différents emplois possibles d'une UL et, surtout, la désambiguïsation des sens d'une unité polysémique. Nous avons montré également le type d'analyse axée sur le verbe chez Mathieu-Colas (2002), où l'auteur attribue des traits sémantico-syntaxiques aux arguments des verbes.

Le verbe est donc envisagé du point de vue de son environnement qui se compose d'actants (les participants à l'action) et de circonstants (les éléments de temps, espace, etc.). Si le sujet d'un verbe est le « prime » actant, on peut dire que le complément d'objet direct est le « second » actant, et le complément d'objet indirect en est le « tiers », alors que les compléments circonstanciels sont les circonstants. D'après la syntaxe structurale de Tesnière, le verbe en tant que nœud de la phrase exprime une scène dramatique dans laquelle les éléments essentiels sont le procès et les acteurs.

En fonction de cette distribution des actants et des circonstants où le verbe règne sur la scène, le verbe fait l'objet d'une classification en fonction du nombre d'actants qu'il admet (ce qu'on appelle la *valence*) : on trouve les verbes avalents (avec zéro actant), les verbes monovalents (avec un actant), les verbes divalents (avec deux actants) et les verbes trivalents (avec trois actants). Le positionnement des actants est déterminé par le rôle sémantique qu'ils jouent par rapport au verbe.

Partant du principe que « la syntaxe structurale a pour objet de révéler la réalité structurale profonde qui se cache derrière l'apparence linéaire du langage sur la chaîne parlée » (Tesnière 1959 : 11), Tesnière s'est penché sur l'analyse des relations entre les mots. Il en distingue trois : connexion, translation et jonction. La relation qui nous intéresse ici est celle de connexion. Dans cette relation, le verbe, qui est le nucléus supérieur ou le nœud d'après la terminologie tesnièreenne, représente le régissant, alors que l'actant prend celle de nucléus inférieur, ou subordonné.

Les idées de Tesnière ont eu une grande résonance en linguistique. Chez Fillmore, et comme nous l'avons mentionné au début de cette section, les travaux sur les verbes et leurs propriétés syntaxiques avaient commencé en s'appuyant sur les principes de la grammaire



transformationnelle de Chomsky. En intégrant la sémantique, les travaux de Fillmore ont pris une autre direction. Fillmore (1982 : 114) constate que :

I had become aware of certain American and European work on dependency grammar and valence theory, and it seemed clear to me that what was really important about a verb was its 'semantic valence' (as one might call it), a description of the semantic role of its arguments. Valence theory and dependency grammar did not assign the same classificatory role to the 'predicate' (or 'VP') that one found in transformationalist work (see, e.g. Tesnière 1959); the kind of semantic classifications that I needed could be more complete and sensible, I believed, if, instead of relying on theoretically separate kinds of distributional statements such as 'strict subcategorization features' and 'selectional features', one could take into account the semantic roles of all arguments of a predication, that of the 'subject' being simply one of them.

Fillmore a traduit ses idées dans une théorie appelée la « grammaire des cas » (1968), qui peut se résumer à une syntaxe basée sur des *cas profonds* (*deep cases*). Les cas chez Fillmore sont les rôles sémantiques que l'on peut associer aux arguments (ou actants) de verbes. La théorie de valence et la grammaire de dépendance, quant à elles, font partie de la description donnée aux verbes sur le plan syntaxique et sémantique, surtout en ce qui concerne le nombre d'actants qu'une unité verbale peut avoir et la centralité du verbe qui a supplanté la division binaire de la phrase entre sujet et prédicat (Osborne 2013).

En ce qui concerne les cas, Fillmore (1968 : 24-25) en propose six :

Agentive (A), the case of the typically animate perceived instigator of the action identified by the verb.

Instrumental (I), the case of the inanimate force or object causally involved in the action or state identified by the verb.

Dative (D), the case of the animate being affected by the state or action identified by the verb.

Factitive (F), the case of the object or being resulting from the action or state identified by the verb, or understood as a part of the meaning of the verb.

Locative (L), the case which identifies the location or spatial orientation of the state or action identified by the verb.

Objective (O), the semantically most neutral case [...] things which are affected by the action or state identified by the verb.

Dans les phrases suivantes, Fillmore (ibid. : 25) donne des exemples pour chacun des cas :

29. John opened the door.
30. The door was opened by John.
31. The key opened the door.
32. John opened the door with the key.
33. John used the key to open the door.
34. John believed that he would win.
35. We persuaded John that he would win.
36. It was apparent to John that he would win.
37. Chicago is windy.
38. It is windy in Chicago.

Dans les exemples 29 et 30, *John* correspond au cas Agentif. Dans les exemples 31, 32 et 33, *key* est un Instrument. Dans les exemples 34, 35 et 36, *John* correspond au cas Datif. Finalement, dans les exemples 37 et 38, *Chicago* est un Locatif. Ces cas représentent des jugements que les êtres humains sont capables de poser par rapport aux réalités qui les entourent. Bien évidemment, la liste reste non exhaustive, puisque d'autres cas peuvent y être ajoutés.

Dans son analyse des cas, Fillmore part de deux principes : la centralité de la syntaxe, c'est-à-dire que l'on part des concepts syntaxiques pour décrire les formes, et l'importance des catégories cachées (*covert categories*). Ces catégories ne se manifestent pas formellement, donc on pourrait penser qu'elles sont purement sémantiques, mais elles ont, en effet, une pertinence syntaxique (Fillmore 1968 : 23-24). Donc, chez Fillmore, le cas est conçu comme une relation syntaxico-sémantique à deux niveaux : la structure de surface qui se manifeste par des formes casuelles, par l'emploi des prépositions ou par l'ordre des éléments, et la structure profonde où la phrase est constituée d'un prédicat lié à un ou plusieurs cas (figure 48).

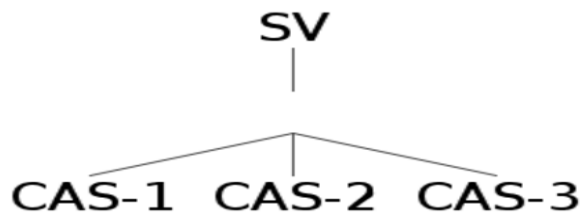


Figure 48 : Structure profonde composée de plusieurs cas

D'après Fillmore (1977), la structure profonde (*deep structure*) est liée à la sémantique interne. Cette sémantique concerne la nature sémantique de la structure interne de la phrase. De plus, la sémantique interne se situe sur l'axe syntagmatique de la phrase et non pas sur l'axe paradigmatique, puisqu'elle concerne des relations qu'entretiennent les éléments de la phrase entre eux. Fillmore discute de ces détails comme suit :

Within semantics in general, the notion of deep cases is a part of what might be called INTERNAL, as opposed to EXTERNAL, semantics; that is, it concerns, not the semantics of truth or entailment or illocutionary force, but rather the semantic nature of the inner structure of a clause. Within internal semantics, the concern is SYNTAGMATIC rather than PARADIGMATIC; that is, deep cases are among the types of semantic relations that elements of sentences structures have with each other in context, rather than with the system of contrasts and oppositions that differentiate constituents paradigmatically. The concern is with the inner structure of clauses rather than with the semantics of interclausal connections through the devices of coordination and subordination (Fillmore 1977: 60).

### 4.3 Sémantique des cadres

Après la grammaire des cas, en 1976, Fillmore présente une réflexion sur la linguistique dans un article intitulé « Frame semantics and the nature of language », dans laquelle il expose sa théorie de la Sémantique des cadres. La notion de « cadre » chez Fillmore est l'extension de ses idées dans la grammaire des cas. Sur la nature de la Sémantique des cadres, Fillmore précise que :

My effort is to look for what can be known about the workings of language through a consideration of the processes of communication. A proposal that I favour is that in characterizing a language system we must add to the description of grammar and lexicon a description of the cognitive and interactional frames in terms of which the language-user interprets his environment, formulates his own messages, understands the messages of others, and accumulates or creates an internal model of his world (Fillmore 1976 : 23).

Les propos de Fillmore reproduits dans cette citation peuvent être résumés dans son slogan : « Meanings are relativized to scenes » (1977 : 59). En d'autres mots, Fillmore estime que la langue fait partie d'un processus de communication où le système langagier peut être décrit en combinant la grammaire, le lexique et les processus cognitifs qui servent à interpréter l'environnement, formuler les messages et comprendre ceux des autres et finalement avoir une compréhension interne du monde. Cette combinaison construit une sorte de scène ou encore *cadre (frame)*.

Quant à l'idée de *cadre*, on peut retracer son origine dans les travaux de Minsky (1974). La théorie de Minsky concerne la représentation des connaissances basée sur des cadres, surtout dans le domaine de l'intelligence artificielle (IA). Selon Minsky, le cadre est une structure des données pour représenter une situation stéréotypée (Minsky 1974 : 1). De plus, à chaque cadre s'attachent un certain nombre d'informations comme la façon d'utiliser le cadre, les attentes qu'on peut avoir avec son utilisation et ce qu'on peut faire si les attentes espérées ne sont pas confirmées.

En outre, l'idée de cadre peut être retrouvée en psychologie cognitive (Fillmore 1976 : 25), en particulier dans la théorie de *schemata* de Bartlett (1937). En psychologie cognitive, la modélisation des connaissances se fait au moyen de cadres qui sont considérés comme étant un mode fondamental de représentation de connaissances. D'après cette théorie, les cadres sont en constante évolution parallèlement à l'expérience humaine : « frames are continually updated and modified due to ongoing human experience, and are used in reasoning in order to generate new inferences » (Evans et Green 2006 : 223). C'est pour cette raison que Fillmore (1982) inscrit sa théorie dans la tradition de la sémantique empirique d'autant plus que la Sémantique des cadres rend compte d'une sorte de « continuities, rather than the discontinuities » (ibid. : 111) entre la langue et l'expérience humaine, et c'est cette continuité qui marque la relation liant la langue au concept des cadres. Cette relation fait en sorte que certains mots, formules ou choix grammaticaux sont associés dans la mémoire humaine à des cadres (Fillmore 1976 : 25).

Fillmore résume l'idée centrale de la Sémantique des cadres comme suit :

Word meaning must be described in relation to semantic frames – schematic representation of the conceptual structures and patterns of beliefs, practices, institutions, images, etc. that provide a foundation of meaningful interaction in a given speech community (Fillmore et al. 2003a : 235).

La notion principale dans la Sémantique des cadres est celle de « cadre sémantique ». Nous signalons que les premiers travaux de Fillmore (1975) distinguaient « scène » de « cadre ». « Scène » relevait des aspects cognitifs et conceptuels, alors que la notion de « cadre » tenait compte des aspects linguistiques (Petrucci 1996). Ainsi, Fillmore (1975 : 124) fait observer ce qui suit :

I would like to say that people associate certain scenes with certain linguistic frames. I use the word scene in a maximally general sense, including not only visual scenes but also familiar kinds of interpersonal transactions, standard scenarios defined by the culture, institutional structures, inactive experiences, body image, and, in general, any kind of coherent segment human beliefs, actions, experiences or imagings. I use the word frame for any system of linguistic choices—the easiest cases being collections of words, but also including choices of grammatical rules or linguistic categories— that can get associated with prototypical instances of scenes.

Avec le temps, le choix s'est fixé sur *cadre* et *scène* a cessé d'être utilisé pour désigner la notion de « cadre sémantique ». D'après la citation reproduite ci-haut et la définition du cadre de Fillmore (2003a), nous croyons que Fillmore a opté pour une fusion entre les deux notions, puisque cette fusion reflète les deux aspects fondamentaux de sa théorie, à savoir la sémantique et la syntaxe.

Enfin, la notion de « cadre » constitue la base de ce que Fillmore appelle la sémantique de la compréhension (*U-semantics*) (Fillmore 1985 : 222, cité dans Petruck 1996). Le but de la sémantique de la compréhension est de fournir le lien entre les textes linguistiques et leurs contextes, d'une part, et le processus et le produit de leur interprétation, d'autre part. En d'autres mots, la sémantique de la compréhension détermine de quelle façon le locuteur interprète une phrase afin de comprendre la situation qui lui correspond. De plus, la sémantique de la compréhension est d'une nature compositionnelle, dans le sens où, afin que la communication s'établisse, son mode opératoire dépend des connaissances des mots, des phrases et des constructions grammaticales (Petruck 1996 : 3).

Pour résumer, un cadre sémantique est : « any system of concepts related in such a way that to understand any one concept it is necessary to understand the entire system; introducing any one concept results in all of them becoming available » (ibid. : 1). À titre d'exemple, les moyens de transport peuvent être décrits via la notion de « cadres sémantiques ». Cette description est basée sur le cadre ou encore sur un système conceptuel où les concepts sont reliés les uns aux autres. Les moyens de transport peuvent être compris comme faisant partie d'un système de termes de transport. Ce qui relie les moyens de transport dans ce système est une schématisation spécifique de l'expérience (Fillmore et Atkins 1992 : 77). En tant qu'unités faisant partie d'un système donné, les moyens de transport peuvent être mis dans un cadre constitué des connaissances suivantes :

1. les différents moyens de transports;
2. la fonction et l'utilité de chacun;
3. les composants essentiels de chaque moyen;
4. la pratique culturelle qui consiste à assigner le mode d'emploi et le système des moyens de transport.

D'après Fillmore et Atkins, cette organisation conceptuelle du système des moyens de transport constitue la base d'un arrière-plan d'un certain nombre d'informations lexicales telles que les bus, les taxis, les trains, les voitures de promenade, etc. Donc, une description d'une unité phraséologique comme « j'ai pris le bus pour aller au travail » doit faire appel à cet ensemble de notions interconnectées (Fillmore et Atkins 1992 : 78).

Dans la section suivante, nous décrivons les principes fondamentaux de la Sémantique des cadres. Pour ce faire, nous donnons l'exemple du cadre *Transfer\_scenario*.

#### **4.3.1 Transfer\_scenario frame**

Comme mentionné dans la section précédente, un cadre est un schéma qui comprend un certain nombre d'informations. Ces informations combinent une description à deux niveaux, sémantique et syntaxique. À chaque item lexical est associée une description de valence. La description de valence a pour rôle de spécifier, sémantiquement et syntaxiquement, les constituants et les contextes pour l'UL. De plus, cette description spécifie la contribution de l'unité à la structure dans laquelle elle se trouve. D'après Fillmore et Atkins (ibid. : 78) le système le mieux développé de la description de valence est celui qui concerne la grammaire et les sens des verbes.

Dans le cadre *Transfer\_scenario*, nous pouvons trouver un certain nombre de verbes, comme *change*, *exchange*, *transfer*, *receive*, *give* et *pass out*. Afin de comprendre leur sens, il faut avoir accès au cadre *Transfer\_scenario* qui fournit : « the background and motivation for the categories which these words represent » (Fillmore 1982 : 116-17). De plus, ces verbes indexent différents aspects du cadre même (Petrucci 1996). Par exemple, le verbe *transfer* exprime le fait qu'un Donneur (l'entité responsable du transfert) fait le transfert d'un Thème

(l'objet qui fait l'objet du transfert) vers un Destinataire (l'entité qui reçoit le Thème). Tandis que le verbe *exchange* exprime l'idée que deux Échangeurs, ou un Échangeur (l'entité qui effectue l'action), se donnent des Thèmes (l'objet échangé) l'un à l'autre.

Comme nous l'avons mentionné au début de cette section, les UL reçoivent une description de valence qui met en lumière leurs propriétés syntaxiques. Cette description comprend les modèles syntaxiques dans lesquels apparaissent les unités qui évoquent le cadre. Par exemple, en ce qui concerne les unités verbales évoquant le cadre, la description tient compte des aspects concrétisés au sein du cadre comme le sujet ou l'objet du verbe, leurs groupes syntaxiques et les éléments obligatoires et ceux qui sont optionnels. Ces éléments s'appellent les éléments de cadre (*frame elements* ou FE). Dans le tableau 1, nous donnons un exemple du cadre *Transfer\_scenario* en montrant les unités verbales qui l'évoquent ainsi que les éléments de cadres et leur réalisation.

	<b>Donor</b>	<b>Exchanger 1</b>	<b>Exchanger 2</b>	<b>Recipient</b>	<b>Theme</b>	<b>Themes</b>
<b>Transfer</b>	NP.Ext			PP[to]Dep.	NP.Obj	
<b>Give</b>	Subj			NP. Obj	NP.Obj	
<b>Swap</b>		NP.Ext	PP[with].Dep		NP.Obj	NP.Obj
<b>Receive</b>	PP[from].Dep			NP.Ext	NP.Obj	

Tableau 1 : Verbes appartenant au *Transfer\_scenario*

Dans le tableau, les verbes qui font partie du cadre sont traités en fonction de leur valence. Par exemple, le verbe *transfer*, en tant qu'unité verbale, admet Donor, qui est réalisé par un syntagme nominal (NP. Ext), Recipient, réalisé par un syntagme prépositionnel (PP[to] Dep.) et Theme, qui est réalisé par un syntagme nominal occupant la fonction d'objet. Nous remarquons que le verbe *swap* n'admet pas les mêmes FE, comme Recipient et Donor. Cependant, tous les verbes admettent le FE Theme.

Certains FE ne figurent pas parmi les FE dans le tableau. Ces éléments ne sont pas des FE obligatoires. Nous mentionnons à titre d'exemple, le FE Manner.

Étant donné que chacun des verbes décrits présente une perspective différente sur le cadre, le cadre *Transfer\_scenario* peut avoir des sous-cadres qui représentent chacun des

aspects dénotés par les différents verbes. Dans ce qui suit, nous donnons deux exemples de deux sous-cadres différents pour les verbes *give* et *receive*.

Le verbe *give* appartient au cadre Giving. Ce cadre admet les FE suivants : Donor, Recipient et Theme. Ce cadre se focalise sur l'entité qui effectue l'action de donner (Donor) et l'objet donné (Theme). Dans les exemples<sup>32</sup> suivants, nous donnons plus de détails.

- *I* [*Donor*] *GAVE* *him* [*Recipient*] *coffee* [*Theme*]
- *We* [*Donor*] *GAVE* *him* [*Recipient*] *the money* [*Theme*]
- *But he* [*Donor*] *would GIVE generously* [*Manner*] *of his time* [*Theme*] *to such organisations as the British Council* [*Recipient*].

Dans le troisième exemple, le FE Manner est optionnel et ne fait pas partie des FE obligatoires du cadre Giving.

En ce qui concerne le verbe *receive*, il appartient au cadre Receiving. Tout comme Giving, ce cadre comprend les FE Donor, Recipient et Theme. Dans ce cadre, les éléments mis en évidence sont le Recipient et le Theme. Considérons les exemples suivants :

- *Pat Thurston , of Kew Rise , Darlington,* [*Recipient*] *RECEIVED a copy of the document* [*Theme*] *through the post* [*Mode\_of\_transfer*]
- *I* [*Recipient*] *RECEIVED my mail* [*Theme*] *at the barracks* [*Place*]
- *I'm afraid a beneficiary* [*Recipient*] *is not entitled to see a will or RECEIVE a copy* [*Theme*]

Dans les exemples, le FE Donor est manquant, son identité est irrécupérable, mais son type est normalement connu (*indefinite null instantiation*, INI). Nous remarquons les différences entre le premier groupe d'exemples et le deuxième groupe, même si le nombre et le type des FE sont identiques. Le cadre Receiving met l'accent sur Recipient et Theme, contrairement au cadre Giving, qui met l'accent sur le Donor et le Theme.

Dans cette optique, afin de comprendre l'ensemble structuré des relations linguistiques régies par le cadre, il est nécessaire d'analyser ce dernier en tant que système de représentation

---

<sup>32</sup> Les exemples sont pris de FrameNet (2016).



de connaissances qui fournit une grande série de séquences d'événements. Sur ce point, Fillmore (1985 : 232, cité dans Petruck 1996) constate que le texte évoque un cadre lorsqu'une forme linguistique est conventionnellement associée à un cadre donné.

## **4.4 Applications de la Sémantique des cadres**

La Sémantique des cadres a fait l'objet de plusieurs applications. Il s'agit de projets accessibles en ligne qui varient selon leur portée et objectifs. Nous organisons cette section de la manière suivante. D'abord, nous parlons des projets réalisés pour la langue générale, notamment le projet de FrameNet. Ensuite, nous discutons des projets conçus pour l'ASM général.

Puis, nous décrivons des projets conçus pour la langue de spécialité. Comme il en existe un certain nombre, notre choix a porté sur ceux qui nous ont inspiré dans le présent travail. Nous signalons que la Sémantique des cadres est un terrain très fertile pour la langue de spécialité. L'aspect cognitif intéresse de plus en plus les spécialistes en terminologie dans la conception de ressources lexicales spécialisées.

### **4.4.1 La langue générale : FrameNet**

L'application de la théorie de la Sémantique des cadres se fait dans FrameNet, de l'Université de Berkeley en Californie (Ruppenhofer et al. 2010). FrameNet est un projet lexicographique computationnel qui analyse les informations sur les propriétés syntaxiques et sémantiques des mots anglais à partir d'un large corpus sous format électronique. Les informations extraites sont présentées dans des rapports publiés sur le web (Fillmore et al. 2003a : 235). Le nom *FrameNet* est inspiré de fameuse base de données WordNet (Fellbaum 1998) : FrameNet reflète le fait qu'il est basé sur la théorie de la Sémantique des cadres et il traite de réseaux de sens auxquels les unités lexicales participent.

Le corpus à partir duquel FrameNet extrait les informations lexicales est le *British National Corpus* (BNC). L'objectif du travail est de décrire les : « frames that underlie their meanings and study the syntactic properties of words by asking how their semantic properties are given syntactic form » (Fillmore et al. 2003a : 235-6). Les unités de l'analyse lexicale dans FrameNet sont :

1. le cadre;
2. l'unité lexicale (UL, ou LU d'après Cruse (1986));
3. les réalisations syntaxiques et la valence; et
4. l'annotation.

#### 4.4.1.1 Le cadre

Le travail lexicographique dans FrameNet consiste, d'abord, à définir un cadre sémantique. La définition d'un cadre se fait par une description informelle du type de situation que le cadre représente. Ensuite, le lexicographe dresse une première liste d'unités lexicales évoquant ce cadre. Le choix de ces unités se fait en faisant appel aux dictionnaires électroniques et papier, aux thésaurus et aux locuteurs natifs. Ici, il convient de faire la différence entre le cadre et le sens. Dans FrameNet, chaque sens lexicalisé correspond à l'emploi d'une UL dans un cadre sémantique donné. La question qui se pose dans ce cas est l'appartenance d'autres mots au même cadre.

Atkins et al. (2003 : 255) proposent quatre étapes qui permettent d'associer une UL à un cadre sémantique. Ces étapes sont :

- a) look at examples of one particular word,
- b) for each semantic type of complement (or frame element – see below) that occurs with that word, look for other words with similar meanings that also take that kind of complement,
- c) notice which complement types cluster together with groups of meaning-sharing words,
- d) given two types of complement that both occur with the target word, if one complement regularly occurs with one group of related words, and the other with a different group of related words, this is strong evidence for a sense distinction (based on a frame distinction).

L'exemple que nous donnons ici est le cadre Placing (figure 49).

## Placing

[Lexical Unit Index](#)

### Definition:

Generally without overall (translational) motion, an **Agent** places a **Theme** at a location, the **Goal**, which is profiled. In this frame, the **Theme** is under the control of the **Agent/Cause** at the time of its arrival at the **Goal**.

**David** **PLACED** **his briefcase** **on the floor**.

This frame differs from Filling in that it focuses on the **Theme** rather than the effect on the **Goal** entity. It differs from Removing in focusing on the **Goal** rather than the **Source** of motion for the **Theme**.

Figure 49 : Cadre Placing dans FrameNet (FrameNet 2015)

Le cadre Placing représente une schématisation d'un processus effectué par un Agent (ou une Cause) qui consiste à mettre une entité (Theme) dans un endroit (Goal). On remarque l'existence de ces trois éléments (les mots colorés dans la figure) qui sont les FE. Dans FrameNet, les FE d'un cadre se divisent en deux catégories : les éléments obligatoires, ou les core-FE, qui délimitent le sens d'une UL; et les éléments non obligatoires, ou encore les non-core FE. Dans le contexte du cadre Placing, les core-FE sont : Agent (ou Cause), Theme et Goal.

Pour identifier les core-FE, trois conditions s'appliquent (Atkins et al. 2003 : 267-268)

:

1. Il faut que l'élément soit réalisé dans l'une des fonctions grammaticales fondamentales;
2. Il faut que la réalisation syntaxique de l'élément soit obligatoire, c'est-à-dire qu'il apparaît dans tout type de valence;
3. Il faut que l'élément soit central au sens du verbe.

En ce qui concerne les non-core FE, FrameNet en considère deux types (ibid. : 268) :

1. les compléments, ce qui inclut les modificateurs adverbiaux qu'ils soient pour exprimer le temps, l'espace, la manière, la raison, etc.;

2. les éléments extrathématiques. Il s'agit ici des éléments sémantiques qui ne sont pas directement introduits par le prédicat, mais plutôt par une structure à l'extérieur du cadre, comme des syntagmes prépositionnels ou une restructuration de la valence syntaxique.

#### 4.4.1.2 L'unité lexicale

Après avoir défini le cadre et établi la liste des unités lexicales qui l'évoquent, le lexicographe procède à l'analyse de ces unités (figure 50).

**Lexical Units:**

*archive.v, arrange.v, bag.v, bestow.v, billet.v, bin.v, bottle.v, box.v, brush.v, cage.v, cram.v, crate.v, dab.v, daub.v, deposit.v, drape.v, drizzle.v, dust.v, embed.v, emplace.v, enclose.v, file.v, gag.v, garage.v, hang.v, heap.v, immerse.v, implant.v, implanted.a, inject.v, insert.v, insertion.n, jam.v, lay.v, lean.v, load.v, lodge.v, pack.v, package.v, park.v, perch.v, pile.v, place.v, placement.n, plant.v, plunge.v, pocket.v, position.v, pot.v, put.v, rest.v, rub.v, set.v, sheathe.v, shelve.v, shoulder.v, shower.v, sit.v, situate.v, smear.v, stable.v, stand.v, stash.v, stick.v, stow.v, stuff.v, tuck.v, warehouse.v, wrap.v*

Figure 50 : Liste des unités lexicales du cadre Placing (FrameNet 2015)

Chaque UL du cadre fait l'objet d'une entrée lexicale. L'entrée lexicale donne des renseignements sur l'unité en ce qui concerne sa partie du discours, sa définition et ses FE. Ce dernier prend la forme d'un tableau récapitulatif montrant les différents FE et leur réalisation syntaxique (figure 51).

Frame Element	Number Annotated	Realization(s)
Agent	(25)	CNI.-- (17) NP.Ext (8)
Duration	(1)	Sub.Dep (1)
Goal	(10)	AVP.Dep (1) PP[in].Dep (7) PP[under].Dep (1) PP[on].Dep (1)
Manner	(8)	AVP.Dep (4) PP[in].Dep (4)
Source	(7)	AVP.Dep (7)
Theme	(25)	NP.Obj (12) NP.Ext (12) CNI.-- (1)

Figure 51 : Tableau récapitulatif des FE et leur réalisation syntaxique pour le verbe *file* (FrameNet 2015)

Un verbe comme *file* en anglais fait partie du cadre Placing. Ce verbe fait l'objet d'une analyse de son sens relatif au cadre. À partir d'un corpus, on extrait tous les contextes où l'UL apparaît. Considérons la figure 52 pour une liste d'occurrences du verbe *file*.

These forms are [filed](#) alphabetically , the expiry date carefully noted and systemat  
Each lexical item is given its own card and is [filed](#) alphabetically . CNI  
rning the mycoplasma 's potential and its location , and [filed](#) away . CNI  
His report to Washington was [filed](#) away . CNI  
ersonnel records , legal contracts and key reports to be [filed](#) away . CNI  
The easiest way to rectify this is to carefully [file](#) away any excess metal until a closer fit is achieved . CNI  
ter claims that photographs of a badly-beaten child were [filed](#) away for at least six years without being passed to police by  
it was one of simply thousands of documents prepared and [filed](#) away in the labyrinths of official Bonn . CNI  
It is important to [file](#) away your classwork notes carefully so that you are able to fi  
by sitting on your rear end behind cosy wooden desks and [filing](#) damn paper all day . CNI  
These cards I [filed](#) in alphabetical order , and I noted on each one details of ne  
atic employees are given letters of counselling which are [filed](#) in an unfavourable information folder . CNI  
Documents or files are given numbers and are [filed](#) in number order . CNI  
ack-to-back with the room history card ( Fig. 3.33 ) and [filed](#) in room order number . CNI  
rly merit publication , and not a few that clearly merit [filing](#) in the waste paper basket . CNI  
Burton [filed](#) it away .  
As a unique event , Karen could [file](#) it away in her snapshot album as one of the interesting things  
I [filed](#) it away in my wallet .  
traced the wound using a clean , clear polythene bag and [filed](#) it in Mrs Allen 's notes to allow regular reassessment of hee  
This evidence should then be [filed](#) so that if it is challenged or its validity is questioned , t  
ords it as an introduction on the INTRODUCE database and [files](#) the papers .  
After lunch I reread and [filed](#) the summary .  
s worth photocopying the more useful reading lists , and [filing](#) them in classified order in a pamphlet bibliography collecti  
Oh , I keep everything at home -- I [file](#) them there , on my own system , " said Mr Kronweiser eagerly .

Figure 52 : Exemple d'une liste d'occurrences d'une UL (FrameNet 2015)

Ces contextes (puisque'ils sont ceux d'une seule forme) servent par la suite à analyser les propriétés sémantiques et syntaxiques de l'UL. Par exemple, l'analyse des contextes permet de voir : 1) les différents sens que le verbe peut avoir; 2) les réalisations linguistiques

des FE; et 3) les propriétés syntaxiques de ces réalisations. Dans la figure 52, la phrase : *The easiest way to rectify this is to carefully file away any excess metal until a closer fit is achieved* montre un sens différent du verbe (le fait de limer le métal). Dans ce cas, le verbe est placé dans un cadre différent. Le contexte qui montre le sens du verbe qui concerne le cadre Placing est celui dénotant le fait de placer une entité dans un endroit, comme le montre la phrase : *As a unique event, Karen could file it away in her snapshot album as one of the interesting things that had happened during her holiday in France*, où Karen est l'Agent, it (event) est le Thème et *in her snapshot album* est le Goal.

Dans cette perspective, il convient de souligner que le choix des unités lexicales appartenant au cadre se fait en fonction d'une approche standardisée tel que décrit dans Atkins et al. (2003 : 254-255). Cette approche comprend les cinq étapes suivantes :

1. look at examples of a word,
2. use semantic intuition to recognize which examples exhibit different meanings,
3. notice syntactic correlates of the meaning differences,
4. record those syntactic correlates when noticed,
5. and perhaps use them in providing a guide to the user by way of an example.

#### **4.4.1.3 La valence**

Par la suite, le lexicographe identifie les participants principaux (les éléments de cadres, ou encore les FE) au cadre évoqué par l'UL qui s'appelle désormais la *cible* (*Target word*), d'après la terminologie de FrameNet. Cette étape du travail consiste à donner une description de valence sur le plan sémantique et syntaxique, à savoir la configuration des FE et le groupe, le modèle et la description de valence (Atkins et al. 2003 : 255). Dans FrameNet, un tableau récapitulatif de l'UL est produit montrant les modèles de valence pour l'ensemble des contextes analysés (figure 53).

Number Annotated	Patterns				
<u>1</u> TOTAL	Agent	Agent	Theme		
(1)	NP Ext	NP Ext	NP Obj		
<u>1</u> TOTAL	Agent	Duration	Goal	Manner	Theme
(1)	CNI --	Sub Dep	AVP Dep	AVP Dep	NP Obj
<u>1</u> TOTAL	Agent	Goal	Manner	Theme	
(1)	CNI --	PP[in] Dep	PP[in] Dep	NP Obj	
<u>2</u> TOTAL	Agent	Goal	Source	Theme	
(1)	CNI --	PP[in] Dep	AVP Dep	NP Ext	
(1)	NP Ext	PP[in] Dep	AVP Dep	NP Obj	
<u>5</u> TOTAL	Agent	Goal	Theme		
(1)	CNI --	PP[in] Dep	NP Ext		
(1)	CNI --	PP[under] Dep	CNI --		
(2)	NP Ext	PP[in] Dep	NP Obj		
(1)	NP Ext	PP[on] Dep	NP Obj		
<u>1</u> TOTAL	Agent	Goal	Theme	Theme	
(1)	CNI --	PP[in] Dep	NP Ext	NP Ext	

Figure 53 : Modèles de valence pour l'UL *file* (FrameNet 2015)

#### 4.4.1.4 L'annotation

Dans FrameNet, ce qui permet de repérer les différents FE et voir leur réalisation dans des phrases est l'annotation. Ainsi, l'annotation aide à (Atkins et al. 2003 : 256-257) :

1. repérer le groupe de valence, qui est l'élément du cadre avec sa réalisation grammaticale (type de phrase et sa fonction);
2. le modèle de valence (*valence pattern*), qui comprend le type de phrase;
3. la description de valence, qui est un ensemble de tous les modèles de valence réalisé par une seule UL.

L'annotation compte parmi les étapes les plus importantes du travail lexicographique dans FrameNet. L'annotation est le fait d'étiqueter les composants syntaxiques importants dans les phrases avec les noms des FE qu'ils expriment. En effet, l'annotation comprend un ensemble de couches (*layers*). Ces couches désignent la cible, les FE, leur fonction

grammaticale et leur groupe syntaxique. L'annotation met également en évidence les verbes support et les verbes copules (Johansson et Nugues 2006 : 136).

Une distinction est faite entre les core-FE et les non-core FE. Les core-FE sont marqués par « core », tandis que les non-core FE sont divisés en deux parties : « Peripheral » et « Extra-Thematic ». Les peripheral non-core FE expriment des notions, comme le temps, le lieu, la méthode, la manière et le degré. Les Extra-thematic non-core FE « situate an event against a backdrop of another state of affairs, either of an actual event or state of the same type » (Ruppenhofer et al. 2010 : 20). À titre d'exemple, dans les phrases suivantes, nous pouvons identifier un non-core FE Peripheral (Manner) et un non-core FE Extra-Thematic (Duration) :

- [1] *These forms are FILED alphabetically [Manner] , the expiry date carefully noted and systematically cleared out after the expiry date has elapsed.*
- [2] *The easiest way to rectify this is to carefully FILE away any excess metal until a closer fit is achieved [Duration].*

Finalement, dans les phrases annotées, les core FE et les non-core FE sont soulignés au moyen des couleurs comme dans la figure 54.

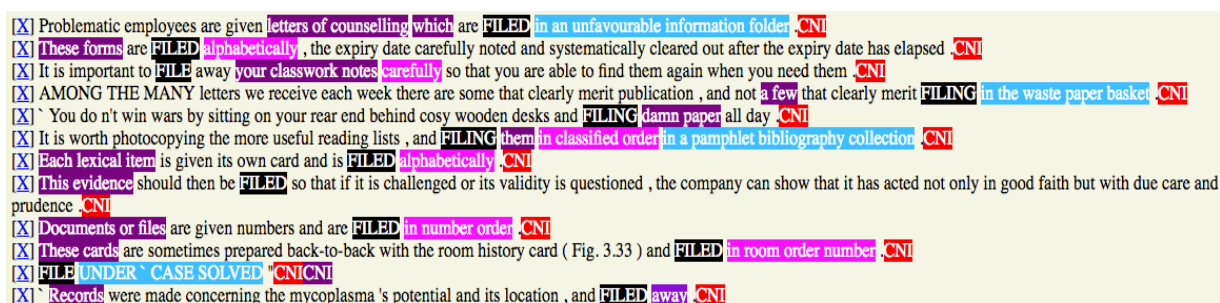


Figure 54 : Phrases annotées pour l'UL file (FrameNet 2015)

#### 4.4.1.5 Relations entre les cadres

Dans FrameNet, les cadres entretiennent des relations entre eux et forment un réseau. Les relations qu'on trouve dans FrameNet comprennent : Inherits from, Is Inherited by, Perspective on, Is Perspectivized in, Uses, Is Used by, Subframe of, Has Subframe(s), Precedes, Is Preceded by, Is Inchoative of, Is Causative of et See also.



Par exemple, la relation « Inherits from » désigne le lien entre deux cadres lorsqu'un second cadre possède toutes les propriétés du premier plus une caractéristique supplémentaire. Les relations dans FrameNet sont représentées au moyen de graphes (figure 55).

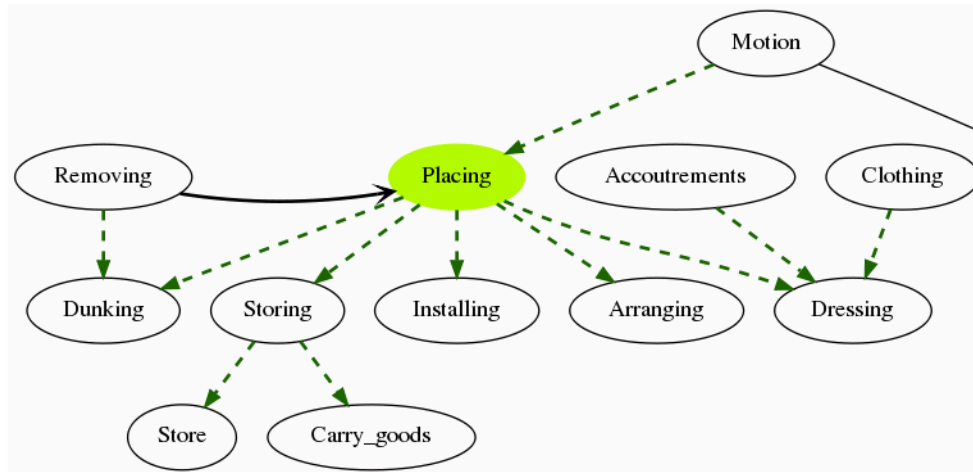


Figure 55 : Relations entre cadres dans FrameNet (FrameNet 2015)

Dans la figure, le cadre Placing se trouve dans un réseau comprenant plusieurs cadres. Les deux relations qu'il partage avec d'autres cadres sont Using of et Precedes. D'après Ruppenhofer et al. (2010 : 78) la relation Using est utilisée lorsqu'une partie de la scène évoquée par un cadre Enfant se rapporte au cadre Parent. Pour la relation Precedes, elle relie deux cadres en fonction d'une séquence d'évènements. Normalement, cette relation réunit deux cadres qui constituent des composants d'un cadre plus complexe (Ruppenhofer et al. 2010 : 77). En d'autres mots, le cadre Placing est précédé par le cadre Removing et comprend une partie de la scène exprimée par Installing et Storing.

#### 4.4.2 FrameNet dans d'autres langues

Dans cette section, nous abordons le projet FrameNet dans d'autres langues, notamment en ASM. Notre objectif est de décrire des initiatives tenant compte de l'ASM en ce qui concerne son traitement de point de vue de la Sémantique des cadres. De plus, nous présentons des projets qui appliquent le modèle de FrameNet à cette langue.

#### 4.4.2.1 Verbes de mouvement du point de vue de la Sémantique des cadres

La seule étude que nous avons trouvée qui présente une analyse de l'ASM de point de vue de la Sémantique des cadres est la thèse de doctorat de Alshehri (2014). Dans son travail, l'auteur explore des cadres sémantiques définis pour l'anglais et la possibilité que ces cadres, initialement conçus pour l'anglais, existent dans d'autres langues, notamment en ASM.

La recherche est basée sur un corpus bilingue anglais-ASM et axée sur l'analyse de cinq unités verbales évoquant le cadre *Self\_motion*. La recherche a donné trois types de résultats, que nous résumons dans ce qui suit. Premièrement, le cadre *Self\_motion* comporte les rôles sémantiques qu'on peut observer dans les verbes ASM de mouvement autopropulsé (self-propelled motion). Deuxièmement, il existe certaines restrictions en ce qui concerne les traits sémantiques (comme Humain, Animal, etc.) des rôles sémantiques en ASM et en anglais dépendamment des verbes. Troisièmement, l'ASM présente certaines particularités concernant les rôles sémantiques, puisque quelques-uns ne sont pas présents.

Dans sa méthodologie, l'auteur a choisi de travailler sur cinq verbes dans chaque langue. Pour l'anglais, les verbes candidats sont *walk*, *run*, *fly*, *climb* et *crawl*. Nous rappelons que ces verbes font partie des unités lexicales qui évoquent le cadre *Self\_motion* dans FrameNet. Selon l'auteur, ces verbes ont la particularité d'avoir un usage fréquent et leur équivalent en ASM sont composés d'un seul morphème. Les verbes ASM sont *ma\$Y* (مشى) (marcher), *jarY* (جری) (courir), *TArA* (طار) (voler), *tasal~aqa* (تسلق) (grimper) et *zaHafa* (زحف) (ramper).

Le corpus utilisé dans cette recherche est bilingue. Le corpus anglais provient de *Corpus of Contemporary American English* (COCA). Le repérage des cinq verbes dans le corpus est fait de façon manuelle. Toutes les formes sont récupérées, à savoir verbe+ $\emptyset$ , verbe+s, verbe+ed, verbe+ing. En ce qui concerne le corpus ASM, les données sont recueillies du Online Arabic Corpus. La même méthode est utilisée pour le repérage des cinq verbes ASM. La seule différence constatée est que la recherche des formes verbales est faite à partir de la forme de la troisième personne au singulier. Les occurrences des verbes obtenues ont subi un filtrage pour délimiter les données en fonction des besoins de la recherche. Par exemple, l'emploi figuré est éliminé comme dans les phrases suivantes :

- [1] *zaHafa aljay\$ tijAha alHiSn*  
 زحف الجيش تجاه الحصن  
*L'armée a rampé vers la forteresse*

- [2] *He will run for president, and even win.*

Finalement, pour chaque verbe dans les deux langues, entre 50 et 62 contextes ont été prélevés.

L'analyse du cadre se limite aux core-FE, à savoir Self\_mover, Direction, Source, Path, Goal et Area. Pour ce faire, l'analyse est effectuée par paire de verbes : deux verbes anglais et leurs équivalents ASM. L'accent est mis sur le type de traits sémantiques associés au FE Self\_mover et la fréquence de la spécification des autres FE. Par exemple, les verbes *ma\$Y* (مشى) (marcher) et *jarY* (جرى) (courir) et *walk* et *run* comportent les composantes « motion » et « manière ». Par conséquent, après avoir analysé les occurrences de ces quatre verbes dans le corpus, l'auteur (Alshehri 2014 : 10) constate ce qui suit. Il existe certaines similitudes entre les verbes analysés en ce qui concerne les notions qu'ils dénotent. Les quatre verbes reçoivent les traits Animé (humain ou animal) et Inanimé. La figure 56 résume les résultats de cette analyse.

Self-Mover	Arabic (ma\$fa)	English (walk)	Arabic (Jarfa)	English (Run)
1. ani. human	53 (1 ghost) (85%)	56 (97%)	47 (84%)	50 (93%)
2. ani. animal	3 (5%)	2 (3%)	5 (9%)	4 (7%)
3. vehicle	6 (10%)	0	3 (7%)	0

Figure 56 : Traits sémantiques et fréquence du FE Self-mover pour les verbes anglais et arabes  
 (Alshehri 2014 : 10)

La figure montre que la fréquence des traits Animé Humain et Animé Animal est presque la même pour les quatre verbes en question. Par contre, le trait Inanimé (*vehicle*) est spécifique aux verbes ASM. D'après FrameNet, le FE Self-Mover ne couvre pas les objets : « Self\_mover is the living being which moves under its own power. Normally it is expressed as an external argument » (FrameNet 2015). Cependant, l'auteur ne justifie pas la raison pour

laquelle il a inclus cet usage des verbes ASM. Considérons les exemples suivants de certains de ces emplois :

- [1] *Ha\*a AltrAm yam\$y fy xaT da}ry*  
 هذا الترام يمشي في خط دائري  
*Ce tram circule dans un mouvement circulaire*
- [2] *EarabAt AlHanTwr tajry bijAnibihi taHt ALA\$jAr*  
 عربات الحنطور تجري بجانبه تحت الأشجار  
*Les diligences passent devant lui sous les arbres*

D'après l'auteur, les véhicules : « are understood to be able to walk and run by itself » (Alshehri 2014 : 11). Il nous semble que, pour cette raison, l'auteur a fait le choix d'intégrer ce genre d'emploi dans son analyse des verbes.

Pour le reste des FE, l'auteur présente les résultats suivants (figure 57).

FEs	Arabic (məʃā)	English (walk)	Arabic (ʒara)	English (Run)
<b>Direction</b>	0	0	1 (1%)	0
<b>source</b>	3 (5%)	14 (24%)	4 (7%)	25 (44%)
<b>Path</b>	13 (26%)	5 (9%)	10 (16%)	4 (7%)
<b>Goal</b>	24 (38%)	32 (55%)	25 (41%)	15 (26%)
<b>Area</b>	11 (18%)	8 (14%)	11 (18%)	5 (9%)

Figure 57 : Fréquence des autres FE du cadre Self\_motion (Alshehri 2014 : 10)

D'après les données présentées dans la figure, les corpus n'attestent pas d'une fréquence pour le FE Direction (à l'exception du verbe *jarY* (جرى) (courir) avec une fréquence très faible de 1 %). Quant au FE Source, les unités verbales anglaises *walk* et *run* évoquent plus fréquemment ce FE que les unités verbales ASM. Par contre, pour le FE Path et Area, les verbes ASM en font un plus grand usage (même si avec une faible fréquence). Le FE Goal a une bonne fréquence pour tous les verbes, sauf pour le verbe anglais *run*. À partir de ces résultats, l'auteur (ibid. : 11-12) tire la conclusion suivante :

The implication of differences between Arabic and English of the frequency of the specification of the spatial role is that generally in the self\_motion events of walk and run, English speakers pay more attention to the beginning and end of the motion more than Arabic speaker. However, Arabic speakers seem to pay slightly more attention to the trajectory and area of the motion events. This difference highlights that walk and run and their Arabic counterparts have some semantic differences.

Pour les quatre verbes combinés avec tous les FE analysés, les résultats suivants sont obtenus (figure 58 et 59).

Semantic role	Fillers	Arabic (total: 280)	English (total: 280)
<b>Self-Mover</b>	1. Human	180 (64%)	201 (72%)
	2. Animal	68 (24%)	58 (21%)
	3. vehicle	15 (5%)	10 (4%)
	4. Inanimate	5 (2%)	5 (2%)

Figure 58 : Traits sémantiques du FE Self\_mover pour tous les verbes analysés (Alshehri 2014 : 12)

Semantic Roles	Arabic (total of 280)	English (total of 280)
<b>Direction</b>	5 (2%)	6 (2%)
<b>source</b>	16 (6%)	70 (25%)
<b>Path</b>	29 (10%)	41 (15%)
<b>Goal</b>	79 (28%)	107 (38%)
<b>Area</b>	92 (33%)	54 (19%)

Figure 59 : Fréquence des non-core FE pour tous les verbes analysés (Alshehri 2014 : 13)

D'après les figures, les traits sémantiques du FE Self\_mover sont presque identiques pour tous les verbes, les plus fréquents étant les traits Humain et Animal. Les traits Vehicle et Inanimate relèvent d'une classification de Talmy (1985) selon laquelle deux catégories existent pour la notion de *motion* : *self-agentive*, qui désigne les entités comme Humain et Animal, et *non-agentive*, qui désigne les entités comme les objets concrets. De ce fait, les entités de type *self-agentive* (Humain et Animal selon l'analyse) sont considérés comme étant prototypiques du cadre Self-motion. Cela explique la raison de la haute fréquence des traits Humain et Animal par rapport à Vehicle et Inanimate. Pour le reste des FE, nous remarquons les différences entre les deux langues, à l'exception du FE Direction, où les deux langues expriment la même réalité.

Enfin, nous signalons que les analyses faites sur le cadre Self\_motion n'ont pas fait l'objet d'une application pratique. De plus, l'auteur a mené une recherche sur un nombre limité de verbes dans les deux langues, ce qui ne permet pas, à notre avis, de faire des généralisations sur les verbes en ASM de point de vue de leur traitement sous l'angle de la théorie de la Sémantique des cadres. Cependant, dans la section 4.4.2.2, nous présentons une tentative de construction d'un FrameNet pour l'ASM.

#### 4.4.2.2 Arabic FrameNet (AFN)

À notre connaissance, il n'existe pas encore de ressource disponible en ligne pour un FrameNet de la langue arabe. Le projet que nous présentons dans cette section est une initiative visant la création d'une telle ressource. Arabic FrameNet (AFN) (Ghneim et al. 2009) est une proposition pour construire une ressource lexicale qui comprend des informations sémantiques et syntaxiques sur des concepts et des mots. Cette ressource constitue la base (ou encore un noyau) pour un futur FrameNet de l'ASM. Le projet est basé sur une ressource lexicale existante appelée *Almukniz Alkabyr* (Omar et al. 2000)<sup>33</sup>, qui sert à extraire les concepts sémantiques, les UL et les FE.

Les auteurs constatent que l'étape la plus importante dans la construction d'AFN est la définition de la structure des cadres et de leurs éléments ainsi que les relations entre eux (cadres). Par exemple, le cadre AntiqAm (انتقام) (Vengeance) peut avoir les FE suivants : *muEtady* (معتدي) (Agresseur), *muntaqim* (منتقم) (la personne responsable de l'acte de vengeance) et |\*Y (أذى) (dégât). Ce cadre comporte, également, les UL suivantes qui l'évoquent : *Antaqama* (انتقم) (se venger), *va>ra* (ثار) (l'acte de vengeance) et *AqtaS~a* (اقتص) (punir). Les relations que les auteurs proposent pour ce cadre est Inheritance et Sub-frame. Le cadre AntiqAm (Vengeance) hérite du cadre EuqwbAt (Châtiments), alors que les cadres AEtiqAl (اعتقال) (Détenion), Ait~ihAm (إتهام) (Accusation), muHAKamap (محاكمة) (Jugement) et AiSdAr Hukm (إصدار حكم) (Verdict) sont des sous-cadres du cadre complexe rafE daEwap tajrym (رفع دعوة تجريم) (Poursuivre\_en\_justice).

---

<sup>33</sup> Almukniz Alkabyr est un dictionnaire moderne de la langue arabe. Ce dictionnaire comprend des concepts sémantiques ainsi que des antonymes et des synonymes.

L'AFN dispose d'une structure de base composée de cadres, de FE, de verbes et de relations (figure 60).

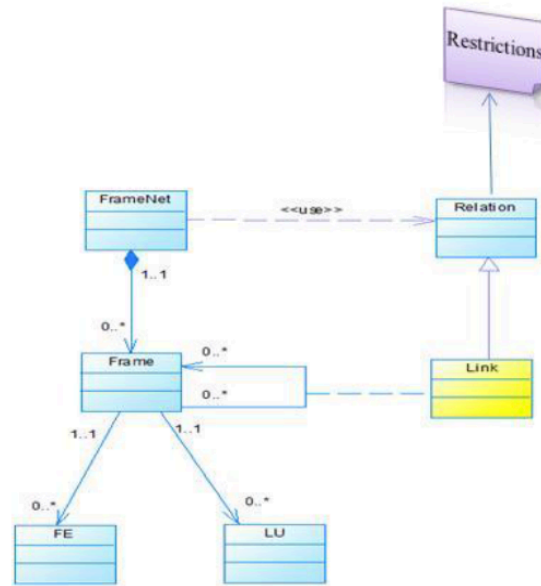


Figure 60 : Structure de base de l'AFN (Ghneim et al. 2009)

Dans cette structure, les relations peuvent être limitées au moyen de restrictions sélectionnelles qui décrivent la nature des cadres qui sont liés au moyen de cette relation (Ghneim et al. 2009). Par exemple, la restriction sélectionnelle d'une relation dite *AftirAs is* (« إفتراس is ») (« prédation » est) est « kA}n Hay » yaftaris « kA}n Hay » (« كائن يفترس » « كائن حي ») (« un être vivant » dévore « un être vivant »).

L'AFN a fait l'objet d'une application avec une interface utilisateur (figure 61).

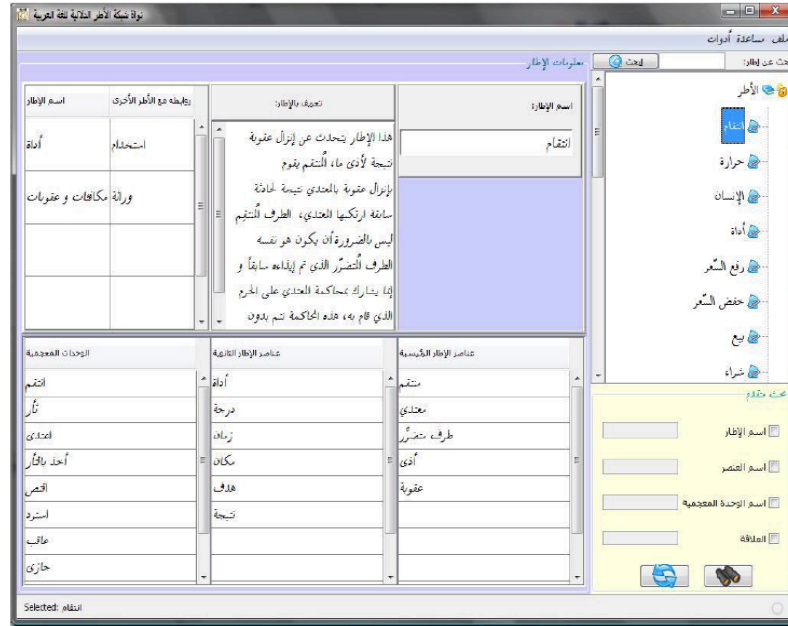


Figure 61 : Interface utilisateur de l'application d'AFN (Ghneim et al. 2009)

Cette application permet à l'utilisateur d'effectuer une recherche dans AFN par cadre, par UL, par FE ou par relation. De plus, à la suite d'une requête, elle affiche tous les renseignements possibles liés à un cadre donné. Nous signalons que cette application n'est pas encore disponible en ligne. Donc, nous n'avons pas eu la possibilité de la télécharger et de la tester.

Étant donné qu'AFN est basé sur une ressource lexicale existante, l'extraction des unités lexicales et des concepts se fait manuellement. Les auteurs ne font pas état d'une procédure automatique ou informatisée<sup>34</sup> quant à la collecte des unités lexicales ou des concepts. De plus, la méthodologie suivie lors du choix de cadres ne nous semble pas très explicite. Nous n'avons pas trouvé d'autres informations plus détaillées sur ce projet, que les auteurs décrivent comme « an initial version of Arabic FrameNet » (Ghneim et al. 2009).

<sup>34</sup> Nous signalons que le corpus n'existe pas sous format électronique.



### 4.4.3 Application de la Sémantique des cadres et le FN en langues de spécialité

Nous avons discuté dans la section 4.4.2 des projets de recherche s’inspirant de la Sémantique des cadres ainsi que FrameNet pour l’ASM. Dans cette section, nous décrivons une application de FrameNet en ASM de spécialité. Nous présentons aussi ce qui a été fait pour d’autres langues, notamment l’anglais et le français.

La Sémantique des cadres et le projet de FrameNet ont fait l’objet d’applications dans certains domaines de spécialité, à savoir Kicktionary (Schmidt 2009) dans le domaine du soccer, BioFrameNet (Dolbey et al. 2006) dans le domaine de la biologie, dans le domaine de l’ingénierie côtière (Faber et al. 2005), dans le domaine du texte coranique (Atwel et Sharaf 2009), JuriDico dans le domaine de droit (Pimentel 2012), la terminologie des cadres, *Frame Based Terminology* (FBT) (Faber 2012) et le Framed DiCoEnviro (L’Homme et Robichaud 2014 et L’Homme et al. 2014) dans le domaine de l’environnement.

Parmi les travaux présentés, nous retenons les projets suivants : Quranic FrameNet Project, pour l’ASM, BioFrameNet et Framed DiCoEnviro. Nous avons décidé de traiter ces trois ressources plutôt que les autres, car nous avons trouvé que BioFrameNet est lié au FrameNet dans la mesure où le premier est une sorte d’intégration d’une ressource spécialisée à la ressource générale. En ce qui a trait au Quranic FrameNet, il s’agit du premier projet inspiré de FrameNet pour l’ASM de spécialité<sup>35</sup>. En ce qui concerne le Framed DiCoEnviro, cette ressource s’inspire, en partie, de FrameNet, surtout en ce qui concerne la méthodologie de création de cadres et d’annotation. De plus, notre méthodologie s’inspire de celle mise au point pour élaborer cette dernière ressource.

#### 4.4.3.1 The Quranic FrameNet Project

The Quranic FrameNet (Atwel et Sharaf 2009) est un projet visant à concevoir un modèle de représentation de connaissances (*Knowledge Representation*, KR) pour le texte

---

<sup>35</sup> Nous signalons que le Quranic FrameNet s’appuie sur la CA (classical arabic), qui est une variante de la langue arabe (section 2.2). Cependant, pour cette section, nous considérons que le texte coranique relève d’une langue de spécialité.

coranique selon la théorie de la Sémantique des cadres. L'objectif de ce projet est de construire une ressource lexicale décrivant les verbes dans la version arabe du Coran. Ce projet de recherche a été initié par l'Université de Leeds en Grande-Bretagne. Le choix de ne traiter que les verbes provient du fait que le verbe joue le rôle prédicatif le plus important dans la phrase (Atwel et Sharaf 2009 : 8). À chaque verbe choisi pour l'étude correspond un cadre dans FrameNet. Afin de vérifier l'exactitude du choix, la valence du verbe arabe doit être étudiée pour voir s'il fait appel aux mêmes core-FE que ceux apparaissant dans le cadre correspondant.

Le texte coranique a ses propres particularités linguistiques, surtout en ce qui concerne les emplois figurés. Par exemple, dans le verset suivant (ibid. : 4), le texte décrit les sentiments de peur des soldats :

*When they came upon you from above you and from below you, and when eyes grew wild and hearts reached to the throats [33:10]*

De plus la métonymie occupe une place importante dans le texte, comme le montrent les exemples suivants :

*Ask the township where we were, and the caravan with which we travelled hither. [12:82]  
And We carried him upon a thing of planks and nails [54:13]*

Dans le premier exemple, la phrase *ask the towns* signifie « demander aux habitants de ces villes », alors que dans le deuxième exemple, *a thing of planks and nails* signifie l'arche de Noé.

La méthodologie suivie dans ce projet consiste à extraire tous les verbes et leurs contextes du texte coranique. Ensuite, l'analyse est effectuée au moyen d'un index informatisé. Chaque forme verbale est classifiée en fonction de son schème (figure 62).

NO	pattern	Semantic significance	Examples
I	فَعَلَ Fa3aLa	<ul style="list-style-type: none"> <li>When the 2<sup>nd</sup> radical is vowelized with (a) it mostly indicates transitive.</li> <li>When the 2<sup>nd</sup> radical is vowelized with (i) it mostly indicates intransitive.</li> </ul>	كَتَبَ (to write) فَرِحَ (to be glad)
II	فَعَّلَ Fa33aLa	<ul style="list-style-type: none"> <li>Intensive or extensive meaning of the first form</li> <li>Convert the intr. In 1<sup>st</sup> form to transitive</li> <li>Estimative or declarative</li> </ul>	كَسَّرَ (to break) and كَسَّرَ (break into pieces) فَرَّحَ (to be glad) فَرَّحَ (to gladden) كَذَّبَ (to lie), كَذَّبَ (to call one a liar)
III	فَاعَلَ Faa3aLa	<ul style="list-style-type: none"> <li>Place effort to perform act upon the obj.</li> <li>Convert prepositional object to direct obj.</li> <li>Use Quality or state to affect another person</li> </ul>	قَاتَلَهُ (he tried to kill him) كَتَبَ إِلَى (write to) = كَتَبَ (write to) حَاشَنَهُ (he treated him harshly)

Figure 62 : Classification des verbes en fonction de leurs formes (Atwel et Sharaf 2009)

Dans la figure, à chaque schème, on a associé les verbes correspondants. Par exemple, les verbes *katab* (كتب) (écrire) et *faraHa* (فرح) (réjouir) correspondent au schème *faE~ala* (فعل). De plus, une explication est fournie pour la signification des diacritiques associés aux formes verbales. Donc, dans le cas du verbe *kataba*, où le deuxième radical est vocalisé avec un *a*, le verbe devient transitif. Nous signalons que cette méthode de traitement des verbes est appliquée dans un certain nombre de ressources lexicales de l'ASM, comme AWN et AVN (voir section 3.2.1.2).

Ensuite, chaque verbe est analysé pour trouver l'UL correspondante dans FrameNet. Après cette étape, le cadre qui comprend cette UL est analysé afin d'en vérifier la bonne correspondance avec le verbe arabe. La condition optimale pour que le verbe *cible* arabe soit retenu pour un cadre donné est lorsqu'il comprend les mêmes core-FE que le cadre choisi.

À titre d'exemple, le cadre Ingestion dans FrameNet dispose de deux core-FE : Ingestibles et Ingestor. Le verbe arabe correspondant dans le texte coranique est *eat*<sup>36</sup>. Pour étudier les propriétés du verbe, une liste KWIC est établie et le lexicographe procède à l'analyse des occurrences (figure 63).

<sup>36</sup> Les auteurs donnent les exemples en anglais, et ils ne font état d'un corpus coranique en langue arabe.

A	the sea to be of service that ye	eat	fresh meat from thence	16:14
B	And	eat	not up your property among	2:188
C	Would one of you love to	eat	the flesh of his dead brother?	49:12
D	seven fat kine which seven lean were	eating		12:43
E	seven hard years which will	devour	all that ye have prepared for them	12:48
F	they	eat	into their bellies nothing else than fire	2:177
G		Devourer	of unlawful	5:42

Figure 63 : Concordances du verbe *eat* dans le Coran (Atwel et Sharaf 2009)

D’après les concordances du verbe, on remarque que certains usages ne correspondent pas au sens décrit dans le cadre *Ingestion*. *Eat* dans [B] n’a pas le même sens que dans les autres exemples. Pour cette raison, les auteurs proposent d’élargir leur projet pour comprendre ces usages particuliers en suivant la stratégie du FrameNet allemand, SALSA (Burchardt et al. 2009), qui propose la création d’un proto cadre (*proto-frame*). L’intérêt du projet SALSA pour les lexicographes du Quranic FrameNet est qu’il permet de traiter des verbes idiomatiques qui sont nombreux dans le texte coranique. C’est pour cette raison que nous trouvons dans les concordances le verbe *devour*. Sa présence, pourtant non justifiée par les auteurs, a pour but la création d’un cadre prototype qui comporte l’usage métaphorique du verbe *eat*.

En ce qui concerne l’annotation, les auteurs entendent annoter le Coran en fonction de la méthode d’annotation suivie dans FrameNet. Donc, elle comprendra trois couches : les FE, les fonctions grammaticales et les groupes syntaxiques. Les auteurs proposent dans les exemples suivants des versets annotés selon le modèle de FrameNet. On y remarque leur utilisation du cadre *Giving*, mais avec une extension pour rendre compte du sens exact du verbe *spend* dans le texte coranique (Atwel et Sharaf 2009).

*and [they DONOR] spend out of [what We have provided for them THEME] [2:3]  
and they ask you what [they DONOR] should spend. Say, “[the excess DONATED\_AMOUNT]”. [2:219]  
and they ask you what they should spend. Say, “Whatever [you DONOR] spend of [good  
DONATED\_AMOUNT] is [for parents and relatives and orphans and the needy and the traveler RECIPIENT].  
[2:215]*

Toutefois, les auteurs ne proposent pas une stratégie pour faire face à la vocalisation dans le texte coranique puisqu’elle constitue un problème important quant au sens des unités lexicales. À titre d’exemple, pour un verbe comme *eat*, en arabe la forme du verbe ne change

pas à la forme nominale. Ce qui change est plutôt la distribution des voyelles courtes : أكل / أكل.  
La translittération du verbe en lettres latines nous montre exactement la vocalisation correcte : Akala/Akl.

Nous signalons que le projet est en cours et il n'existe pas encore une application sur le web. De plus, dans leur article, les auteurs ne donnent pas d'exemples tirés de la version en langue arabe du Coran.

#### 4.4.3.2 BioFrameNet

Inspiré du projet de FrameNet, le projet de BioFrameNet vise la construction d'une ressource lexicale dans le domaine de la biologie. Il est considéré comme une extension de FrameNet puisqu'il comporte des liens qui renvoient aux cadres de FrameNet. Comme dans FrameNet, BioFrameNet comprend des cadres avec des définitions. Ces cadres sont liés les uns aux autres par des liens bien définis, à l'instar des liens entre les cadres dans FrameNet. En outre, pour chaque cadre il existe une définition pour les FE ainsi qu'une liste d'unités lexicales. Les sources primaires pour le BioFrameNet sont une collection de textes annotés par des biologistes de Hunetlab de l'Université du *Colorado Health Sciences Center*.

À titre d'exemple, dans BioFrameNet, nous trouvons le cadre de Transport\_intracellular (Dolbey et al. 2006 : 88) qui décrit le processus biologique du transport intracellulaire des entités moléculaires. Pour ce cadre les FE sont :

- Cargo : l'entité transportée;
- Carrier : l'entité transporteuse;
- Origin : le point de départ du transport;
- Destination : le point d'arrivée du transport.

Quant aux unités lexicales qui évoquent le cadre, nous trouvons :

*Translocate.v, Translocation.n, Transport.v, Transport.n, Shift.v, Shuttle.v, Export.v*

Il est à noter que, souvent, les nouveaux cadres ajoutés dans BioFrameNet trouvent leur origine dans FrameNet. Par exemple, le cadre de Transport\_intracellular fait partie du cadre Bringing dans FrameNet. Dans la figure 64, nous donnons un exemple de six cadres (Protein Activation, Necessity, Molecular Assembly, Molecular Complex, Part inner outer et Activation) avec des phrases annotées comportant des unités lexicales. Nous signalons que les cadres avec astérisques sont des cadres nouveaux qui n'ont pas d'équivalent dans FrameNet.

* Protein Activation Frame: The ACTIVATION [Activated Entity of factor Xa] requires the assemblage of the tenase complex (Ca <sup>2+</sup> and factors VIIIa, IX, and X) on the surface of the activated platelets.
Necessity Frame: [Enabled Situations The activation of factor Xa] requires [Precondition the assemblage of the tenase complex (Ca <sup>2+</sup> and factors VIIIa, IX, and X) on the surface of the activated platelets.]
* Molecular Assembly Frame: The activation of factor Xa requires the assemblage [Resulting Assembly of the tenase complex (Ca <sup>2+</sup> and factors VIIIa, IX, and X)] [Location on the surface of the activated platelets.]
* Molecular complex Frame: The activation of factor Xa requires the assemblage of the [Function tenase] complex (Ca <sup>2+</sup> and factors VIIIa, IX, and X) [Location on the surface of the activated platelets].
Part inner outer Frame: The activation of factor Xa requires the assemblage of the tenase complex (Ca <sup>2+</sup> and factors VIIIa, IX, and X) on the surface [Whole of the activated platelets].
* Activation Frame: The activation of factor Xa requires the assemblage of the tenase complex (Ca <sup>2+</sup> and factors VIIIa, IX, and X) on the surface of the activated [Activated Entity platelets].

Figure 64 : Cadres dans BioFrameNet avec des contextes annotés (Harabagiu et Bejan 2010)

Toutefois, BioFrameNet et FrameNet se distinguent l'un de l'autre sur les plans suivants. Dans un premier temps, BioFrameNet fait appel à des ontologies préexistantes et à des banques terminologiques pour le vocabulaire technique, tandis que FrameNet fait appel à un corpus de langue générale. En ce qui concerne les cadres, dans BioFrameNet, les cadres et leurs éléments sont liés à des ontologies externes. Dans un second temps, les définitions dans BioFrameNet sont fournies par des experts, tandis que les définitions de FrameNet sont fournies en faisant appel à l'intuition du lexicographe, au corpus et aux dictionnaires.

### 4.4.3.3 Framed DiCoEnviro : A framed version of DiCoEnviro<sup>37</sup>

Le DiCoEnviro est une ressource terminologique multilingue (anglais, français et espagnol) de l'environnement. Tout comme le DiCoInfo (voir section 3.1.3), cette ressource est basée sur les principes de la Lexicologie explicative et combinatoire (LEC) de Mel'čuk (Mel'čuk et al. 1995). Dans le DiCoEnviro, les termes sont traités en fonction de leur appartenance au domaine (ce qui couvre plusieurs sous-domaines comme le changement climatique, l'énergie renouvelable et la gestion des matières résiduelles) et leurs propriétés lexico-sémantiques. La description des termes propose une définition de la structure actancielle du terme et ses liens lexicaux. De plus, jusqu'à vingt contextes annotés sont associés au terme montrant comment les actants et les circonstants se combinent avec celui-ci. Considérons la figure 65 pour l'entrée *recycler*.

**recycler** <sub>1</sub>, v. tr.

Statut : 2

Domaine : gestion des matières résiduelles

recycler : [entreprise](#) ~ [matières](#)

**Contextes**

**Contextes annotés**

- Nous sommes loin derrière les Pays-Bas qui recyclaient déjà 75 % de ses matières résiduelles CRD en 1999 (Saskatchewan waste réduction council, 2005). (Source : PITRE)
- Comme nous l'avons mentionné ci-dessus, 8,1 % des matières résiduelles générées sur le site sont récupérées en vue d'être recyclées, ce qui équivaut à 10,2 tm par année. Le papier et le carton représentent respectivement 60 % et 30 % de ce tonnage. (Source : RIBESSE)
- L'Allemagne, la Belgique, la Suède, l'Estonie et l'Irlande sont les pays qui recyclent en plus grande proportion les déchets municipaux. (Source : RQBILAN08)

**Liens lexicaux**

Explication	Lexie reliée
<b>Voisins</b>	
=	<a href="#">récupérer</a> <a href="#">réutiliser</a> <sub>1</sub> <a href="#">valoriser</a> <sub>1</sub>
<b>Contraires</b>	
Opposé	<a href="#">éliminer</a> <sub>1</sub>
Opposé	<a href="#">jeter</a> <sub>1</sub>
<b>Autres parties du discours et dérivés</b>	
Nom	<a href="#">recyclage</a> <sub>1</sub>
Des <b>matières</b> qui ont été r.	<a href="#">recyclé</a> <sub>1</sub>
Des <b>matières</b> qui peuvent être r.	<a href="#">recyclable</a> <sub>1</sub>
Nom de l' <b>agent</b>	<a href="#">recycleur</a> <sub>1</sub>
<b>Autres</b>	
Lieu où on r.	centre de recyclage entreprise de recyclage usine de recyclage

[anglais : recycle](#)<sub>1</sub>

Figure 65 : Entrée du verbe *recycler* dans le DiCoEnviro (DiCoEnviro 2015)

La structure actancielle du verbe est représentée dans le schéma : entreprise [Agent] ~ matières [Patient]. Les liens lexicaux décrivent, entre autres, des relations d'opposition, comme

<sup>37</sup> Nous signalons que les deux ressources (DiCoEnviro et sa version en cadres) sont en construction.

*éliminer*, et de voisinage, comme *réutiliser*. Finalement, les contextes annotés mettent en évidence les actants et les circonstants, comme dans l'exemple suivant :

*On [Agent] est donc habileté à transformer, à RECYCLER et à récupérer, ici [Lieu], toutes sortes de matières plombifères [Patient].*

Le DiCoEnviro a fait l'objet d'une modélisation à la FrameNet, concrétisé dans une ressource appelée DiCoEnviro (*Framed version*) (L'Homme et Robichaud 2014). Dans cette ressource, la terminologie de l'environnement est décrite au moyen des cadres en utilisant comme point de départ des cadres existants dans FrameNet. La ressource propose également des cadres nouveaux. Dans la réalisation de ce travail, un certain nombre d'adaptations ont été faites. Par exemple, dans la version Framed, les noms des rôles sémantiques des actants ont été conservés et les actants (et circonstants) sont définis au niveau des unités lexicales et non pas des cadres (rappelons que les éléments du cadre sont définis au niveau du cadre dans FrameNet). Cela étant dit, les auteurs (L'Homme et al. 2014 : 1365) constatent que leur approche permet de :

1. Unveil frames in specific subject fields based on predicative terms such as verbs and nouns (more specifically, frames in the field of the environment) that share lexico-semantic properties.
2. Define new frames specific to the field of the environment: Although some frames might correspond to frames already defined in FrameNet, it is likely that new frames will be discovered in the data analyzed.
3. Validate the notion of "frame" on data that was not previously encoded for that purpose.

L'identification des cadres dans le DiCoEnviro se base sur une méthodologie comportant les étapes suivantes (L'Homme et Robichaud 2014 : 190) : 1) extraire les données de la ressource terminologique ; et 2) utiliser les cadres de FrameNet comme référence pour identifier les cadres existants que les termes dans DiCoEnviro peuvent évoquer. Pour ce faire, une première étape consiste à comparer les données comprises dans les deux ressources, DiCoEnviro et FrameNet. Une page a été mise en place qui permet de montrer les correspondances entre les UL de DiCoEnviro et les cadres de FrameNet (figure 66).



DiCoEnviro			FrameNet																																																																				
<p><b>emit_1</b> (vt)</p> <p>SA: emit: <b>Source</b>{ atmosphere 1 , Earth } ~ <b>Patient</b>{ energy } to <b>Destination</b>{ atmosphere 1 }</p> <p><b>Emitting</b> (nombre d'actants différent) :</p> <p>Definition(s):</p> <p>en: An <b>Agent</b> or a <b>Source</b> discharges a <b>Patient</b> into a <b>Destination</b>.</p> <table border="1"> <thead> <tr> <th>DiCoEnviro</th> <th></th> <th>FrameNet</th> </tr> </thead> <tbody> <tr> <td><b>Agent</b></td> <td>↔</td> <td>?</td> </tr> <tr> <td><b>Source</b></td> <td>↔</td> <td>Source_emitter</td> </tr> <tr> <td><b>Patient</b></td> <td>↔</td> <td>Emission</td> </tr> <tr> <td><b>Destination</b></td> <td>↔</td> <td>?</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>en</th> <th>fr</th> <th>es</th> </tr> </thead> <tbody> <tr> <td>emission.1</td> <td>injecter.1</td> <td>emitir.2</td> </tr> <tr> <td>emission.2.1</td> <td>libération.1</td> <td></td> </tr> <tr> <td>emit.1</td> <td>libérer.1a</td> <td></td> </tr> <tr> <td>emit.2</td> <td>libérer.1b</td> <td></td> </tr> <tr> <td>inject.1</td> <td>émettre.1</td> <td></td> </tr> <tr> <td>radiate.1</td> <td>émettre.2</td> <td></td> </tr> <tr> <td>release.1a</td> <td>émission.1</td> <td></td> </tr> <tr> <td>release.1b</td> <td>émission.2.1</td> <td></td> </tr> <tr> <td>release.1b.1</td> <td></td> <td></td> </tr> </tbody> </table> <p><a href="#">16 contextes...</a></p>			DiCoEnviro		FrameNet	<b>Agent</b>	↔	?	<b>Source</b>	↔	Source_emitter	<b>Patient</b>	↔	Emission	<b>Destination</b>	↔	?	en	fr	es	emission.1	injecter.1	emitir.2	emission.2.1	libération.1		emit.1	libérer.1a		emit.2	libérer.1b		inject.1	émettre.1		radiate.1	émettre.2		release.1a	émission.1		release.1b	émission.2.1		release.1b.1			<p><b>emit</b> (V) <span style="float: right;">ID: 6233</span></p> <p>COD: discharge; send forth or give out.</p> <p><b>Emitting</b>: In this frame a <b>Source_emitter</b> discharges its <b>Emission</b> along a <b>Path</b> or to a <b>Goal</b> .</p> <p><a href="#">3 Examples</a></p> <p><a href="#">Annotated Contexts</a></p> <table border="1"> <thead> <tr> <th>Type</th> <th>FE</th> </tr> </thead> <tbody> <tr> <td rowspan="2">Core</td> <td>Emission</td> </tr> <tr> <td>Source_emitter</td> </tr> <tr> <td rowspan="4">Extra-Thematic</td> <td>Distance</td> </tr> <tr> <td>Form</td> </tr> <tr> <td>Reason</td> </tr> <tr> <td>Sub_source</td> </tr> <tr> <td rowspan="10">Peripheral</td> <td>Area</td> </tr> <tr> <td>Carrier</td> </tr> <tr> <td>Degree</td> </tr> <tr> <td>Duration</td> </tr> <tr> <td>Goal</td> </tr> <tr> <td>Manner</td> </tr> <tr> <td>Means</td> </tr> <tr> <td>Path</td> </tr> <tr> <td>Speed</td> </tr> <tr> <td>Time</td> </tr> </tbody> </table> 			Type	FE	Core	Emission	Source_emitter	Extra-Thematic	Distance	Form	Reason	Sub_source	Peripheral	Area	Carrier	Degree	Duration	Goal	Manner	Means	Path	Speed	Time
DiCoEnviro		FrameNet																																																																					
<b>Agent</b>	↔	?																																																																					
<b>Source</b>	↔	Source_emitter																																																																					
<b>Patient</b>	↔	Emission																																																																					
<b>Destination</b>	↔	?																																																																					
en	fr	es																																																																					
emission.1	injecter.1	emitir.2																																																																					
emission.2.1	libération.1																																																																						
emit.1	libérer.1a																																																																						
emit.2	libérer.1b																																																																						
inject.1	émettre.1																																																																						
radiate.1	émettre.2																																																																						
release.1a	émission.1																																																																						
release.1b	émission.2.1																																																																						
release.1b.1																																																																							
Type	FE																																																																						
Core	Emission																																																																						
	Source_emitter																																																																						
Extra-Thematic	Distance																																																																						
	Form																																																																						
	Reason																																																																						
	Sub_source																																																																						
Peripheral	Area																																																																						
	Carrier																																																																						
	Degree																																																																						
	Duration																																																																						
	Goal																																																																						
	Manner																																																																						
	Means																																																																						
	Path																																																																						
	Speed																																																																						
	Time																																																																						

Figure 66 : Correspondance entre l'UT *emit* et l'UL *emit* du cadre Emitting dans FrameNet

Dans le cas du verbe *emit*, le cadre correspondant est Emitting. D'après cette page de comparaison, les éléments affichés mettent en lumière les composants suivants de chaque ressource : pour le DiCoEnviro, on trouve l'unité terminologique (UT), la structure actancielle, la définition, les équivalents et les liens lexicaux en anglais, français et espagnol; pour FrameNet, on trouve l'UL, le cadre, les FE avec leurs types et les relations. Dans notre exemple, nous pouvons remarquer que l'UT dans DiCoEnviro ne s'accorde pas avec l'UL dans FrameNet en termes de nombre d'actants/core-FE.

Cependant, cette méthodologie doit faire face à un certain nombre de défis, surtout lorsqu'il est question des différences entre les deux ressources en question. Les différences constatées par les auteurs (L'Homme et al. 2014 : 1368-1369 et L'Homme et Robichaud 2014 : 191-192) sont :

1. Dans FrameNet, les FE sont définis au niveau du cadre, alors que dans le DiCoEnviro, les actants sont définis au niveau de l'UT;
2. Le nombre d'actants peut être différent du nombre de core-FE dans un cadre donné;

3. Les noms d'actants sont génériques dans le DiCoEnviro, alors que dans FrameNet les noms des FE sont spécifiques aux cadres;
4. Dans FrameNet, un FE peut avoir différents noms. Dans DiCoEnviro, un actant peut être divisé, par exemple, Cause et Agent.

Pour rendre compte des spécificités de chaque ressource, cinq catégories sont définies pour les correspondances entre les UT et les cadres (L'Homme et Robichaud 2014 : 193) :

1. Entièrement compatible : correspondance parfaite entre les UT et les cadres, en terme de définition et de nombre d'actants/FE;
2. Alternance : cette catégorie rend compte d'une particularité du DiCoEnviro, c'est-à-dire que lorsqu'une UT a deux entrées dans la ressource mais que toutes les deux évoquent le même cadre. Par exemple, *emit* a deux structures actanciennes : *emit*<sub>1</sub> (Source ~ Patient to Destination) et *emit*<sub>2</sub> (Agent or Cause ~ Patient into Destination). L'alternance qui existe dans ce cas est entre Agent (et Cause) et Source;
3. Partiellement compatible : présence d'une différence entre les UT et les cadres, en termes de nombre d'actants/core-FE;
4. Nouveau : lorsqu'il n'existe aucune correspondance entre les données dans les deux ressources, de nouveaux cadres sont créés;
5. En attente : certains UT sont associés d'une manière provisoire à des cadres.

Finalement, une interface web a été mise en ligne pour afficher les données encodées dans le DiCoEnviro (figure 67).

Definition:  
An **Agent** or a **Source** discharges a **Patient** into a **Destination**.

Example(s):  
when flourishing, **forests** absorb about as much **carbon** as they **EMIT**  
**Aircrafts** **EMIT** **gases and particles** directly into the upper troposphere and lower stratosphere .  
Currently, **warm tropical ocean waters** brought north eastwards by this system **RELEASE** enough heat to the air to keep temperatures over much of Europe some 10 °C warmer than land areas at similar latitudes in other parts of the Northern Hemisphere

Notes:  
This frame is based on **Emitting** in FrameNet. The number of actants vs. core FEs differs. There is an Agent/Source alternation.

Participants (1): 1. <b>Agent</b> 2. <b>Source</b> 3. <b>Patient</b> 4. <b>Destination</b>	Participants (2): 1. Manner (10) 2. Direction (9) 3. Substitute (5) 4. Time (4) 5. Method (3) 6. Frequency (3) 7. Purpose (3) 8. Instrument (2) 9. Duration (1) 10. Location (1) 11. Value (1) 12. Result (1) 13. Reason (1) 14. Path (1) 15. Degree (1) 16. Beginning (1) 17. End (1)	Frame Relations:  Legend: Precedes →
English LUs: • <u>emission_1</u> • <u>emission_2_1</u> • <u>emit_1</u> • <u>emit_2</u> • <u>inject_1</u> • <u>radiate_1</u> • <u>release_1a</u> • <u>release_1b</u> • <u>release_1b_1</u>	French LUs: • <u>injecter_1</u> • <u>libération_1</u> • <u>libérer_1a</u> • <u>libérer_1b</u> • <u>émettre_1</u> • <u>émettre_2</u> • <u>émission_1</u> • <u>émission_2_1</u>	Spanish LUs: • <u>emitir_2</u>

Figure 67 : Affichage du cadre Emitting dans DiCoEnviro *framed version* (DiCoEnviro Framed Version 2015)

Dans le DiCoEnviro *framed version*, le cadre est défini en mettant en évidence les actants compris dans celui-ci (Agent ou Source, Patient et Destination). Des exemples sont également donnés comprenant l'UT en question et soulignant les core-FE. À la suite des exemples, une note explicative est ajoutée pour préciser l'origine du cadre. Cette note est conforme aux catégories déjà discutées ci-dessus. Dans le cas de l'exemple dans la figure 67, le cadre est basé sur un cadre existant dans FrameNet, mais le nombre d'actants diffère de celui des core-FE. De plus, il y existe un cas d'alternance (catégorie deux). Une section avec tous les actants et circonstants est ajoutée. On y trouve deux colonnes : Participant (1), qui regroupe les actants ; et Participants (2), qui regroupe les circonstants. Bien entendu, les différents actants et circonstants sont distingués au moyen des couleurs. La dernière section est dédiée aux UT (ou UL) qui appartiennent au cadre. Ici, nous pouvons remarquer que les terminologies du DiCoEnviro ont mis dans cette section les UT des trois langues représentées dans la ressource, à savoir l'anglais, le français et l'espagnol.

Tout comme FrameNet, les relations entre les cadres sont définies dans le DiCoEnviro *framed version*. Dans la partie à droite de chaque cadre, les relations sont représentées via des graphes. Les flèches reliant les cadres précisent la nature des liens entre chacun. Dans la définition des relations entre les cadres, les terminologies de DiCoEnviro se sont basés sur les relations de FrameNet (L’Homme et Robichaud 2014 : 193). Par exemple, dans la figure 67, le cadre Emitting est relié à deux cadres, Reflecting et Soaking\_up. Le type des relations reliant ces trois cadres est Precedes.

Finalement, le Framed DiCoEnviro propose des scénarios reliant plusieurs cadres. Par exemple, on y trouve le scénario Cycle\_of\_existence\_scenario (figure 68).

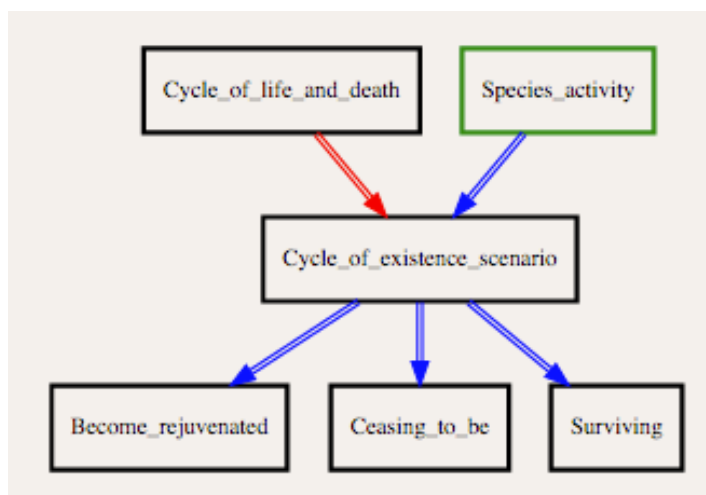


Figure 68 : Cycle\_of\_existence\_scenario dans DiCoEnviro Framed Version (DiCoEnviro Framed Version 2015)

Le cadre Cycle\_of\_existence\_scenario est un cadre général qui a pour but de relier sémantiquement les autres cadres. Ce scénario relie cinq cadres : Cycle\_of\_life\_and\_death, Species\_activity, Become\_rejuvnated, Ceasing\_to\_be et Surviving. Chacun de ces cadres est relié au scénario par une relation. Les cadres Species\_activity, Become\_rejuvnated, Ceasing\_to\_be et Surviving sont des sous-cadres (Subframe) du scénario Cycle\_of\_existence\_scenario. Tandis que

Cycle\_of\_life\_and\_death entretient une relation d'héritage (Is inherited by) avec le premier.

#### **4.4.4 Choix de cadre théorique et ce qui distingue notre projet**

Dans le présent chapitre, nous avons présenté le cadre théorique de notre travail de recherche. Nous avons discuté les différents aspects de la Sémantique des cadres et décrit son application sur le web, FrameNet. Nous avons également présenté quelques applications de la Sémantique des cadres en ASM général et en ASM de spécialité. Enfin, nous nous sommes attardé sur des applications de FrameNet en langue de spécialité pour les domaines de l'environnement et de la biologie. Nous trouvons pertinent à ce stade de justifier notre choix de cadre théorique et d'expliquer pourquoi à notre avis la théorie de la Sémantique des cadres correspond parfaitement à notre sujet.

La théorie de la Sémantique des cadres nous semble adaptée à notre projet pour deux raisons. Premièrement, étant donné que nous travaillons sur les unités prédicatives verbales et nominales, la Sémantique des cadres fournit des principes théoriques et méthodologiques particulièrement utiles pour l'étude de verbes et de noms déverbaux. En effet, dans le contexte des cadres évoqués par ce genre d'unités, Ruppenhofer et al. (2010 : 5-6) constatent que : « the frame-evoking LU is a verb and the FEs are its syntactic dependents » et « the frames evoked by artifact and natural kind nouns rarely dominate the clauses in which they occur, and so are seldom selected as targets of annotation ». De ce point de vue, nous estimons que cette théorie nous aidera à déceler les différents composants de la structure actancielle des verbes et des noms déverbaux pour pouvoir, par la suite, les structurer dans des cadres spécifiques au domaine de la technologie informatique. En outre, la Sémantique des cadres nous permettra de procéder à l'étude et à l'analyse des propriétés linguistiques de l'ASM de spécialité.

Deuxièmement, la Sémantique des cadres est une approche appropriée pour l'étude de la langue de spécialité. Nous avons montré dans les sections 4.4.3 comment différents projets s'appuient sur cette théorie. À notre avis, FrameNet propose un modèle tout à fait approprié pour la représentation du lexique d'un domaine de spécialité. Sur ce sujet, Fillmore constate que : « I suppose a reasonable « finished » FrameNet could be a model for exploring the vocabulary of specialists areas, by showing how each area's lexical units provide the means of

introducing related concepts (through their syntactic /semantic valence) » (Andor 2010 : 166). Partant des résultats obtenus dans la conception de DiCoEnviro *framed version*, L'Homme (2015 : 14) appuie le point de vue de Fillmore en concluant que « frames can be a useful way to capture a specific kind of structure in specialized domains ».

En ce qui concerne notre recherche, nous estimons que la Sémantique des cadres est un modèle d'analyse et de structuration d'unités terminologiques prédicatives (UTP). Les UTP, qui sont réalisées par les verbes et les noms déverbaux, sont des unités transmettant des contenus spécialisés relatifs au domaine de l'informatique. Nous nous appuyons sur ce modèle afin de voir dans quelle mesure nous pouvons organiser ces unités en fonction des cadres que nous définissons. Donc, il ne s'agit pas uniquement d'étudier ou d'analyser les unités prédicatives du domaine en montrant leurs propriétés sémantiques et syntaxiques et leur structure actancielle, mais aussi de proposer un modèle d'organisation qui se base sur les cadres pour structurer le domaine de la technologie informatique. De plus, notre choix du modèle de FrameNet pour notre projet, plutôt qu'un autre modèle, comme WordNet ou VerbNet, provient du fait que FrameNet se distingue des autres ressources par ce qui suit (Boas 2009 : 20) :

it makes use of independent organizational units that are larger than words, i.e., semantic frames (see also Atkins 2002, Ohara et al. 2003, Boas 2005b, Atkins and Rundell 2008). As such, FrameNet facilitates a comparison of the comprehensive lexical descriptions and their manually annotated corpus-based example sentences with those of other LUs (also of other parts of speech) belonging to the same frame. Another advantage of the FrameNet architecture lies in the way lexical descriptions are related to each other. Using detailed semantic frames which capture the full background knowledge evoked by all LUs of the same frame makes it possible to systematically compare and contrast their numerous syntactic valence patterns (see Atkins 2002, Boas 2005a).

Finalement, nous avons montré comment les données, dans les différentes ressources étudiées de l'ASM, sont importées de projets déjà existants par voie de traduction. Nous avons également montré les projets en cours pour implémenter un modèle de description de l'ASM (général et de spécialité) basé sur la Sémantique des cadres et le projet FrameNet. Or, ces projets s'appuient souvent sur des ressources existantes et les corpus utilisés ne sont pas compilés par les lexicographes. La particularité de notre projet est qu'il fait appel à un corpus que nous avons constitué nous-même. Ce corpus est le reflet de la langue de spécialité tel qu'elle est utilisée actuellement.

## 5. Méthodologie

Partant du principe que tout travail terminologique doit être basé sur une méthodologie qui oriente le projet selon ses objectifs, nous présentons dans le présent chapitre notre méthodologie de traitement et d'analyse des unités terminologiques nominales et verbales (UTN et UTV). Nous parlerons de différentes étapes de collecte et d'encodage des données et, par la suite, nous aborderons la question des cadres sémantiques (Atkins et al. 2003) dans le contexte de notre projet de thèse. Nous allons procéder comme suit : d'abord, à la section 5.1, nous décrirons notre corpus, ses caractéristiques et sa constitution; puis, à la section 5.2, nous présentons notre méthode semi-automatisée d'extraction des UTP. Ensuite, à la section 5.3, nous validons les UTP candidates obtenues suite à l'extraction. À la section 5.4, nous parlons de notre choix de contextes pour les UTP; et à la section 5.5 nous abordons la question de l'annotation des contextes. La dernière étape, décrite à la section 5.6, est consacrée à l'organisation des UTP dans des cadres, selon la méthodologie suivie dans le projet FrameNet (Ruppenhofer et al. 2010). Pour étudier les UTP de la technologie informatique, nous partons des données trouvées dans le corpus. Donc, nous adoptons une approche dite *bottom-up* (Teubert 2009).

### 5.1 Corpus

Contrairement à la linguistique chomskienne qui met l'accent sur la « compétence » afin d'explorer la langue, l'analyse linguistique basée sur le corpus s'intéresse à l'usage de la langue dans des contextes réels (Adolphs et Lin 2011 : 597). Cela se vérifie d'autant plus en terminologie, puisque l'analyse et la description des UT d'un domaine nécessitent l'extraction de ces dernières. La source de base de tout travail terminologique est le corpus (Cabré 1999 : 121). Ainsi, L'Homme (2004 : 119) décrit cette réalité comme suit :

La recherche terminographique repose principalement sur le contenu de textes de spécialité. La collecte d'une documentation représentative du domaine dont on souhaite décrire la terminologie et son exploitation constituent les premières étapes d'une recherche en bonne et due forme.

Pearson (1998 : 43) fait observer que le corpus est un : « artefact; it is selected, chosen or assembled according to explicit criteria. It is stored in electronic form. It consists of pieces of naturally occurring language ». À partir de ce constat, nous comprenons qu'il existe un certain nombre de critères selon lesquels les textes de corpus sont choisis. Le fait qu'il soit sous format électronique ouvre la voie à différents processus de traitement automatique. De plus, le corpus est un échantillon d'une langue dans son état naturel et peut avoir une variété de formes; chacune a son propre usage et sa finalité. À titre d'exemple, nous mentionnons le corpus général, le corpus de spécialité, le corpus parallèle, le corpus comparable et le corpus de référence.

Ce qui nous intéresse ici est le corpus de spécialité. Ce genre de corpus, contrairement au corpus général, porte sur un sujet précis. Cabré (1999 : 121) présente trois critères auxquels doivent répondre les corpus de spécialité :

1. they must be representative of the subject matter, in accordance with the objectives of the task and the delimitation of the topic, so that they permit the preparation of a sufficiently large initial list of units;
2. they must be up-to-date both regarding the designations that experts really use and the topic;
3. they must be explicit enough to allow retrieval of the identification and information at any point in time during the process or during dissemination of the terminological data.

Donc, un corpus de spécialité est monté en fonction des objectifs spécifiques du travail terminologique; il est délimité en ce qui concerne son sujet; il est à jour, pour refléter l'usage courant des concepts et des termes; et il est accessible afin de permettre l'identification et la récupération d'informations.

Dans cette perspective, nous avons monté notre corpus en tenant compte des éléments qui suivent. Notre corpus porte sur la technologie informatique. Par *technologie informatique*, nous entendons tout ce qui concerne l'ordinateur et ses différents composants, les tablettes, les cellulaires et leurs applications, les systèmes d'exploitation ainsi que les réseaux informatiques, ce qui inclut l'Internet. La langue de notre corpus est la langue arabe, ou, plus spécifiquement, l'arabe standard moderne (ASM). Avant de discuter de ces éléments, il convient tout d'abord de nous attarder sur la particularité d'un corpus de spécialité pour l'ASM.



### 5.1.1 Particularité et importance d'un corpus arabe de spécialité

Dans la littérature, on parle peu de la constitution de corpus à des fins terminographiques pour l'ASM de spécialité. En ce qui concerne la langue générale, il n'existe pas, à notre connaissance, un dictionnaire moderne de l'ASM constitué à partir d'un corpus (à l'exception d'un dictionnaire arabe-néerlandais/néerlandais-arabe développé par Van Mol 2000).

À partir des années 1990, les spécialistes en traitement automatique des langues et en linguistique de corpus ont commencé à s'intéresser à la langue arabe (Al-Sultani et Atwell 2006; Abdelali et al. 2005). Ils ont vite constaté un manque de ressources, ce qui les a poussés à entamer des efforts réels pour procéder à la compilation de corpus de l'ASM. Nous mentionnons à titre d'exemple le corpus de l'Université de Leeds, le corpus de l'Université de Lancaster (KALIMAT) (El-Haj et Koulali 2013), le corpus de *International Corpus of Arabic* de la bibliothèque d'Alexandrie (Alansary et al. 2007) et le corpus de ar-TenTen, qui est considéré comme étant le plus vaste et qui comprend presque 6 milliards de mots (Arts et al. 2014).

Cependant, peu d'efforts comparables sont déployés dans le domaine de la terminologie. À notre connaissance, il n'existe pas un corpus de la langue arabe de spécialité diffusé par un centre de recherche ou une université à des fins terminologiques ou de traitement automatique des langues. Al-Sultani et Atwell (2006) font état de deux corpus généraux tirés de deux journaux, *Al-Nahar* et *Al-Hayat*, qui comprennent des sous-corpus portant sur certains domaines de spécialité, dont l'informatique. Ces corpus sont accessibles par abonnement<sup>38</sup>.

Même s'il ne s'agit pas de son principal objectif, notre projet veut contribuer au développement de corpus spécialisés dans le domaine de la terminologie et de la linguistique de corpus, compte tenu de l'importance accordée au corpus dans tout travail terminographique de qualité.

---

<sup>38</sup> Les deux corpus sont disponibles sur le site internet de l'ELRA (European Language Ressources Association) à l'adresse : [http://catalog.elra.info/product\\_info.php?products\\_id=632](http://catalog.elra.info/product_info.php?products_id=632)

Par ailleurs, le contenu textuel arabe sur la Toile est en pleine expansion. Van Mol (2000 : 3) constate qu'un travail lexicographique en langue arabe basé sur corpus a l'avantage de rendre possible la découverte de nouveaux sens : « that are not represented in the available dictionaries and to define more precisely the modern vocabulary of a language ». À l'heure actuelle, les particularités que nous pouvons observer dans la compilation d'un corpus en ASM de spécialité sont d'ordre linguistique et régional. Sur le plan linguistique, nous avons montré dans Section 2.2 la particularité de l'ASM par rapport aux autres variantes de la langue arabe. Nous croyons qu'il serait difficile de monter un corpus de spécialité en l'une des variantes régionales de l'espace arabophone, puisqu'elles ne sont presque pas employées dans les documents officiels, les sites web, les recherches scientifiques, etc.

Sur le plan régional, nous pouvons dire que les variations en langue arabe de spécialité diffèrent selon le domaine. Dans le cadre de l'informatique, nous croyons que les variations sont nombreuses, dû aux raisons que nous avons discutées à la section 2.4.1. Cependant, à l'heure actuelle, deux facteurs interviennent qui stabilisent cette variation. Premièrement, la dominance de certains pays arabophones (notamment la Jordanie, l'Égypte, la Syrie et l'Arabie Saoudite) dans la production scientifique et technologique. Cela signifie que le contenu textuel visible sur la Toile provient majoritairement de ces pays (donc, pas question d'une diversité importante). Deuxièmement, le rôle important joué par la presse électronique dans la standardisation et la dissémination de la terminologie. De nos jours, nous croyons que la normalisation des termes se fait par la presse électronique (et évidemment les entreprises) et non pas par les académies ou les instances officielles<sup>39</sup> (voir section 2.4.1). C'est pour cette raison que les nouveaux sites technologiques, les manuels ou les recherches portant sur des questions techniques en ASM tentent d'employer une terminologie standardisée afin de produire un contenu de qualité.

Dans ce qui suit, nous décrivons notre corpus et nous détaillons ses propriétés. Nous allons nous attarder sur un certain nombre de critères (Bowker et Pearson 2002 : 45-52) qui décrivent le mieux notre corpus et sa construction. Ces critères sont : le sujet, la langue, la

---

<sup>39</sup> Par exemple, les académies arabes, notamment celles de Damas, de Bagdad et du Caire.

provenance géographique, les types de textes, la taille, les textes qui composent le corpus et leur nombre, le format, la paternité (*authorship*) et la date de publication.

### **5.1.2 Le sujet**

Le premier critère que nous abordons concerne le sujet traité dans les textes. Comme mentionné au début de ce chapitre, notre corpus porte sur la technologie de l'informatique. Nous avons décidé d'inclure une grande variété de sujets liés à ce domaine. Étant donné que l'informatique est devenue un domaine multidisciplinaire, notre décision d'inclure des textes portant sur différents aspects provient du fait que nous croyons que les frontières entre ses sous-domaines ne sont pas très étanches. À titre d'exemple, en ce qui concerne les systèmes d'exploitation, que ce soit pour les cellulaires, les tablettes ou les ordinateurs, la terminologie est la même (avec de petites nuances, bien entendu). Nous avons remarqué, tout au long de notre lecture sur ce domaine assez large, que le système d'exploitation de la technologie cellulaire se rapproche de plus en plus de celui des ordinateurs. Dans notre recherche, nous ne faisons pas la distinction entre les différents sous-domaines. Par contre, nous montrons au moyen de contextes, la provenance de chaque terme. Dans le tableau 2 nous montrons pour chaque source, les sujets abordés dans les textes choisis.

SOURCE	DÉPANNAGE	TUTORIEL	LOGICIEL	MATÉRIEL	APPLICATION CELLULAIRE	TABLETTE	INTERNET
AITNEWS	Oui	Oui	Oui	Non	Oui	Oui	Oui
TECH-WD	Oui	Oui	Oui	Non	Oui	Oui	Oui
ASK MOST	Oui	Oui	Oui	Non	Oui	Non	Oui
ARABHARDWARE	Non	Non	Oui	Oui	Non	Non	Oui
KEEFTECH	Oui	Oui	Non	Non	Oui	Non	Oui
ABAHE	Oui	Non	Non	Non	Non	Non	Non
LINUX	Non	Oui	Oui	Non	Non	Non	Non
PYTHON-AR	Non	Oui	Oui	Non	Non	Non	Non
UBUNTU-AR	Non	Oui	Oui	Non	Non	Non	Non
WIKIPÉDIA	Non	Non	Oui	Oui	Non	Non	Non
AL-JAZEERA	Non	Non	Oui	Oui	Oui	Non	Oui
TAIBAHU	Non	Oui	Oui	Non	Non	Non	Oui
APPLE-WD	Non	Oui	Oui	Non	Oui	Non	Non
ITP	Non	Non	Oui	Non	Oui	Non	Oui
TIQANIAT	Non	Oui	Oui	Oui	Oui	Non	Oui
OS	Non	Oui	Oui	Non	Non	Non	Non
MICROSOFT	Non	Oui	Oui	Non	Non	Non	Non
A-PLUS	Oui	Oui	Non	Oui	Non	Non	Non
UNLIMIT-TECH	Oui	Oui	Oui	Non	Oui	Oui	Oui

Tableau 2 : Sujets abordés dans les textes du corpus

### 5.1.3 Langue du corpus

Notre corpus est monolingue. Il est composé entièrement de textes rédigés en ASM. Ces textes sont soit des documents originaux, soit des traductions. Dans le cas des manuels de programmation, il s'agit de traductions de l'anglais (à l'exception d'un manuel portant sur Python, qui est une traduction du français). En ce qui concerne les documents provenant de sites web, ils sont originaux et rédigés par des natifs de la langue arabe.

Notre choix d'inclure des documents traduits dans notre corpus repose sur deux raisons. Dans un premier temps, étant donné que les connaissances liées à l'informatique sont surtout diffusées en anglais, il est difficile d'éviter les documents traduits de cette langue,

surtout lorsqu'il s'agit de constituer un corpus en ASM de spécialité de bonne qualité. Dans un second temps, afin de produire un corpus diversifié, couvrant plusieurs aspects de la terminologie ASM de l'informatique, nous avons trouvé nécessaire d'inclure ce genre de documents.

La question qui se pose maintenant est celle de l'homogénéité du corpus. Avec une telle diversité, nous constatons que le corpus reste homogène. En ce qui a trait aux documents traduits, la traduction de ces documents est faite par des spécialistes du domaine, c'est-à-dire des ingénieurs, des informaticiens, des développeurs, etc. La terminologie employée par ces spécialistes est standardisée puisqu'il s'agit de documents diffusés par une communauté informatique arabophone (Arab Linux Community). Pour ce qui est des textes provenant de sites Internet, ils sont également rédigés et diffusés par des institutions journalistiques en ligne (presse électronique). Ces institutions ont leur propre politique terminologique, dans le sens où la terminologie employée est standardisée et normalisée selon des normes déterminées par le corps journalistique.

L'autre question importante reliée à la langue est la qualité. Au cours du montage de notre corpus, nous nous sommes fixé les critères suivants quant à la qualité de la langue arabe dans les textes et la terminologie qui s'y trouve :

1. Les textes doivent être rédigés en ASM; donc ni les variantes (nous donnons l'exemple de variations dialectales qui existent entre les pays arabophones), ni la langue classique (par exemple, la langue du Coran) sont acceptées;
2. La langue doit être utilisée correctement, c'est-à-dire que les textes doivent comporter très peu de fautes de grammaire ou d'orthographe; et
3. Les auteurs doivent employer les termes arabes les plus répandus et connus des lecteurs, et, dans le cas de création terminologique, les néologismes doivent être faciles à comprendre.

### 5.1.4 Provenance géographique

Notre corpus ne couvre pas toutes les régions arabophones. Nous nous sommes concentré sur les sources les plus importantes et les plus répandues sur la Toile. Une partie de notre corpus est constituée d'articles de presse électronique, notamment des sites internet basés dans la région du Golfe, au Moyen-Orient (la Syrie et la Jordanie) et en Égypte. Par exemple, le site AitNews est basé à Dubaï, Émirats Arabes Unis. Son corps journalistique est composé de rédacteurs professionnels, essentiellement des informaticiens, des développeurs et des programmeurs arabophones (de plusieurs pays où la langue arabe est une langue officielle). Le site ne fait pas état de sa politique langagière ou terminologique, mais il apparaît assez évident que des normes de rédaction journalistiques sont imposées. Par exemple, nous pouvons y trouver le terme *maElwmatiy~a* (معلوماتية) (*informatique*), qui est considéré comme un terme normalisé dans plusieurs pays arabophones. Le terme même peut avoir des variantes, comme *iElAm Aliy* (إعلام آلي) (traduction littérale de *informatique*), que l'on trouve dans les pays du Maghreb, mais il n'a pas fait l'objet d'un consensus dans une communauté linguistique, et il n'est pas répandu, et souvent méconnu, dans les sociétés scientifiques arabophones. Donc, cette variante n'est pas retenue par les rédacteurs du site.

Il reste à dire que, dans le contexte de l'ASM de spécialité, la question reste limitée aux pays où l'on témoigne d'une effervescence technologique et une présence médiatique importante qui favorisent la production et la diffusion du contenu textuel en ASM.

### 5.1.5 Type de textes

Ce critère s'intéresse au type des textes ajoutés à un corpus. Nous avons retenu ce que Pearson appelle « communicative settings » (1999 : 35). Il s'agit des paramètres qui servent à définir les différents niveaux des textes dans le corpus. Ils sont au nombre de quatre : expert à expert, expert à initié, relativement expert à non-initié et professeur-étudiant. Nous passons rapidement en revue ces paramètres, puis nous expliquons comment ces niveaux ont été pris en compte dans notre corpus.

- Le niveau expert à expert concerne les communications scientifiques que partagent les spécialistes du domaine. Il s'agit d'un langage standardisé que les deux parties, lecteur et auteur, comprennent, et il est caractérisé par un degré élevé de spécialisation.
- Le niveau expert à initié survient dans une communication entre les experts du domaine et des personnes n'ayant pas le même niveau d'expertise. À ce niveau, les experts ont tendance à simplifier leur terminologie ou donner davantage d'explications sur les termes employés. Pour ce faire, on a normalement recours à la langue générale.
- Quant au niveau expert à non-initié, on s'attend à ce que les textes de ce niveau soient moins riches en termes, étant donné qu'ils sont destinés aux personnes n'ayant pas de formation dans le domaine en question. Lorsque l'auteur emploie un terme, soit il l'explique, ou bien il signale qu'il s'agit d'un concept connu de la part des lecteurs.
- Le dernier niveau est celui de professeur à élève. Ici, le lecteur présumé est une personne qui veut apprendre sur un domaine particulier. La terminologie employée est appropriée au domaine, mais l'auteur donne davantage d'explications et de définitions en langue générale ou dans une langue technique simplifiée.

Dans la compilation de notre corpus, nous avons diversifié nos sources pour garantir un maximum de représentativité. Dans tableau 3, nous montrons le niveau de chacune de nos sources, en fonction de ce que nous avons discuté dans cette sous-section.

	<b>EXPERT À EXPERT</b>	<b>EXPERT À INITIÉ</b>	<b>EXPERT À NON-INITIÉ</b>	<b>PROFESSEUR-ÉTUDIANT</b>
<b>AITNEWS</b>	Non	Non	Oui	Non
<b>TECH-WD</b>	Non	Non	Oui	Non
<b>ITP</b>	Non	Non	Oui	Non
<b>KEEF-TECH</b>	Non	Non	Oui	Non
<b>UNLIMIT-TECH</b>	Non	Non	Oui	Non
<b>ARABHARDWARE</b>	Non	Oui	Non	Non
<b>TIQANIAT</b>	Non	Non	Oui	Non
<b>APPLE-WD</b>	Non	Non	Oui	Non
<b>WIKIPÉDIA</b>	Non	Oui	Non	Non
<b>AL-JAZEERA</b>	Non	Non	Oui	Non
<b>ASKMOST</b>	Non	Non	Oui	Non
<b>ABAHE</b>	Non	Oui	Non	Non
<b>TAIBAHU</b>	Non	Non	Oui	Non
<b>LINUX</b>	Non	Oui	Non	Non
<b>PYTHON-AR</b>	Non	Oui	Non	Non
<b>UBUNTU-AR</b>	Non	Oui	Non	Non
<b>OS</b>	Non	Oui	Non	Non
<b>MICROSOFT</b>	Non	Non	Oui	Non
<b>A-PLUS</b>	Non	Oui	Non	Non

Tableau 3 : Correspondance des quatre paramètres communicatifs aux sources de notre corpus

D'après les données présentées dans le tableau, aucune de nos sources appartient au niveau expert à expert ou professeur-étudiant. Nous estimons que notre corpus se situe entre deux niveaux : expert à initié et expert à non-initié. La langue technique employée est simple, et aucune expertise (ou peu d'expertise) n'est requise de la part de lecteur pour comprendre le contenu. Dans le cas des sites Internet choisis, les auteurs ont tendance à expliquer les termes et parfois à donner des définitions. L'objectif n'est pas toujours de simplifier la technologie au lectorat, mais plutôt de présenter de nouveaux concepts peu familiers en ASM, étant donné que la technologie est plus facilement accessible en anglais.

En ce qui concerne les manuels de dépannage informatique, ils sont aussi destinés aux débutants (avec des connaissances de base en informatique) qui veulent s'initier ou acquérir de nouvelles compétences dans le domaine de l'informatique. La terminologie employée est simple et familière. Finalement, les manuels de programmation sont, quant à eux, rédigés par des spécialistes. Ils représentent un niveau un peu plus élevé que les manuels et les sites internet. Ils sont destinés aux amateurs de la programmation, donc, un niveau débutant ou



intermédiaire. Pour chaque manuel, il existe un glossaire des termes, et les équivalents des termes anglais sont donnés en ASM avec des définitions.

### **5.1.6 Taille du corpus**

La taille « idéale » d'un corpus spécialisé est toujours une question non résolue en terminologie (Pearson 1999 : 58). Pour nous, la taille occupe une place particulière, étant donné que nous travaillons sur des unités terminologiques moins fréquentes dans les textes, c'est-à-dire les UTP, par rapport aux unités nominales dénotant des entités. Pour cette raison, nous avons constitué un corpus de presque 500 000 mots<sup>40</sup>.

Une telle taille nous permet de trouver un nombre significatif d'occurrences d'UTN et d'UTV spécialisées. Dans cette perspective, Bowker et Pearson (2002 : 46) précisent que la taille de corpus dépend de : « factors such as the needs of your project, the availability of data and the amount of time that you have ». Notre but est de nous assurer d'avoir cerné la question de la variété du domaine de l'informatique en intégrant des textes qui portent sur plusieurs sujets. En outre, nous voulons avoir recours à un corpus représentatif de l'ASM de spécialité, c'est-à-dire nous assurer d'avoir à notre disposition des textes provenant des meilleurs sites techniques et des documents issus d'auteurs professionnels et connus de la part de la communauté technophile. Bien entendu, notre corpus à l'état actuel est fermé, ce qui veut dire que nous n'entendons pas l'élargir au fur et à mesure de l'avancement de notre recherche.

### **5.1.7 Les textes du corpus**

Un autre critère concerne les textes composant le corpus. Une première distinction est faite entre le texte complet ou des extraits de textes. Dans notre cas, les textes composant notre corpus sont des textes entiers. Nous avons opté pour cette méthode pour garantir la disponibilité des données et de toutes leurs formes. À titre d'exemple, si nous cherchons un verbe donné, il est probable que nous trouvions dans un même texte un synonyme pour ce verbe qui désigne la même action ou processus. Cela nous permet aussi d'avoir une meilleure

---

<sup>40</sup> Nous avons calculé le nombre de mots à l'aide du logiciel *aConCorde* (Rogers et al. 2005), qui est un concordancier comprenant un outil de calcul.

idée du style des auteurs surtout que l'ASM de la technologie est une langue en pleine évolution sur la Toile. De plus, choisir un texte entier nous donne accès à davantage d'informations sur nos candidats UTP, c'est-à-dire que nous pouvons récupérer autant d'informations définitives que nous désirons (Pearson 1999 : 59).

Finalement, notre choix de textes composant notre corpus est basé sur trois critères que nous nous sommes fixés, car nous ne nous intéressons qu'à certains aspects de la langue de spécialité. Ces critères sont :

1. les textes qui portent sur les spécifications techniques des appareils, par exemple, le mode de fonctionnement d'un ordinateur, d'une tablette ou d'une imprimante;
2. la description d'une plateforme d'un ensemble d'appareils, comme dans le cas de la plateforme Android ou Windows; et
3. des cours donnés par des spécialistes (programmeurs, développeurs, concepteurs, etc.) sur des logiciels, à titre d'exemple, un cours de plusieurs parties sur XML. Enfin, des rapports techniques sur certains aspects d'un environnement informatique.

### **5.1.8 Le nombre de textes**

Le nombre des textes est déterminé par la diversité des textes inclus dans le corpus et la diversité des auteurs de ces mêmes textes (Bowker et Pearson 2002 : 49). Dans notre corpus, nous avons choisi de diversifier nos sources dans le but de mieux comprendre les concepts et les termes les plus utilisés en ASM de spécialité. Par exemple, dans le cas des articles de presse électronique, les articles sont rédigés par plusieurs auteurs, et chaque site d'information technologique a sa propre politique langagière et terminologique.

### **5.1.9 Le format du corpus**

Ce critère s'intéresse au format du corpus. Notre corpus est de type texte. Nous nous intéressons au format écrit, car le corpus fait l'objet de divers traitements informatiques. Nous signalons que nous avons converti les textes du corpus en format *.txt*, ce qui facilite leur

traitement par différents logiciels de traitement automatique des langues (TAL). La forme écrite nous intéresse en particulier, car nous jugeons important de présenter le contenu textuel numérique de l'ASM sur la Toile.

### **5.1.10 Paternité**

Ce critère concerne l'auteur des textes. Bowker et Pearson (2002 : 51) estiment que : « the author of each text should be an acknowledged subject-field expert ». Elles ajoutent : « Although it is relatively easy to identify the author of a printed document, this is not always the case when dealing with electronic media, such as the Web ». De ce point de vue, notre corpus a fait l'objet d'une analyse rigoureuse afin d'identifier la provenance des textes. Pour ce faire, nous nous sommes fixé les critères suivants :

1. La source doit être entièrement dans la langue faisant l'objet de l'étude, donc pas de translittération ni de contenu en d'autres langues;
2. Les textes sélectionnés doivent mentionner la provenance, la date et le nom du rédacteur;
3. Dans le cas des sites internet, le site doit être mis à jour constamment;
4. Dans la rubrique « qui sommes nous », nous nous sommes assuré de trouver des informations relatives à l'identité du site (ou l'entreprise responsable du site, ou même le gestionnaire, etc.);
5. Les administrateurs du site doivent être joignables via la rubrique « contacter nous »;
6. Le site est présent sur les différents réseaux sociaux (Facebook, Twitter, Google plus, etc.);
7. Dans le cas des manuels, ils doivent être publiés ou diffusés au grand public par une maison d'édition afin de nous assurer qu'ils sont issus d'un travail professionnel (voir l'Annexe 1 pour davantage d'informations sur les sources de notre corpus).

### 5.1.11 Date de publication

Le dernier critère est celui qui concerne la date de publication des textes. L'intérêt de choisir la date de publication est toujours en fonction des objectifs du projet (Bowker et Pearson 2002 : 52). Dans notre cas, l'objectif est d'étudier l'état actuel de la langue de spécialité, surtout les termes et les concepts en usage à l'heure actuelle, et ceux qui apparaissent suite au développement dans le domaine de la technologie.

Notre corpus comprend des textes qui datent de 2009 jusqu'à 2014. Donc, nous n'avons pas l'intention de faire une étude diachronique de la langue de spécialité. Étant donné que nous travaillons sur les UTP, nous trouvons nécessaire que le corpus soit récent et à jour afin que nous puissions décrire la structure actancielle de ces unités où les actants sont de nouveaux termes. À titre d'exemple, le verbe *Had~ava* (حَدَّثَ) (*mettre à jour*) dans le domaine de la technologie cellulaire et des réseaux sociaux exprime-t-il la même notion que dans le domaine de la technologie informatique ?

En somme, nous estimons avoir composé un corpus reflétant un dynamisme qui se traduit par une richesse terminologique. Nous disposons de données linguistiques et conceptuelles récentes ce qui nous permet d'analyser et d'identifier les aspects de la langue de spécialité qui nous intéressent. Nous rappelons que la langue de spécialité : « is typically dynamic-concepts in specialised subject field are constantly evolving and the terms used to describe these concepts also change » (ibid. : 48).

Dans le tableau qui suit, nous résumons les principales caractéristiques de notre corpus. Nous divisons nos sources en deux groupes, les sources provenant des sites Internet et les sources provenant des manuels.

	<b>Sites Internet</b>	<b>Manuels</b>
<b>Langue</b>	ASM	ASM
<b>Taille</b>	205 687	296 352
<b>Prov. Géo</b>	Moyen-Orient	Moyen-Orient
<b>Date</b>	Entre 2009-2014	2009-2014
<b>Sujet</b>	Application et systèmes d'exploitation mobiles, tablette, ordinateurs, Internet et réseaux	Dépannage informatique, programmation et tutoriels
<b>Type de corpus</b>	Textes écrits	Textes écrits
<b>Textes</b>	articles entiers	documents entiers
<b>Types de textes</b>	Expert à initié/non-initié	Expert à initié
<b>Paternité</b>	Sites authentiques	Manuels publiés par des maisons d'édition

Tableau 4 : Résumé des principales caractéristiques de notre corpus

## 5.2 Extraction de candidats termes

Après avoir constitué le corpus, nous avons procédé à l'extraction des candidats termes. À notre connaissance, il n'existe pas un système adapté à nos besoins, à savoir un système qui se concentre sur les termes simples et qui extrait, plus spécifiquement, les UTV et les UTN. La littérature sur le sujet (Boukandel et al. 2008, Al-khatib et Badarneh 2010, El-Mahdaouy et al. 2013, Hadni et al. 2014 et Abed et al. 2013) montre que les extracteurs conçus pour l'ASM de spécialité ne sont pas disponibles sur Internet et ils ne sont pas conçus pour l'extraction des UTV et UTN (sauf celui de Abed et al. 2013 qui est basé sur une méthode hybride pour l'extraction des unités simples et complexes).

Par conséquent, nous<sup>41</sup> avons décidé d'exploiter un système d'extraction de termes simples existant, mais qui n'avait pas encore été adapté à la langue arabe. Il s'agit de

---

<sup>41</sup> Cette partie de notre travail était collaborative. La programmation était prise en charge par Benoit Robichaud, membre de l'OLST.

l'extracteur de l'Observatoire de Linguistique Sens-Texte (OLST) de l'Université de Montréal. Cet extracteur est conçu et géré par Drouin (2003) et s'appelle TermoStat.

### 5.2.1 TermoStat et la méthode de comparaison

TermoStat est un logiciel de statistique textuelle qui repose sur une méthode d'extraction par comparaison de corpus (Drouin 2004; Lemay et al. 2005). Cette méthode (statistique) consiste à comparer les fréquences des formes lexicales d'un corpus d'analyse (CA) à celles d'un corpus de référence (CR) afin de faire émerger le vocabulaire propre au CA.

Le corpus de référence est un corpus de langue générale comprenant un large ensemble de textes sur plusieurs sujets, tandis que le corpus d'analyse porte sur un domaine de spécialité. L'avantage de cette méthode est qu'elle se présente comme une solution éventuelle à un problème récurrent dans l'extraction des unités terminologiques simples (UTS) ; un bon nombre d'UTS ne se différencient pas au niveau de surface d'autres unités lexicales (Lemay et al. 2005 : 228). La comparaison se fait entre les deux corpus afin de découvrir quels items du CA : « deviate from a « standard » or « theoretical » frequency represented by the RC » (ibid. : 208).

À titre d'exemple, si nous travaillons sur un CA du domaine de l'informatique, un terme comme *télécharger* pourrait retenir notre attention. Ce terme aurait probablement un bon nombre d'occurrences dans le CA. Le système d'extraction le proposerait comme candidat terme s'il ne présente pas un bon nombre d'occurrences dans le CR. En d'autres mots, l'UL, pour être retenue comme terme, doit avoir une certaine spécificité dans le CA par rapport au CR. Nous allons nous attarder sur cette question dans la section 5.2.4.

TermoStat permet d'isoler trois types de spécificités : les spécificités positives, les spécificités négatives et les formes banales. Les spécificités positives représentent les formes qui sont sur employées dans le CA. Les spécificités négatives représentent les formes qui sont sous-employées dans le CA. Quant aux formes banales, ce sont les formes qui ne sont considérées ni négatives ni positives. Le type de spécificités qui nous intéresse est celui de spécificités positives, étant donné que les unités se caractérisant par ce type de spécificité sont censées être représentatives d'un domaine de spécialité. Il convient de noter que les

algorithmes utilisés dans TermoStat font appel à la loi normale et la « valeur-test » qui sont utilisées pour calculer les spécificités positives. En ce qui concerne la loi normale, Drouin (2002 : 142-143) stipule que :

La loi normale nous permet d'identifier à la fois les formes qui sont significativement fréquentes et celles qui sont significativement peu fréquentes dans le corpus d'analyse par rapport au corpus de référence. Nous sommes donc face à deux cas de figure où soit la fréquence observée dans le corpus d'analyse est supérieure ou égale à la valeur théorique, soit la fréquence observée est inférieure à sa valeur théorique.

Quant à la valeur-test, Drouin (ibid. : 147) la décrit comme étant la « quantité permettant d'apprécier la signification de la position d'un élément illustratif sur un axe factoriel. En d'autres termes, il s'agit de la position d'un point sur la courbe normale ». Pour retenir les formes les plus significatives, un seuil de 3,09 est adopté pour la valeur-test. Cela veut dire que ce seuil permet d'assurer qu'il n'y a qu'une chance sur 1000 que la référence observée dans le texte du corpus soit due au hasard (ibid. : 147).

### **5.2.2 Intégration de l'ASM à TermoStat**

TermoStat est conçu pour le français, l'anglais, l'espagnol, l'italien et le portugais. Afin de bâtir notre système d'extraction, nous avons décidé d'y intégrer l'ASM. Pour ce faire, nous avons commencé à réunir les éléments nécessaires : un étiqueteur, un analyseur morphosyntaxique et un corpus général (ou corpus de référence, CR). Nous détaillons dans la section 5.2.4 les différentes étapes de notre travail qui nous ont permis d'intégrer l'ASM à TermoStat et d'effectuer, par la suite, l'extraction des candidats termes.

Avant de commencer, il importe de nous attarder sur nos premières tentatives faites avant d'avoir fait notre dernier choix sur les outils. Ces tentatives nous ont permis de comprendre la particularité de l'ASM, de point de vue du traitement automatique des langues (TAL). De plus, elles nous ont permis de voir les outils de TAL disponibles pour cette langue, et lesquels parmi eux sont plus adaptés à nos besoins.

### 5.2.3 Le système de MADA de Habash

Nos premières tentatives<sup>42</sup> portaient sur le système de MADA (Habash et al. 2009). MADA (Morphological Analysis and Disambiguation for Arabic) est un ensemble d'outils capables d'effectuer la diacritisation, la segmentation, l'analyse morphosyntaxique, l'étiquetage en parties du discours (POS) et la désambiguïisation morphologique. Pour effectuer une analyse morphologique, MADA s'appuie sur l'analyseur de BAMA (Buckwalter Arabic Morphological Analyzer). Son mode de fonctionnement consiste à choisir, pour chaque mot, l'analyse la plus probable à partir d'une liste d'analyses potentielles fournie par BAMA. MADA ne prend pas en charge les caractères arabes. Donc, pour fonctionner, il convertit les textes à analyser en caractères latins, ou ce que l'on appelle *translittération* (voir section 2.3.1). Une sortie de MADA a la forme suivante, figure 69.

```
;;; SENTENCE mn Almmkn tHsyn >dA' $bkAT <vrnt btqsym Alqsm AlmzdHm ELY Al$bkp <LY qsmyn vm rbThmA mEA bAstxdAm jsr @LATE
;;;WORD mn
;;;SVM_PREDICTIONS: mn asp:na cas:na enc0:0 gen:na mod:na num:na per:na pos:prep prc0:na prc1:0 prc2:0 prc3:0 stt:na vox:na
*1.033497 diac:min lex:min_1 bw:+min/PREP+ gloss:from pos:prep prc3:0 prc2:0 prc1:0 prc0:na per:na asp:na vox:na mod:na g
_0.913320 diac:man lex:man_2 bw:+man/INTERROG_PART+ gloss:who/whom pos:part_interrog prc3:0 prc2:0 prc1:0 prc0:na per:na
_0.691333 diac:man lex:man_1 bw:+man/REL_PRON+ gloss:who/whom pos:pron_rel prc3:0 prc2:0 prc1:0 prc0:0 per:na asp:na vox:
-----
;;;WORD Almmkn
;;;SVM_PREDICTIONS: Almmkn asp:na cas:g enc0:0 gen:m mod:na num:s per:na pos:adj prc0:AL_det prc1:0 prc2:0 prc3:0 stt:d vo
*0.964535 diac:Almmokin lex:mumokin_1 bw:AL/DET+mumokin/ADJ+ gloss:possible pos:adj prc3:0 prc2:0 prc1:0 prc0:AL_det pe
-----
;;;WORD tHsyn
;;;SVM_PREDICTIONS: tHsyn asp:na cas:n enc0:0 gen:m mod:na num:s per:na pos:noun prc0:0 prc1:0 prc2:0 prc3:0 stt:c vox:na
*0.907318 diac:taHosiyn lex:taHosiyn_1 bw:+taHosiyn/NOUN+ gloss:improving/making_better pos:noun prc3:0 prc2:0 prc1:0 prc
^0.904070 diac:taHosiyn lex:taHosiyn_2 bw:+taHosiyn/NOUN+ gloss:improvement/beautification pos:noun prc3:0 prc2:0 prc1:0
_0.721090 diac:tuHas~lyna lex:>aHas~_1 bw:tu/IV2FS++Has~/IV_PASS++lyna/IVSUFF_SUBJ:2FS_MOOD:I gloss:be_felt/be_perceived
_0.719417 diac:tuHis~lyna lex:>aHas~_1 bw:tu/IV2FS++His~/IV++lyna/IVSUFF_SUBJ:2FS_MOOD:I gloss:feel/perceive pos:verb prc
_0.719417 diac:taHis~lyna lex:Has~_1 bw:ta/IV2FS++His~/IV++lyna/IVSUFF_SUBJ:2FS_MOOD:I gloss:feel pos:verb prc3:0 prc2:
_0.719417 diac:taHas~lyna lex:Hasis_1 bw:ta/IV2FS++Has~/IV++lyna/IVSUFF_SUBJ:2FS_MOOD:I gloss:feel pos:verb prc3:0 prc2:0
_0.719417 diac:taHosiyna lex:HasA~u_1 bw:ta/IV2FS++Hos~/IV++lyna/IVSUFF_SUBJ:2FS_MOOD:I gloss:drink/sip pos:verb prc3:0 pr
_0.697215 diac:taHas~ayona lex:taHas~aY_1 bw:+taHas~ay/PV++na/PVSUFF_SUBJ:3FP gloss:drink/sip pos:verb prc3:0 prc2:0 prc1
_0.678538 diac:tuHosayona lex:HasA~u_1 bw:tu/IV2FP++Hosay/IV_PASS++na/IVSUFF_SUBJ:FP gloss:be_sipped pos:verb prc3:0 prc2
-----
;;;WORD >dA'
;;;SVM_PREDICTIONS: AdA' asp:na cas:g enc0:0 gen:m mod:na num:s per:na pos:noun prc0:0 prc1:0 prc2:0 prc3:0 stt:c vox:na
*0.927430 diac:>adA' lex:>adA'_1 bw:+adA'/NOUN+ gloss:performance/fulfillment pos:noun prc3:0 prc2:0 prc1:0 prc0:0 per:na
-----
;;;WORD $bkAT
;;;SVM_PREDICTIONS: $bkAT asp:na cas:g enc0:0 gen:f mod:na num:p per:na pos:noun prc0:0 prc1:0 prc2:0 prc3:0 stt:i vox:na
*0.963717 diac:$abakat lex:$abakap_1 bw:+$abak/NOUN++At/NSUFF_FEM_PL gloss:networks/systems pos:noun prc3:0 prc2:0 prc1:0
-----
;;;WORD <vrnt
;;;NO-ANALYSIS
;;;SVM_PREDICTIONS: Avrnt asp:p cas:na enc0:0 gen:m mod:i num:s per:3 pos:verb prc0:0 prc1:0 prc2:0 prc3:0 stt:na vox:a
NO-ANALYSIS [<vrnt]
-----
;;;WORD btqsym
;;;SVM_PREDICTIONS: btqsym asp:na cas:u enc0:0 gen:m mod:na num:s per:na pos:noun prc0:0 prc1:bi_prep prc2:0 prc3:0 stt:i
*1.000064 diac:bitaqosym lex:taqosym_1 bw:bi/PREP++taqosym/NOUN+ gloss:partition/division/distribution pos:noun prc3:0
```

Figure 69 : Partie d'une phrase analysée avec MADA

Comme montré dans la figure 69, une fois le texte converti en forme translittérée, MADA donne toutes les analyses possibles pour chaque item dans la phrase. On peut

<sup>42</sup> Nous avons également essayé d'autres outils comme AntConc et Unitex. Avec ces outils, nous n'étions pas satisfait des résultats, surtout que les versions que nous avons n'étaient pas stables, ce qui nous a causé un certain nombre de soucis techniques.



remarquer que pour la forme *tHsyn* (تحسين) (amélioration) dans la figure, il existe 9 analyses possibles, et l'analyse la plus probable est marquée par un astérisque (\*). En ce qui concerne l'analyse, pour chaque élément, on peut trouver les informations suivantes : la forme diacritisée de l'item, la forme lemmatisée, la racine (la forme sans clitiques ni affixes), la partie du discours, le genre, le nombre et l'équivalent en anglais.

Nous avons remarqué que MADA présente un certain nombre de limites, que nous allons décrire dans ce qui suit.

#### *L'analyseur BAMA*

Étant donné que MADA est basé sur BAMA pour l'analyse morphologique, nous avons constaté que, souvent, MADA ne retient pas la bonne analyse, car il n'est pas entraîné sur un corpus moderne de la langue arabe. Notre champ d'intérêt est la langue de spécialité, et BAMA comporte des unités lexicales obsolètes qui ne sont plus en usage en ASM<sup>43</sup>. À titre d'exemple, dans les résultats de l'extraction, nous avons trouvé des mots dont les significations sont ambiguës pour nous, comme *fasA*, *HAf*, *bakar*, *hakir* et *tAbal*<sup>44</sup>. De plus, un verbe comme *naqara* (نقر) (cliquer) n'est jamais retenu dans la bonne forme. Il est toujours présenté comme *naqira* (نقر) (nerveux).

#### *Translittération*

La translittération constitue un problème pour nous, à cause de la difficulté de lecture qu'elle représente. Nous comprenons que la translittération facilite la diacritisation des formes, mais cela constitue toujours un handicap surtout que notre objectif était d'intégrer MADA dans TermoStat. Dans ce cas, l'utilisateur devait reconverter la sortie en caractères arabes, ce qui pourrait relever d'un processus compliqué.

---

<sup>43</sup> Habash et al. (2009) recommandent l'usage de SAMA (Standard Arabic Morphological Analyzer) avec MADA, qui est une source lexicale plus moderne que BAMA. Nous n'avons pas pu l'obtenir, compte tenu du coût du logiciel.

<sup>44</sup> Ces mots relèvent de la langue arabe classique; nous n'avons pas pu vérifier leur signification exacte, puisqu'ils sont très polysémiques.

### *Plusieurs analyses*

Comme MADA propose plusieurs analyses probables pour un item, l'inconvénient lors de l'extraction est quand l'analyse la plus probable ne se trouve pas en première position, c'est-à-dire qu'elle n'est pas celle marquée par un astérisque. Donc, cette analyse ne sera pas celle retenue par l'extracteur.

## **5.2.4 Étapes pour bâtir un système d'extraction**

### **5.2.4.1 Corpus général**

Après des recherches sur la Toile, nous sommes parvenu à trouver un corpus général de presque 113 millions de mots. Ce corpus se trouve dans le site Internet de AraCorpus<sup>45</sup>. Il est constitué d'articles de presse extraits de journaux en ASM, parus en 2004 et 2005. Il comprend 28 fichiers, chacun de ces fichiers comprend un ensemble d'articles extraits d'un journal. Il importait pour nous de trouver un corpus de langue générale en ASM. Cela signifie que les deux corpus, général et spécialité, sont homogènes, puisque tous deux sont écrits dans la même variante.

Donc, ce corpus est de nature journalistique, les articles portent sur plusieurs sujets, principalement la politique, les sports, la météo, la nature, la société et la culture. En ce qui concerne nos besoins, nous n'avons pas pris tous les fichiers. Nous n'en avons pris que cinq, ce qui fait un total de 16 millions de mots. La décision de prendre juste une partie est justifiée par le fait que nous avons estimé qu'un corpus général de 16 millions de mots est suffisant pour que l'extraction fonctionne (Drouin 2004). Dans le tableau 5, nous donnons un aperçu des sources de ce corpus.

---

<sup>45</sup> <http://aracorpus.e3rab.com>

SOURCE	NOMBRE DE MOTS 1	NOMBRE DE MOTS 2	NOMBRE DE LIGNES	PROVENANCE
ARABIC.CNN	5 437 167	4 024 149	423 105	États-Unis
ARABIC.BBC	2 926 970	3 086 838	248 246	Grande-Bretagne
AHRAM	15 281 728	4 676 306	725 934	Égypte
THAWRA	4 433 879	4 557 379	247 054	Syrie
SAHARA.MEDIA	375 053	393 608	24 024	Mauritanie

Tableau 5 : Détails du corpus général

Dans le tableau, nous avons ajouté deux colonnes qui spécifient le nombre de mots. Dans la première, nous donnons le nombre de mots pris tel quel du fichier *readme* fourni avec le corpus téléchargé. Dans la deuxième, nous donnons notre propre décompte, que nous avons effectué à l'aide du concordancier *aConCorde* (Roberts et al. 2005). Nous ne pouvons pas juger lequel des deux décomptes est le plus exact, mais nous estimons que la taille de corpus est d'environ 16 millions de mots. En ce qui concerne le nombre de textes dans le corpus, l'auteur ne fait pas état de cette question, mais il précise le nombre de lignes pour chaque corpus.

#### 5.2.4.2 Étiquetage des corpus

Afin que l'extraction fonctionne, il fallait étiqueter les deux corpus, le corpus général et le corpus spécialisé. Nous avons effectué l'étiquetage à l'aide de l'étiqueteur *Stanford POS tagger* (Toutanova et al. 2003) de l'Université Stanford. Cet étiqueteur est compatible avec l'ASM, et il existe en plusieurs versions. Pour notre travail, nous avons utilisé la version la plus récente, 3.4.1.

Avant de procéder à l'étiquetage, nous avons converti les textes en format *.txt*, ou texte brut. En ce qui concerne le corpus général, il n'était pas difficile de convertir les fichiers comprenant les textes. Ils étaient en format HTML et il nous a suffi de faire la conversion en utilisant une simple manipulation<sup>46</sup>. Par contre, la conversion des fichiers de notre corpus de spécialité n'était pas une tâche facile à réaliser.

---

<sup>46</sup> Les fichiers étaient au format *html*. Après les avoir ouverts avec le navigateur Google Chrome, nous avons choisi l'option « enregistrer le fichier sous » pour les convertir en *txt*.

Nous avons éprouvé certaines difficultés avec les documents, comme les manuels, en format *PDF*. Le genre de difficulté rencontré concerne la disparition des caractères arabes lors de la conversion, un changement dans l'ordre de mots, une grande répétition des lettres pour certains mots, et si le document comportait des termes en anglais, nous avons remarqué un chaos dans le texte entier. Nous avons essayé plusieurs méthodes pour trouver une solution à ce problème, mais toujours en vain. Nous montrons quelques exemples des fichiers mal convertis dans les figures 70 et 71.

Finalement, nous avons décidé d'écarter certains documents que nous n'avons pas pu convertir et de garder ceux dont la conversion était réussie à 75 %<sup>47</sup>. Pour ces documents, nous avons procédé à quelques étapes de remise en forme et de nettoyage qui consistaient à éliminer certains éléments dans les textes comme les numéros de pages, les en-têtes des titres, les pieds de pages, etc. À l'exception de ces étapes de prétraitement, notre corpus n'a subi aucune modification une fois converti en format *.txt*.

```

@@ 6+.pa@yil"n0a@bàãácé@ @té+áié 12
- PCs
Apple
10.7 OSX Lion
Windows 7 Macintosh Apple
h
@@@
Operating Systems @ @@ @ Linux 9x N N N N AT & T C Novell 1993
1995 Ø N N
7
N x
N N
-
N - N
ë .
Apple N N N N
N ĞN N N

```

Figure 70 : Texte mal converti, problème d'encodage

---

<sup>47</sup> Les documents écartés couvrent des sujets déjà compris dans le corpus, tel que le dépannage informatique, la programmation, etc. Le fait d'avoir écarté ces documents a eu un impact sur le nombre de mots de notre corpus. De plus, pour les documents restants, nous avons trouvé important et nécessaire de procéder à une vérification manuelle pour nous assurer que les unités analysées (suite à l'extraction et l'analyse morphosyntaxique) ne sont pas déformées suite à la conversion.

إذا قررا أنك ال ترمذ الاحتفاظ بجهاز الكمبيوتر الآخر أو استعماله، من المفضل إلغاء تصحضه بتشغيل املحسبى، مقاطع الصيديمج، أو الكتب الفحتية التي قمت بشرائها من iTunes Store. إن من شأن إلغاء تصحض الكمبيوتر أن محلل دون تشغيل الأغاني، مقاطع الصيديمج، أو الكتب الفحتية من قبل شخص آخر ومحذر تصحض استعمال آخر. للزمذ من املعلماء حل إلغاء التصحض، قم باختيار iTunes Help من قائمة املعلماء في iTunes Store.

## الصقل اجاهز، إعداء، اطلق II

قوخلال تخصيصا سطح املكتبا قواصا يكاوتفضيا لك. ميكنك أن تدع سطح املكتب مبدو، بسرعة، كما ترمذ باستخدام تصضيا لا النظام، وهي مركز الأوامر اخلص بك لمعظم الإعداء على iMac اخلص بك. قم باختيار Apple ( > تصضيا لا النظام من شريط القائمة، أو انقر على أمقنة تصضيا لا النظام في Dock.

ماهه ميكنك إعداء خيار العادة تعيني كلمة السر اخلصا بك في تصضيا لا املستخدمني واجلمعما. في حال نسيت كلمة سر دخلك للزمذ من املساعدة حل تصضيا لا النظام، قم بصتح مركز املساعدة وابحث عن "تصضيا لا النظام" أو عن التصضيل اوند الذي ترغب في تغييره.

## إسبا iMac اخلص بك أو أمقاف تشغيله

عندما تنتهي من العمل على iMac، ميكنك أن تدخه إلى مرحلة الإسبا أو أن تحقف تشغيله.

Figure 71 : Texte mal converti, problème de lisibilité

L'étiqueteur de Stanford fonctionne par méthode statistique. En ce qui concerne l'ASM, il est entraîné sur le corpus de *Penn Treebank* de cette langue (Maamouri et al. 2004). Ce corpus est compilé à partir des textes provenant de journaux en ASM. De plus, l'étiqueteur se distingue par sa capacité d'attribuer des étiquettes même à des mots ambigus et de donner des détails sur les verbes. Nous montrons dans la figure 72 une sortie d'un texte étiqueté.

الحاسب/ NNP هو/ PRP آلة NN تعامل/ VBP مع/ NN البيانات/ DTNNS وفقا/ NN لمجموعة/ JJ من/ IN التعليمات/ DTNNS - البرامج تتعدد/ DTJJ أنواع/ NN الحواسيب/ DTNN من/ IN حيث/ WRB طريقة/ NN عملها/ NN وحجمها. NN/ أوائل/ NN الحواسيب/ DTNN الإلكترونية/ DTJJ كانت/ VBD في/ IN حجم/ NN غرفة/ NN كبيرة/ JJ وتستهلك/ NN طاقة/ NN مماثلة/ JJ لا/ IN يستهلكه/ VBP بضعة/ NOUN مئات/ NNS من/ IN الحواسيب/ DTNN الشخصية/ DTJJ اليوم/ VBD/ (1) DTJJ من/ IN الممكن/ DTNN اليوم/ DTNN صنع/ IN حواسيب/ NN داخل/ NN ساعة/ NN يد/ NN تأخذ/ NN طاقتها/ JJ من/ IN بطارية/ NN الساعة/ DTNN ينظر/ VBP المجتمع/ DTNN إلى/ VBD الحاسب/ DTNN الشخصي/ DTJJ - ونظيره/ NNP المتفعل/ DTNN الحاسب/ DTNN المحمول/ DTJJ على/ IN أنهما/ NN رمزي/ NNP عصر/ NN المعلومات/ DTNN فهما/ NN ما/ WP يفكر/ VBP به/ NN معظم/ NOUN الناس/ DTNN عند/ NN الحديث/ DTNN عن/ IN الحاسب/ DTNN على/ IN الرغبة/ DTNN من/ IN هذا/ DT فاكتر/ NN أشكال/ NN الحاسب/ DTNN استخدم/ NN يوما/ DTNN هي/ PRP الحواسيب/ DTNN المضمّنة/ DTJJ الحواسيب/ DTNN المضمّنة/ DTJJ حي/ NN أجهزة/ NN صغيرة/ JJ وبسيطة/ JJ تستخدم/ VBN عادة/ NN للتحكم/ JJ في/ IN أجهزة/ NN أخرى/ NN فعلى/ NN سبيل/ NN المثال/ DTNN يمكنك/ VBP أن/ NN تجدها/ NN في/ IN آلات/ NNS تتراوح/ VBP من/ IN الطائرات/ DTNNS المقابلة/ DTJJ و/ CC والأيون/ DTNN والآت/ NNS التصوير/ DTNN الرقمية/ DTJJ ولعب/ NN الأطفال/ DTNN حاسوب/ NN في/ IN ساعة/ NN يد. لا/ NN يمكن/ VBP القول/ DTNN بأن/ NN الحاسوب/ DTNN هو/ PRP اختراع/ NN بحد/ NN ذاته/ NN له/ NNP كان/ VBD نتاج/ NN الكثير/ DTNN من/ IN الابتكارات/ DTNNS العلمية/ DTJJ و/ CC التطبيقات/ DTNNS الرياضية/ DTJJ الحاسبات/ DTJJ متنوعة/ VBN في/ IN الواقع/ DTNN إنها/ NN الآت/ NNS معالجة/ NN بيانات/ NNS عالمية/ PUNC. JJ طبقا/ NN لفرض/ NN Church-Turing NN فإن/ NN حاسب/ NN له/ NN قدرة/ NN ذات/ NN حد/ NN أدنى/ JJR معين/ NN يكون/ VBP ببساطة/ NN قادر/ JJ على/ IN إنجاز/ NN المهام/ DTNN الخاصة/ DTJJ بأي/ NNP حاسب/ NNP آخر/ VBD يد/ NN من/ IN المساعدة/ DTNN الرقمي/ DTJJ الشخصي/ DTJJ إلى/ VBD الحاسب/ DTNN الفائق/ DTJJ طابا/ IN أن/ NN الوقت/ DTNN وسعة/ NN الذاكرة/ DTNN ليست/ VBD في/ IN الاعتبار. DTNN لذلك/ NN فإن/ NN التصميمات/ DTNNS المتماثلة/ DTJJ من/ IN الحاسب/ DTNN من/ IN الممكن/ DTNN أن/ NN تضبط/

Figure 72 : Sortie d'un texte étiqueté avec Stanford

L'étiqueteur Stanford utilise les étiquettes de *Penn Treebank*. Chacun des mots reçoit une étiquette désignant sa partie du discours. Pour les noms, Stanford indique si le nom est accompagné d'un article défini ou d'une préposition et s'il s'agit d'un nom propre au singulier ou au pluriel. En ce qui concerne les verbes, on leur assigne des étiquettes distinctes indiquant le lemme, le temps, la personne et le nombre. Nous détaillons dans le tableau 6 les composantes de ces étiquettes avec des exemples.

<b>VB (base form)</b>	<b>taEAmala (تعامل) (traité)</b>
<b>VBD (past tense)</b>	taEAmala (تعامل) (traité)
<b>VBG (gerund present)</b>	yataEAmal (يتعامل) (traitant)
<b>VBN (past participle)</b>	taEAmala (تعامل) (traité)
<b>VBP (non-3rd ps. Sing. Present)</b>	tataEAmal (تتعامل) (tu traites)
<b>VBZ (3rd. ps. Sing. Present)</b>	taEAmala (تعامل) (traite)

Tableau 6 : Étiquette VB de Stanford et ses variations

### 5.2.4.3 Analyse morphosyntaxique

L'étiqueteur Stanford n'est pas équipé pour donner la forme lemmatisée ou l'analyse morphologique des formes lexicales. Pour cette raison, nous avons à trouver un autre outil pour effectuer cette tâche. Notre choix a porté sur une trousse à outils de traitement morphologique appelée *AraComLex* (Attia et al. 2011). *AraComLex* est un transducteur<sup>48</sup> à états finis, basé sur un corpus modèle d'ASM. Afin d'élargir la portée de son corpus, il se sert d'une ressource lexicale existante, SAMA (*Standard Arabic Morphological Analyser*) (Maamouri et al. 2010) et il procède, par la suite, à une étape de filtrage qui consiste à consulter le web par une stratégie de requêtes et de statistiques afin de déterminer quelles seront les informations lexicales à retenir ou à exclure de SAMA. Ces informations doivent refléter l'ASM en usage.

L'avantage que possède *AraComLex* est qu'il représente une ressource lexicale moderne pour l'ASM, contrairement au BAMA qui contient un certain nombre de mots obsolètes qui ne sont plus en usage (Attia et al. 2011 : 103). Dans la section 5.2.4.4, nous faisons une comparaison entre les deux outils testés pour notre recherche (MADA et *AraComLex*) et nous justifions notre choix en ce qui concerne l'analyseur morphosyntaxique retenu.

---

<sup>48</sup> Un transducteur morphologique à états finis, ou *finite state technology transducer*, est un automate qui a pour tâche de traiter une entrée (lemmes ou racines suivis par des étiquettes correspondant à des morphèmes) pour produire une sortie (séquence d'étiquettes remplacée par des morphèmes).

Nous montrons dans la figure 73 une sortie d'AraComLex appliquée à notre corpus. Pour mieux présenter les résultats de l'analyse, nous avons importé la sortie dans un fichier Excel.

	A	B	C	D	E	F	G	H	I	J	K	L
1	NN/	كروم/	عن/	طريق/	كروم/	NN/	كروم/	NN/	كروم/	NN/	كروم/	NN/
2												
3	Forme	PdD	Lemme	Analyse								
4	ثبتت	VBD	ثبتت	+verb+imp2+ثبتتpers+masc+sg@								
5	إضافة	NN	إضافة	+noun+inhumanإضافة+fem+sg@								
6	Secure	DTNN	??	+?								
7	Profile	DTJJ	??	+?								
8	من	IN	من	+prepمن@								
9	متجر	NN	متجر	+noun+inhumanمتجر+masc+sg@								
10	كروم	NN	كروم	@autogen+pl كروم								
11	أو	NNP	أو	+con @								
12	اضغط	VBD	اضغط	+verb+imp2+اضغطpers+masc+sg@								
13	على	IN	على	+prepعلى@								
14	اسم	NN	اسم	+noun+inhumanاسم+masc+sg@								
15	الإضافة	DTNN	إضافة	@fem+sg+إضافة defArt@+noun+inhuman+								
16	من	RB	??	+?								
17	إن	NN	إن	+part:إن@								
18	كنت	VBD	كنت	@masc+ت prep@+noun+inhuman+ك								
19	تتصرف	VBP	تتصرف	+verb+pres+active+3pers+fem(v@taSaf^aH_1@)								
20	عن	IN	عن	+prepعن@								
21	طريق	NN	طريق	+noun+inhumanطريق+masc+sg@								
22	كروم	NN	كروم	@autogen+pl كروم								
23												

Figure 73 : Phrase analysée avec AraComLex

Afin d'analyser les textes du corpus, AraComLex découpe les textes en phrases. Pour chaque phrase, le système place chaque forme sur une ligne avec les analyses correspondantes. Dans la figure 73, les quatre colonnes présentent les informations suivantes :

- Dans la première colonne (de gauche), on présente la forme telle qu'elle apparaît dans la phrase analysée;
- Dans la deuxième, les parties du discours sont attribuées aux formes. Étant donné qu'AraComLex ne fait pas l'étiquetage, nous avons intégré Stanford pour effectuer cette tâche;
- Dans la troisième colonne, le lemme est fourni pour chaque forme;
- Dans la dernière colonne, on trouve l'analyse morphosyntaxique qui comporte les informations suivantes. Par exemple, dans le cas de la première forme, il s'agit d'un verbe, *vab~it* (ثبتت) (installés), donc on a le schéma : **+verb+imp2+ثبتتpers+masc+sg@**, où le premier élément correspond à la partie du discours (verbe), le deuxième élément correspond

au type de verbe (impersonal verb), le troisième élément correspond au lemme (*vab~ata* (ثبت) (installer)), le quatrième élément correspond à la personne (deuxième personne), le cinquième élément correspond au genre (masculin), et le dernier élément correspond au nombre (singulier). Lorsqu'il s'agit d'un nom, le cas du deuxième élément dans la figure, le schéma se présente comme suit : **+noun+inhumanإضافة+fem+sg@**. Dans ce schéma, on trouve la partie du discours (nom), le trait (non-humain), le lemme (*iDAfap* (إضافة) (extension)), le genre (féminin) et le nombre (singulier).

Dans la figure 73, on observe quelques formes non analysées par AraComLex (les points d'interrogation). Ces formes correspondent à deux choses : soit une forme que le système n'a pas reconnue, car elle n'existe pas dans son corpus, soit une forme étrangère, comme un mot en anglais (dans la figure *secure profile*).

Cependant, nous avons remarqué que Stanford, qui s'occupe de l'étiquetage, attribue, malgré leur ambiguïté, des étiquettes à ces formes étrangères, étant donné qu'il fonctionne par méthode statistique. On peut également observer qu'il existe une forme dont le lemme est donné en translittération Buckwalter, *taSaf~aH\_1* pour la forme *tataSaf~aH* (تتصفح) (tu navigues). Ce genre de lemmatisation s'explique par le fait qu'AraComLex comprend une liste des lemmes récemment ajoutée par le concepteur du logiciel. Ces lemmes sont donnés en translittération Buckwalter pour une raison que nous ignorons. Finalement, la dernière forme analysée dans la figure est un nom propre, *kruwm* (كروم) (le navigateur Chrome de Google). AraComLex ne l'a pas retenu comme un nom propre, mais comme un nom commun, ce qui signifie en arabe « des vignes ». Cependant, AraComLex est capable de bien analyser certains noms propres de l'informatique comme *Windows*.

#### 5.2.4.4 AraComLex versus MADA

Notre choix d'analyseur morphosyntaxique repose sur les résultats d'un test que nous avons fait opposant AraComLex à MADA. Nous avons extrait un échantillon de presque 900 mots de notre corpus. Dans cet échantillon, nous avons sélectionné environ 24 formes



susceptibles d'être relatives au domaine de l'informatique. L'objectif était d'évaluer le degré de précision en ce qui concerne l'étiquetage et l'analyse morphosyntaxique des deux outils. Ces formes sont de nature verbale et nominale. À titre d'exemple, nous avons analysé des verbes comme *Ham~ala* (حمل) (télécharger) et *rafaEa* (رفع) (téléverser), et leur nominalisation, et des noms comme *qA{imap* (قائمة) (menu) et *mustaxdim* (مستخدم) (utilisateur).

Ces formes étaient analysées selon les critères suivants : sur les 216 formes choisies pour le test à partir de l'échantillon, le nombre des formes correctement analysées par MADA et AraComLex; le nombre des formes correctement analysées par MADA, mais pas retenues en premier (marquées par astérisque). Les chiffres obtenus montrent que :

Pour MADA :

- sélection par 1<sup>re</sup> analyse correcte retenue: 162/216 (75%)<sup>49</sup>
- sélection par analyse correcte comprises dans les analyses proposées:  
194/216 (89%)

Pour AraComLex :

- sélection par analyse: 163/216 (75,46%)
- cas spécifiques (erreurs d'étiquetage): 31/216 (14,62%)

Enfin, nous résumons nos remarques sur la performance des deux systèmes dans le tableau 7.

---

<sup>49</sup> Nous rappelons que pour MADA, la sortie d'une analyse morphologique d'une forme consiste à proposer une liste de toutes les analyses possibles. L'analyse la plus probable est placée en premier et marquée par un astérisque. Pour cette raison, nous avons constaté que, parfois, l'analyse correcte n'est pas celle marquée par un astérisque, mais l'une des analyses proposées.

<b>AraComLex</b>	<b>MADA</b>
Les formes diacritisées dans le corpus ne sont pas reconnues par AraComLex	Le taux d'analyses correctes (pas celles retenues en premier) est plus élevé avec MADA
AraComLex donne tout simplement la partie du discours et le lemme sans proposer trop d'analyses	MADA propose trop d'analyses pour certains items, ce qui est parfois excessif et augmente l'ambiguïté
AraComLex est limité, en termes d'analyses (il propose une seule analyse)	L'avantage de MADA est qu'il est plus précis qu'AraComLex en ce qui concerne les choix d'analyse, étant donné qu'il diacritise les items, donc il y a plus de chances que l'on trouve la forme correcte parmi les propositions même si cette forme n'est pas celle retenue
Étant donné qu'AraComLex est jumelé avec l'étiqueteur de Stanford, il existe un certain conflit entre l'analyse morphologique d'AraComLex et l'annotation de Stanford, surtout par rapport à l'alternance verbe/forme nominale du verbe	Il n'existe pas de confusion avec MADA, étant donné qu'il fait la diacritisation, la segmentation, l'analyse morphosyntaxique, l'étiquetage de parties du discours (POS) et la désambiguïsation morphologique

Tableau 7 : Comparaison entre MADA et AraComLex

Par conséquent, nous avons, en fin, opté pour AraComlex pour les raisons suivantes :

- il se limite à donner le POS avec le lemme;
- il ne fait pas la translittération pour faire l'analyse : il produit des sorties en caractères arabes;
- il serait probablement possible de l'améliorer afin de résoudre le conflit avec Stanford.

La combinaison des deux systèmes présentés dans les sections 5.2.4.2 et 5.2.4.3, AraComLex et Stanford, forme ce que nous avons appelé MAMA (*Modern Arabic Morphological Analyser*). Après avoir traité nos deux corpus avec MAMA, nous avons intégré le tout dans TermoStat afin de générer une liste de candidats termes.

#### 5.2.4.5 Gestion de spécificité

Comme mentionné à la section 5.2.1, le seuil de spécificité par défaut dans TermoStat est de 3,09. Pour notre recherche, le seuil de spécificité utilisé était de 1,96<sup>50</sup>. Un tel seuil nous

---

<sup>50</sup> Drouin (2002 : 148) constate qu'un seuil de 1,96 « correspond à une probabilité de 5 % ». D'ailleurs, l'auteur soutient que ce seuil a été utilisé par certains spécialistes dans l'identification de leur vocabulaire.

a permis d'élargir nos choix de candidats termes, surtout que nous cherchons des UTP dans le corpus. De plus, nous nous intéressons aux UTP à spécificité positive, c'est-à-dire des unités avec une spécificité de 1,96 et plus. Notre hypothèse est que ces unités sont susceptibles d'être spécifiques au domaine de l'informatique.

#### 5.2.4.6 Résultats de l'extraction avec TermoStat

Après avoir effectué l'extraction, nous avons obtenu deux listes, une pour les UTN et l'autre pour les UTV (figures 74 et 75) :

Fréq_réf	Fréq-spéc	Spécificité	Lemmes	Équivalent	POS	Formes							
2984	3557	233,561221	نظام	systeme	N	لنظام	بنظام	الأنظمة	بالأنظمة	لنظامك	بالنظام	لأنظمة	
1947	2342	189,960692	مع	avec	N	مع	معى	معها	معهم	معك	معه		
9007	3246	145,677526	أو	ou	N	أو							
3411	2053	143,450261	تشغيل	fonctionnement	N	بتشغيل	وتشغيله	تشغيل	بالتشغيل	والتشغيل	بتشغيلها	التشغيل	
818	1100	133,836875	شكل	forme	N	وأشكال	وشكله	وبالشكل	بالشكل	بشكل	شكلًا	شكلت	
4013	2029	133,358544	استخدام	utilisation	N	استخدامك	استخدام	استخدام	الاستخدام	استخدامات	استخدامهما	باستخدامهم	
9521	2957	129,41738	ملف	dossier	N	بملفاته	ملفين	بالملفات	للملف	ملفًا	كملف	ملف	
1843	1241	116,132189	أكل	manger	N	كلاهما	كل	كلها	كله	كلا			
4172	1766	115,524668	مستخدم	utilisateur	N	مستخدمين	والمستخدم	مستخدمون	لمستخدمين	مستخدمين	مستخدميه	مستخدم	
454	687	108,700829	خدمة	service	N	بخدمات	خدمات	وخدمات	للخدمات	وخدماتها	خدماتها	الخدمات	
8718	2368	108,110718	جهاز	appareil	N	الجهاز	جهازان	جهازًا	والأجهزة	الجهازين	للأجهزة	بجهاز	
504	664	103,5045	نسخة	copie	N	ونسخ	والنسخة	بنسخته	بنسخ	بنسخة	للتسخ	للتسخة	
492	626	99,6545144	أداة	outil	N	الأداتين	الأدوات	أداتان	أداة	بأدوات	للأدوات	كأدوات	
970	792	98,9452444	حالة	état	N	الحالة	حالات	الحالتين	حالتها	حالة	للحالة	الحالات	
653	683	98,8922829	عدد	nombre	N	والأعداد	الأعداد	والعدد	لعدد	عدهم	أعد	عدد	
1529	928	96,7659906	عند	chez	N	عند	عنده	عندي	عندها	عندك			
1748	957	94,5202246	عد	énumérer	N	وعدا	أعد	أعدت	بعد	بعدها	العد	أعدتها	
1845	942	91,242645	آخر	dernier	N	والآخرى	وآخر	الآخران	والآخرين	كآخر	وآخرهما	آخرين	
1760	829	82,8420111	صورة	image	N	صوّر	كصورة	بصور	كالصورة	الصورة	لصورة	بصورة	
1	209	78,8948397			N	612500	90	798	818	0	810	812	
3536	1070	76,8508948	Oay*_1	quel	N	بأي	أيهما	أي	لأي	أيها	وأي	وأي	
1083	577	72,6461548	وصول	accès	N	وللوصول	والوصول	للوصول	الوصول	لوصول			
962	536	71,2462811	رسالة	message	N	كرسائل	الرسائل	لِلرسائل	والرسائل	ورسالة	والرسالة	برسالة	
455	381	69,2227673	نص	texte	N	نصا	كنص	ونصوصا	نصوصها	ونصوص	للتص	والنص	
222	289	68,156709	وظيفة	fonction	N	بالوظائف	وظائف	وظيفتها	وظائفهم	الوظيفة	للوظائف	وظيفة	
550	391	66,4763019	lum*_ap_1	mot	N	كلمة	وكلمة	كلمتي	كلمته				
540	387	66,3176305	حصول	obtention	N	الحصول	حصولك	والحصول	حصول	فالحصول	للحصول	بالحصول	

Figure 74 : Résultats de l'extraction des UTN

freq_ref	freq_spec	specificite	lemme	Équivalent	pos	formes					
6458	5012	244,812676	Oan~i_1	X	V	لأنه	فأنها	لأنها	وأنه	لأن	لأنهم
316	1340	178,412735	تم	achever	V	لنتم	سنتم	سينم	فتنم	تم	أنتم
5377	2946	165,847954	ممكن	permettre	V	يمكنهم	تمكنك	تمكنه	يمكنني	لمكنها	يمكنها
1095	1411	150,015444	yatum-u_1	achever	V	يتمكن	ينم	لينم	ليتمكن		
8586	3219	147,745305	إلى	à	V	إليها	إليك	إلى	إليه		
9191	3127	139,144891	كان	être	V	يكونون	وسنكون	سنكون	نكون	فكونوا	لكن
2014	1245	112,822555	أستخدم	utilisé	V	لاستخدم	بستخدمان	سنستخدم	فأستخدمه	بستخدمه	أستخدم
800	833	109,062826	تكون	former	V	تكون	سنتكون	سيتكون	وتكون	فتتكون	تتكون
9763	2261	96,7918452	فام	faire	V	وقفمت	سنقوم	وأقوم	وفامت	وسنقوم	يقومان
2197	1024	91,6596936	جاب	X	V	أجاب	يجب	ويجب	فيجب		
5259	1480	87,076487	عمل	tourner	V	لتعمل	لعملها	عملها	سنعمل	عملك	بعمله
3117	900	68,8240039	غير	X	V	لتغير	وتغيره	لتغير	يتغير	سنتغير	وتغيرها
2205	629	57,1917687	سمح	permettre	V	نسمح	سنسمح	سمحت	اسمح	فتسمح	يسمحون
9962	1398	55,0427221	أو	ou	V	أو					
1581	493	52,9853327	وضع	X	V	بالأوضاع	ونضع	بوضعها	بأوضاع	ووضع	وضعه
150	161	48,4762044	واري	X	V	لتوز	توز	توازي	وتوز	فيوز	
9492	1131	43,585978	wahiy-a_1	X	V	أنه	وهي	فيه	نه		
347	192	42,6045135	qud~_1	que	V	فلقد	ولقد	وقد	فقد	فقدها	
875	292	42,1885997	شكل	forme	V	وأشكال	وشكله	وبالشكل	بالشكل	بشكل	شكلًا
614	236	40,5385074	حد	frontière	V	لحد	يحدود	الحد	أحدًا	كحد	الحدود
1	52	39,3453001	تشغيل	fonctionnement	V	بتشغيل	وتشغيله	تشغيل	بالتشغيل	والتشغيل	وتشغيل
1839	378	36,9452699	دعم	supporter	V	تدعمه	بدعم	لدعم	بدعم	ودعم	دعم
4180	576	34,8728788	عرض	afficher	V	وتعرض	نعرض	الأعراض	فسيعرض	وسنعرض	العرض
661	199	33,1317012	حوي	contenir	V	وسيحوي	وتحوي	تحوي	تحويها	يحو	سيحوي
4122	536	32,1894861	mA_7	X	V	يم	موا	فيما	أمان	أماكن	أماكن
2047	349	31,5508592	آخر	autre	V	والأخرى	وآخر	الأخران	والآخرين	كآخر	وآخرهما

Figure 75 : Résultats de l'extraction des UTV

Les listes des résultats prennent la forme d'un tableau à sept colonnes<sup>51</sup> :

- dans la première, TermoStat indique la fréquence du lemme dans le corpus général;
- dans la deuxième colonne, on donne la fréquence du lemme dans le corpus de spécialité;
- dans la troisième colonne, on présente la spécificité;
- dans la quatrième colonne, on donne le lemme;
- dans la cinquième colonne, on trouve l'équivalent
- dans la sixième colonne, la partie du discours est attribuée;
- dans la septième colonne, on trouve les formes du lemme dans le corpus.

<sup>51</sup> La colonne « Équivalent » a été intégrée par nous, et elle ne fait pas partie des sorties de TermoStat.

Pour mieux gérer les sorties de TermoStat, nous avons trié les résultats de l'extraction par ordre décroissant de spécificité, c'est-à-dire que les candidats avec la spécificité la plus élevée occupent la tête de la liste.

#### 5.2.4.7 Traitement des résultats : filtrage

La première étape dans le traitement des résultats consiste à éliminer certaines unités que nous jugeons être non pertinentes au domaine de l'informatique. Ces unités comprennent les erreurs de l'étiquetage, les chiffres, les mots grammaticaux, etc. De plus, nous éliminons les unités qui relèvent de la langue générale. Dans le tableau 8, nous donnons quelques exemples et dans les annexes 2 et 3, nous donnons deux listes partielles des unités candidates.

Type d'unité éliminée	Exemples
Erreurs d'étiquetage	maE (مع) (avec)
Mot grammatical	ha*A (هذا) (ce)
Unité de la langue générale	kalimap (كلمة) (mot)
Chiffres	818

Tableau 8 : Unités éliminées

#### 5.2.4.8 Résultats de filtrage pour les UTN

Après avoir filtré les résultats, nous avons obtenu une liste de 215 unités (sur 553 unités extraites). Ces unités comprennent les UTN et les unités nominales non prédicatives (noms d'entité). À partir de cette liste, nous avons extrait 62 UTN. Dans la liste de départ, nous avons détecté un certain nombre d'erreurs d'étiquetage. Nous avons remarqué que dans la liste, des syntagmes prépositionnels sont souvent proposés comme des candidats. À titre d'exemple, *kalHawAsyb* (كالحواسيب) (comme les ordinateurs), *lil>zrAr* (للأزرار) (pour les boutons) et *bilnaqr* (بالنقر) (en cliquant).

#### 5.2.4.9 Résultats pour les UTV

En ce qui concerne les UTV, nous avons obtenu une liste de 375 lemmes, dont nous avons gardé seulement 18 UTV. Les résultats sont surprenants, surtout que nous avons observé une grande quantité d'erreurs d'étiquetage dans la liste. Pour cette raison, nous avons refait l'extraction, mais cette fois-ci sans seuil de spécificité. Cela veut dire que nous avons

voulu vérifier si les autres unités, dont la spécificité est banale ou négative, sont susceptibles d’avoir un contenu spécialisé. Dans le tableau 9, nous montrons les résultats de cette nouvelle extraction.

Total	S+ 1,96	S=	S-
1194	18	5	35

Tableau 9 : Extraction des UTV sans seuil de spécificité

Avec la nouvelle extraction, nous avons obtenu un total de 58 UTV candidats. Comme montré dans le tableau 9, seulement 18 unités sont à spécificité positive. Quant aux autres, elles sont à spécificité négative, avec 5 unités à spécificité banale.

#### 5.2.4.10 Analyse des résultats

*Pour les UTN*

La plupart des unités extraites pour les UTN sont des noms d’entité. Par exemple, nous trouvons dans la liste des termes comme *niZAm* (نظام) (système), *malaf* (ملف) (fichier), *mustaxdim* (مستخدم) (utilisateur) et *jihAz* (جهاز) (appareil). Quant aux UTN prédicatives, nous avons remarqué que les unités les plus significatives, occupant la tête de la liste, sont des unités comme *ta\$gyl* (تشغيل) (fonctionnement), *AistixdAm* (استخدام) (utilisation), *wSwl* (وصول) (accès), *tanfy\** (تنفيذ) (exécution) et *tavbyt* (تثبيت) (installation).

Cependant, le nombre total extrait reste un point à considérer. Nous avons remarqué l’absence de certaines UTN comme *tazAmun* (تزامن) (synchronisation) et *taEryf* (تعريف) (reconnaissance). Ces noms sont très polysémiques et leur sens spécialisé n’a pas été retenu par l’extracteur. Nous rappelons que notre corpus général est de 16 millions de mots et constitué d’articles des journaux portant sur une diversité de sujets. Cela peut être une raison possible pour laquelle un certain nombre de noms d’activité n’est pas reconnu comme spécialisé par un système d’extraction automatique.

### *Pour les UTV*

Les résultats de l'extraction révèlent que les UTV sont particulièrement problématiques. D'abord, la qualité de l'extraction n'est pas satisfaisante, notamment en raison du grand nombre d'erreurs d'étiquetage. Nous avons détecté un conflit entre l'analyseur morphologique (AraComLex) et l'étiqueteur (Stanford), surtout lorsqu'il s'agit de l'alternance entre les verbes et les noms. TermoStat nous a proposé certaines UTN comme candidats UTV. Par exemple, la liste présente des UTN comme *ta\$giyl* (تشغيل) (exécution) et *muEAlajap* (معالجة) (traitement).

Ensuite, 58 candidats ne représentent qu'une partie des termes d'activité en informatique. D'après les chiffres présentés dans tableau 9, 35 verbes se distinguent par une spécificité négative, pourtant, nous estimons qu'ils véhiculent des connaissances spécialisées dans le domaine. L'explication pourrait se trouver dans le fait que ces verbes se distinguent par une spécificité sémantique et non pas lexicale (Drouin 2004 : 81). À titre d'exemple, un verbe spécialisé comme *télécharger*, qui est relatif au domaine de l'informatique, possède une forme lexicalisée avec un sens distinct. Par contre, son équivalent ASM, le verbe *Ham~ala* (حمل) ne présente aucune distinction sur le plan lexical. La forme même est ambiguë. Comme discuté à la section 2.3, l'absence des voyelles courtes, ou les diacritiques, augmente l'ambiguïté. De plus, cette absence des voyelles courtes crée une situation d'homographie (voir la section 2.3). Par exemple, la forme même du verbe se confond avec d'autres unités ayant la même forme; à savoir *Ham~ala* (حمل) (charger), *Hamala* (حمل) (porter) et *Himl* (حمل) (fardeau). Toutefois, aucune de ces autres unités ne véhicule un sens spécialisé. À cet effet, Drouin (2004 : 81) constate que :

Multiple phenomena come into play in this case, including homography, polysemy and terminologization (Galissou 1978; Meyer and Mackintosh 2000). In order to be able to identify the lexical items that were missed, one must also look at semantic aspects. Without an additional level of tagging that could take meaning into account, these items cannot be accurately retrieved using a purely statistical approach.

Enfin, une question que nous jugeons importante se pose par rapport aux autres verbes que nous ne trouvons nulle part dans la liste, c'est-à-dire ni parmi les verbes à spécificité positive ni parmi les verbes à spécificité négative ou banale. Ici, un autre phénomène joue un rôle dans la terminologie ASM de l'informatique, surtout lorsqu'il s'agit de traduire de

l'anglais. La construction morphologique des certains verbes change complètement quand la création terminologique par calque est introuvable. C'est le cas plus précisément de verbes comme *redémarrer* et *décompresser*. Souvent ces verbes se traduisent par une construction à verbe support. Par exemple, *décompresser* devient *yafuk~u AlD~agT* (فك الضغط) (défaire la compression) et *redémarrer* se traduit par *yuEydu Al>iqlAE* (يعيد الإقلاع) (refaire le démarrage). En effet, l'ASM dispose d'un système d'affixation relativement pauvre par rapport au français, c'est pour cette raison que les affixes comme *able*, *re*, et *dé* sont souvent traduits par une expression multilexémique.

Dans certains cas, la forme verbale est supplantée par une construction que nous avons remarquée dans notre corpus. Il s'agit d'une structure où un verbe support accompagne la forme nominale du verbe. Par exemple, on peut trouver des phrases comme *yaqwm bilTibAEap* (يقوم بالطباعة) (faire l'impression). Cette structure est récurrente lorsqu'il s'agit d'expliquer un processus ou une manipulation en informatique. De plus, étant donné que l'ASM de spécialité de l'informatique est influencé par l'anglais, on a toujours tendance à reproduire des phrases comme :

[1] *le système vous permet de configurer l'imprimante*

par :

[2] *yasmaH laka AlniZAm bita\$kyI AlTAbiEap*  
 يسمح لك بالنظام بتشكيل الطابعة  
*Le système permet la configuration de l'imprimante*

C'est pour cette raison que, dans la liste initiale, TermoStat a proposé ce genre de verbes comme candidats. Considérons la table 10.

Verbes	Spécificité	Équivalent
<b>tam~a</b> (تم)	178,41	Accomplir
<b>mak~ana</b> (مكن)	165,84	Permettre
<b>qAm</b> (قام)	96,79	Faire
<b>samaHa</b> (سمح)	57,19	Permettre

Tableau 10 : Verbes support



### 5.3 Validation des candidats UTP

Après avoir obtenu les deux listes d'UTP, nous avons procédé à la validation de ces dernières afin de nous assurer qu'elles étaient spécifiques au domaine de l'informatique. La validation se fait en appliquant les critères de validation de L'Homme (2004) (voir section 3.1.3). Dans cette démarche, nous avons utilisé le concordancier *AConcorde* afin de repérer le terme dans son contexte. Les critères de validation terminologique en question sont :

- Le sens de l'UTP est lié au domaine de spécialité;
- Les actants sémantiques de l'UTP sont des termes;
- La parenté morphologique de l'UTP avec d'autres UL déjà admises comme termes;
- Les relations paradigmatiques entretenues par les UTP avec d'autres unités déjà admises comme termes.

En fonction du premier critère, le candidat doit avoir un sens lié au domaine. Ce critère s'applique plus facilement aux UT de nature nominale, étant donné que la plupart de ces unités désignent des entités. En ce qui concerne les UTP, la question devient plus épineuse. Pour cette raison, L'Homme (2004) suggère de combiner à ce critère d'autres critères lexicosémantiques (ibid. : 69). Ces critères sont :

- la cooccurrence compatible;
- la substitution par un synonyme;
- la dérivation morphologique différentielle; et
- la présence de liens paradigmatiques différentiels.

La consultation des dictionnaires spécialisés peut être aussi utile pour certaines UTP. Pour notre recherche, nous avons fait appel à un dictionnaire général en ligne, Almaany. Ce dictionnaire comprend des vocabulaires de différents domaines de spécialité, dont l'informatique. Par exemple, nous avons vérifié le terme *taSaf~aHa* (تصفح) (naviguer) dans ce

dictionnaire. Les résultats obtenus (figure 76) montrent que ce terme existe sous différentes formes. Bien entendu, certains des résultats obtenus relèvent de la langue générale (la colonne de gauche indique le domaine : « général », « computing », « accounting », etc.). Il faut noter que les dictionnaires consultés n'attestent pas toujours de l'emploi spécialisé des UTP. Pour cette raison, nous nous appuyons sur les autres critères pour la validation terminologique de nos termes.

مصطلحات	المعنى	النص الاصيل
عامة	Browsing	تصفح
عامة	Browse; Browsed; Riffle	تصفح
عامة	Skim	تصفح
تقنية	Cleavage; Fora Laminate; Laminate	تصفح
عامة	Browses	يتصفح
مالية	Navigation	تصفح / تجول
مالية	Browse	إستعراض / تصفح
تقنية	Asymmetric Laminate	التصفح الامتثال
الامم المتحدة	browser; Web browser	برنامج تصفح
اعلامية	Zapping	تصفح سريع
الحاسوب	Hierarchical navigation	تصفح شجري
الحاسوب	Surf & Fun	تصفح وامرح
تقنية	Antisymmetric Laminate	التصفح الغير متناظر
تقنية	Anisotropic Lamina	التصفح المتباين الخواص
الحاسوب	Browsing the pages of a WAP service	تصفح صفحات خدمة WAP
الحاسوب	surf	تصفح معلومات على الإنترنت

Figure 76 : Résultats pour la requête *taSaf~aHa* (تصفح) (naviguer) dans Almaany

Le deuxième critère concerne les actants de l'UTP qui sont eux-mêmes considérés comme termes dans le domaine. Si l'UTP se trouve dans un contexte où elle est accompagnée par des termes, il est fort probable qu'elle véhicule un sens spécialisé. À titre d'exemple :

- [1] *Almubarmiju yaktubu barnamajaN*  
المبرمج يكتب برنامجاً  
*Le programmeur écrit un programme*

Dans cet exemple, les actants de l'UTV sont réalisés par des termes du domaine; elle est précédée par *mubarmij* (مبرمج) (programmeur) et succédée par *barnAmij* (برنامج) (programme), qui sont tous deux considérés comme des termes en informatique.

Le troisième critère concerne la parenté morphologique de l'UTP avec des unités qui sont admises comme termes. Par exemple, si le dérivé nominal d'une UTV est un terme, il est fort probable que cette UTV soit un terme elle-même. De même, si le dérivé adjectival d'une UTN est un terme, on peut dire que cette UTN a un sens spécialisé relatif au domaine. Il faut noter que la parenté morphologique d'une UTP doit être accompagnée d'une parenté sémantique.

Le dernier critère s'intéresse aux relations paradigmatiques que l'UTP peut avoir, à titre d'exemple, l'antonymie et la synonymie.

Dans ce qui suit, nous allons montrer avec des exemples l'application de chacun de ces critères pour les UTN et les UTV.

#### *Pour les UTN*

Le premier exemple que nous donnons est l'UTN *tanfy\** (تنفيذ) (exécution). Cette unité présente une spécificité de 61,09 et a 492 occurrences dans le corpus.

#### **Critère 1**

Cette unité est propre au domaine de l'informatique, puisqu'elle exprime l'activité de faire fonctionner une commande informatique. Le dictionnaire que nous avons consulté atteste de cet emploi spécialisé de l'UTN. De plus, les contextes relevés dans le corpus ne présentent pas de nuances de sens.

[1] *Kernel hwa barnAmij yataHak~am bitaxzyn AlbayanAt wa tanfy\* alkwd*  
و الكرنل هو برنامج يتحكم بتخزين البيانات و تنفيذ الكود  
*Kernel est le programme qui contrôle la sauvegarde des données et l'exécution du code*

## Critère 2

Les actants de cette unité sont des termes. Dans l'exemple cité ci-haut, la structure actancielle de l'unité est composée de termes en informatique, à savoir *Kernel* (noyau de système d'exploitation) et *kwd* (كود) (code).

## Critère 3

L'UTN est reliée à deux unités morphologiquement apparentées. Étant donné que cette unité est retenue comme terme en vertu des critères 1 et 2, ses dérivés sont forcément spécialisés. Donc, *tanfy\** (تنفيذ) (exécution) a une parenté morphologique et sémantique avec un verbe, *naf~a\*a* (نفذ) (exécuter) et un adjectif, *tanfi\*y* (تنفيذي) (exécutable). Le verbe renvoie au fait d'amorcer un processus informatique (une commande), tandis que l'adjectif désigne la possibilité qu'un composant informatique (un fichier) accepte de se faire manipuler dans un environnement informatique.

## Critère 4

Les relations paradigmatisées de l'unité avec d'autres termes sont exprimées, entre autres, au moyen de la synonymie ou de l'antonymie. Pour l'UTN en question, nous avons détecté un antonyme, *>inhA'* (إنهاء) (terminer) et un synonyme, *ta\$gyl* (تشغيل) (mise en marche).

Le deuxième exemple est l'UTN *muEAlajap* (معالجة) (traitement). Cette unité a une spécificité de 23,90 et 252 occurrences dans le corpus.

## Critère 1

En fonction de ce critère, l'UTN figure dans une entrée dans le dictionnaire mentionné au début de cette section. Cette unité désigne l'activité effectuée par le processeur dans le traitement des données ; *muEAlajap AlbayanAt >wtwmAtyky~aN* (معالجة البيانات أوتوماتيكيا) (le traitement automatique des données).

## Critère 2

La structure actancielle de cette unité est composée de termes qui relèvent de l'informatique. À titre d'exemple, nous trouvons des termes comme *muEAlij* (معالج) (processeur), *bayanAt* (بيانات) (données) et *naS* (نص) (script) comme dans :

- [1] *Photoshop barnAmijuNyqwm bimuEAlajap AlSwr*  
فوتوشوب برنامج يقوم بمعالجة الصور  
*Photoshop est un programme qui fait le traitement des photos*

## Critère 3

L'UTN a une parenté morphologique avec le nom d'instrument, *muEAlij* (معالج) (processeur), qui est un terme en informatique. De plus, la forme verbale de l'UTN, *EAlaja* (عالج) (traiter) est un terme, puisqu'elle dénote l'activité réalisée par l'instrument.

## Critère 4

Nous n'avons pas trouvé de relations paradigmatiques pour cette unité.

*Pour les UTV*

Le premier exemple que nous donnons est l'UTV candidat *TabaEa* (طبع) (imprimer). Cette unité présente une spécificité négative de -11,18 et a 55 occurrences dans le corpus.

## Critère 1

D'après notre corpus (et les résultats de l'extraction), l'unité a un sens spécialisé propre au domaine de l'informatique. Le sens véhiculé concerne un processus qui consiste à reproduire des données textuelles sur du matériel, notamment du papier, comme le montre l'exemple suivant.

- [1] *taxtalif TAbiEAt All~ayzar Ean gayrihA fiy >an~aha taTbaE AlSafHap kAmilap*  
تختلف طابعات الليزر عن غيرها في أنها تطبع الصفحة كاملة  
*Les imprimantes laser se distinguent des autres imprimantes par leur capacité à imprimer une page entière*

Dans tous les contextes relevés pour cette UTV, nous avons trouvé le même sens qui concerne un processus de reproduction des données par un instrument, à savoir *l'imprimante*. De plus, cette unité apparaît dans le dictionnaire de référence.

## Critère 2

Les actants de cette unité sont eux mêmes des termes. Si l'on considère que *TAbiEap* (طابعة) (imprimante), *HAsuwb* (حاسوب) (ordinateur), *SafHap* (صفحة) (page), *bayAnAt* (بيانات) (données), et *Suwrap* (صورة) (image ou photo) sont des termes en informatique, alors on peut dire que l'UTV a un sens spécialisé.

## Critère 3

L'UTV en question a une parenté morphologique qui la relie à deux unités, à savoir sa forme nominale *TibAEap* (طباعة) (impression) et la forme adjectivale *maTbuwEap* (مطبوعة) (imprimé). Si on considère que l'UTV est spécialisée en fonction des critères 1 et 2, on peut constater que ses dérivés sont aussi spécialisés.

## Critère 4

L'UTV *TabaEa* (طبع) (imprimer) a une seule relation de synonymie. Il s'agit de l'UTV *>xraja* (أخرج) (sortir), comme dans l'exemple suivant :

[1] *AITAbiEap tuxriju waraqAN fArigAN duwn TibAEap*  
الطابعة تخرج ورقا فارغا دون طباعة  
*L'imprimante sort des papiers vierges sans impression*

Le deuxième exemple que nous donnons est l'UTV candidat *naqara* (نقر) (cliquer). En termes de spécificité, cette unité a une spécificité positive de 18,32 et 475 occurrences dans le corpus.

### Critère 1

En ce qui concerne son appartenance au domaine, cette unité apparaît souvent dans des contextes où des directives ou des instructions sont données pour guider l'utilisateur dans une démarche quelconque. Par exemple, lorsqu'on explique aux utilisateurs comment installer un certain logiciel ou télécharger une vidéo de l'Internet. De plus, cette unité apparaît dans le dictionnaire en tant que verbe spécialisé.

### Critère 2

Elle admet dans sa structure actancielle un ensemble de termes, notamment l'instrument utilisé pour effectuer l'action, à savoir *fa>rap* (فأرة) (souris), ou le moyen par lequel l'on effectue l'action *bizir Alfa>rap Al>iman* (بزر الفأرة الأيمن) (avec le bouton droit), ou tout autre lien (sur Internet), bouton (virtuel ou physique), ou icône (représentant une application informatique) avec lequel on peut interagir par ce moyen. À titre d'exemple :

- [1] *liltarAjuE Ean AltagiyyrAt All~atiy tam~a <ijrA&uhA, >unqur fawqa Cancel*  
للتراجع عن التغييرات التي تم إجراؤها, انقر فوق Cancel  
*Pour annuler les modifications, cliquez sur Cancel*

### Critère 3

Dans le contexte spécialisé, cette UTV candidate a une parenté morphologique avec une unité nominale, qui est le terme *Alnaqr* (النقر) (clic) et une autre adjectivale, *qAbil lilnaqr* (قابل للنقر) (cliquable). Puisque cette unité est retenue comme terme en vertu des critères 1 et 2, ses dérivés nominal et adjectival sont considérés comme des termes. Dans l'exemple suivant, nous montrons un contexte où le dérivé nominal de l'UTV est utilisé.

- [1] *AlqA}imap AlmuxtaSarap : wa taZhar Einda Alnaqr biAlzir Al>iman EalY Almawas*  
القائمة المختصرة: و تظهر عند النقر بالزر الأيمن على الماوس  
*Le menu abrégé : il apparaît lorsqu'un clic est fait avec le bouton droit de la souris*

#### Critère 4

Cette unité a une relation de quasi-synonymie avec l'UTV *DagaTa* (ضغط) (appuyer). Nous avons remarqué que les deux UTV véhiculent le même sens dans le discours spécialisé. Notre corpus atteste de 523 occurrences de cette UTV, avec une spécificité positive de 16,28.

- [1] *fy HAl ragiba Almustaxdim <izAlap ha\*A Altavbiyt, yadgaT mar~ap vAniyap biAlzir Al>iman Ealy EinwAn Altabwiyb wa yaxtAr >izAlap tavbiyt*  
في حال رغب المستخدم إزالة هذا التثبيت، يضغط مرة ثانية بالزر الأيمن على عنوان التثبيت و يختار إزالة تثبيت  
*Si l'utilisateur désire annuler l'installation, il appuie encore une fois avec le bouton droit sur le titre de l'onglet et choisit « désinstaller »*

Pour résumer, nous présentons un tableau récapitulatif des verbes et des noms obtenus pour l'analyse.

	UTN	UTV
<b>Total extrait</b>	553	1194
<b>Total filtré</b>	215	375
<b>Total retenu</b>	65	58
<b>Total validé</b>	45	45

Tableau 11 : Tableau récapitulatif des UTP analysées

D'après les chiffres présentés dans le tableau 11, le passage d'*extrait* à *filtré* représente les formes écartées. Comme montré dans le tableau 8 (section 5.2.4.7), ces formes sont les erreurs d'étiquetage et d'autres formes qui relèvent de la langue générale. Quant à *retenu*, cette étape représente les candidats termes. Finalement, *validé* représente les UTV et les UTN que nous avons discutées dans la section 5.3.



### 5.3.1 UTP candidates écartées

Un certain nombre de candidats proposés par TermoStat sont écartés de la validation. Certaines de ces unités répondent aux critères que nous avons énumérés, mais d'autres n'y répondent pas. D'abord, nous avons écarté certains candidats qui n'ont pas suffisamment d'occurrences dans le corpus. Nous avons écarté les candidats dont les occurrences sont inférieures à 5 : nous ne sommes pas en mesure de vérifier clairement leur nature ou leur structure actancielle. Nous mentionnons à titre d'exemple, l'UTV *laSaqa* (لصق) (coller), 3 occurrences, et *>tlafa* (اتلف) (endommager), 4 occurrences.

Ensuite, nous avons écarté les unités qui présentent une ambiguïté. Ces unités sont très polysémiques, et, dans le corpus, elles sont utilisées de différentes façons. Par exemple, le verbe *EaT~ala* (عطّل) (désactiver) peut désigner plusieurs processus en discours spécialisé, comme le fait de désactiver une option, un composant informatique qui tombe en panne ou empêcher le bon déroulement d'un processus. De plus, pour ces verbes, nous n'avons pas observé une normalisation d'utilisation en discours spécialisé. Par conséquent, l'usage de ces verbes reste un choix personnel de l'auteur.

Enfin, certaines unités ne répondent pas aux critères. À titre d'exemple, *Astaxdama* (استخدم) (utiliser) et *AistixdAm* (استخدام) (utilisation). Nous avons obtenu ces unités dans les deux listes de TermoStat, mais nous avons estimé qu'elles relèvent de la langue générale, puisque (en fonction du critère 2) leur sens ne change pas lorsqu'elles admettent des unités de la langue générale dans leur structure actancielle. De plus, nous avons écarté des unités comme *qAma* (قام) (dans le sens de *faire*) et sa forme nominale, *qyAm* (قيام). Cette unité verbale fonctionne comme un verbe support et elle ne véhicule aucun sens spécialisé.

### 5.3.2 Distinction des sens

La validation terminologique est suivie d'un autre processus qui consiste à vérifier davantage les UTP qui sont susceptibles d'avoir plusieurs sens spécialisés dans le domaine. Donc, les résultats présentés dans le tableau 11 ont fait l'objet d'une autre validation. Nous avons voulu qu'un sens corresponde à une UT ; chaque sens doit donner lieu à une entrée.

Nous avons montré dans la section traitant la validation des UTP candidates que les actants admis comme termes peuvent servir d'indice du contenu spécialisé d'une UTV ou d'une UTN. Partant de ce principe, L'Homme (2004 : 69) précise que les contextes des UL dans les textes spécialisés sont d'une grande utilité pour confirmer les connaissances qu'on a du domaine de spécialité et lever une ambiguïté. Donc, en s'inspirant des travaux de Mel'čuk et al. (1995), l'auteure propose quatre tests lexico-sémantiques qui aident le terminographe dans la distinction des sens. Ces tests servent à la fois pour distinguer un sens général d'un sens spécialisé et les sens spécialisés entre eux.

Nous allons appliquer ces tests sur une UTP de notre choix, et, par la suite, nous allons voir comment les différents sens se distinguent selon les contextes. Nous commençons par présenter quatre contextes pris d'une liste de concordances de l'UTN *tavbyt* (تثبيت) (installation). Dans les exemples, nous sommes censé pouvoir distinguer deux sens différents relatifs au domaine de l'informatique.

- [1] *ytwaf~ar maE AlHASib Al~wHy lawHap mafAtyH lAsilkiy~ap taHtwy El|fArp lASMiY~ap wa yumkin tavbyt AljihAz Elayh| litashyl AlAstixdAm*  
 يتوفر مع الحاسب اللوحي لوحة مفاتيح لاسلكية تحتوي على فأرة لمسية و يمكن تثبيت الجهاز عليها لتسهيل الاستخدام  
*Il y a un clavier sans fil qui vient avec la tablette. Il comporte une souris tactile et ce sera possible de faire l'installation de l'appareil dessus pour faciliter l'utilisation.*
- [2] *yanbagy tavbyt malaf~At AlniZAm EIY AlqurS AlSalb*  
 ينبغي تثبيت ملفات النظام على القرص الصلب  
*Il faut faire l'installation des fichiers du système sur le disque dur*

Nous présentons ces deux sens dans le tableau 12.

UTV	Sens
<b>tavbyt 1</b> <b>(installation)</b>	(تثبيت) Le fait de relier deux composants matériels compatibles ensemble afin qu'ils fonctionnent simultanément.
<b>tavbyt 2</b> <b>(installation)</b>	(تثبيت) Le fait de ajouter un composant logiciel à une plateforme (ou environnement) pour qu'ils fonctionnent ensemble.

Tableau 12 : Distinction sémantique de l'UTN *tavbyt* (تثبيت) (installation)

Alors que *tavbyt<sub>1</sub>* (تنبيت) (installation) désigne la relation entre deux composants matériels, *tavbyt<sub>2</sub>* (تنبيت) (installation) désigne le paramétrage d'un composant logiciel afin de l'intégrer à un environnement informatique.

Parmi les quatre tests lexico-sémantiques proposés par L'Homme (2004 : 69), nous allons en appliquer deux à l'unité en question.

En ce qui concerne le premier test, il s'agit de substituer l'UTN par un synonyme afin de vérifier les sens distincts des unités. Nous appliquons ce test au premier et au deuxième contexte que nous avons donnés ci-haut. Dans le cas du premier contexte, l'UTN peut être remplacée par l'UTN *tawSyl* (توصيل) (connexion), comme dans l'exemple suivant :

- [1] *ytwaf~ar maE AlHASib Al~wHy lawHap mafAtyH lAsilkiy~ap taHtwy El|fArp lASMiY~ap wa yumkin tawsyl AljihAz ilayh| litashyl AlAstixdAm*  
 يتوفر مع الحاسب اللوحي لوحة مفاتيح لاسلكية تحتوي على فأرة لمسية و يمكن توصيل الجهاز إليها لتسهيل الاستخدام  
*Il y a un clavier sans fil qui vient avec la tablette. Il comporte une souris tactile à la quelle ce sera possible de faire la connexion de l'appareil pour faciliter l'utilisation.*

En revanche, si l'on fait la même substitution pour le deuxième contexte, le sens de l'UTV risque de changer.

- [2] *\*yanbagy tawSyl malaf~At AlniZAm El| AlqurS AlSalb*  
 \*ينبغي توصيل ملفات النظام على القرص الصلب  
*\*Il faut faire la connexion des fichiers du système sur le disque dur*

Le deuxième test que nous appliquons est la *dérivation morphologique différentielle* (L'Homme 2004 : 71). Il s'agit de dégager « des ensembles de dérivés morphologiques correspondant à des sens distincts ». Par exemple, l'UTV *naqara* (نقر) (cliquer) est liée à *qAbil lilnaqr* (قابل للنقر) (cliquable), comme dans :

- [3] *naqara Ely rAbiT*  
 نقر على رابط  
*il a cliqué sur un lien*
- [4] *rAbiT qAbil lilnaqr*  
 رابط قابل للنقر  
*un lien cliquable*

Cependant, ceci ne s'applique pas à *mutaSaf~iH* (متصفح) (navigateur) :

[5] *naqara EIY almutaSaf~iH*  
نقر على متصفح  
*Il a cliqué sur le navigateur*

[6] *\*mutaSaf~iH qAbil lilnaqr*  
متصفح قابل للنقر  
*\*un navigateur cliquable*

Finalement, pour certains des noms déverbaux, il faut distinguer le sens d'activité du sens concret. Les unités concernées par cette distinction ne sont pas nombreuses. Nous donnons l'exemple de *TibAEap* (طباعة) (impression) :

#### ***TibAEap* (طباعة) (impression) :**

[1] *yumkin AlHuSwl EIY TibAEap mulaw~anap Ean taryq AltaHak~um binisabap xalT*  
*AlAlwAn AlAsAsiyap likul qaTrap wabl wSwlihA iLY alwaraqap*  
يمكن الحصول على طباعة ملونة عن طريق التحكم بنسبة خلط الألوان الأساسية لكل قطرة قبل وصولها إلى الورقة  
*Vous pouvez obtenir une impression couleur en contrôlant le taux de mélange des couleurs primaires pour chaque goutte avant qu'ils atteignent le papier*

Dans le contexte, l'unité en question n'exprime pas l'action; on a affaire à une sorte d'impression (impression couleur) plutôt que l'action d'imprimer. Pour cette raison, nous avons trouvé nécessaire de faire la distinction entre ces unités et les unités exprimant une action. Pour l'instant, et en ce qui concerne l'état actuel de notre travail, nous n'avons pas pris en compte le sens concret exprimé par ces unités. Nous souhaitons mettre en lumière les unités au sens d'activité.

Suite à la distinction de sens, nous avons obtenu un total de 47 UTV et 59 UTN. Dans les annexes 4 et 5, nous donnons deux listes complètes de ces unités avec leurs équivalents anglais et français dans le DiCoInfo.

## **5.4 Choix de contextes pour les UTP retenues**

Après avoir extrait et validé les candidats termes et distingué les sens des unités validées, nous avons procédé au choix des contextes pour chaque UTP retenue. Nous avons retenu jusqu'à vingt contextes pour chaque unité. Nous estimons que ce nombre est suffisant

pour montrer clairement le comportement d'une UTP dans un domaine de spécialité et les différents participants potentiels dans une structure actancielle. Cependant, pour certaines unités, nous n'en avons attribué que cinq. Il s'agit du nombre maximal de contextes trouvés dans le corpus pour ces unités.

Afin d'extraire les contextes, nous avons utilisé un outil disponible sur Internet<sup>52</sup> qui nous a permis de créer des concordances *KWIC* (*Key Words in Context*). Cet outil s'appelle *Sketch Engine* (Kilgarriff et al. 2004) : il s'agit d'un outil de traitement de corpus qui aide à créer une base de donnée textuelle et, par la suite, à extraire des mots clefs ou des candidats termes<sup>53</sup>. Il est compatible avec plusieurs langues, dont l'ASM.

Pour nous, l'usage de cet outil se limite à la gestion de corpus, c'est-à-dire le stockage des textes de notre corpus, la création des listes KWIC et la recherche des UTP.

Nous nous sommes fixé des critères quant à nos choix de contextes. Ces critères sont :

1. Le contexte doit être une phrase complète, ou, à la limite, une phrase faisant partie d'une phrase plus longue à condition que le message transmis soit clair;
2. Le contexte doit refléter le sous-domaine auquel appartient l'UTV;
3. Le contexte doit comprendre un maximum de réalisations des participants de l'UTP;
4. Le contexte doit être entièrement en ASM.

Nous détaillons ces critères dans ce qui suit.

### **Critère 1**

Afin de faciliter la lecture et la compréhension des contextes où figurent les UTV et les UTN, nous optons pour des phrases complètes et non abrégées. Le type de phrase qui nous

---

<sup>52</sup> L'utilisation de Sketch Engine exige une licence. Nous l'avons obtenue suite à un accord conclu avec le laboratoire OLST.

<sup>53</sup> Cette fonctionnalité n'est pas disponible pour la langue arabe.

intéresse ici est celle qui donne une idée claire du rôle de l'UTV ou de l'UTN dans la phrase. Nous avons trouvé des phrases très longues, à tel point qu'il était difficile de les retenir. Ci-après, nous donnons des exemples<sup>54</sup> de contextes relatifs à ce critère. Dans les exemples, nous avons mis les verbes et les noms qui nous intéressent en gras.

- [1] *qabla <yqAf Alta\$giyl, Hawil >an **tugliq** jamyE AlbarAmij AlmaftwHap*  
 قبل إيقاف التشغيل، حاول أن تغلق جميع البرامج المفتوحة  
*Avant l'arrêt de l'exécution, essayez de **fermer** tous les programmes ouverts*
- [2] *AlmutaSaf-iH **yHam~il** AltaHdyvAt tilqA}y~AN*  
 المتصفح يحمل التحديثات تلقائياً  
*Le navigateur **télécharge** les mises à jour automatiquement*
- [3] *yanbagy **tavbyt** malaf~At alniZAm EIY alqurS AlSalb*  
 ينبغي تثبيت ملفات النظام على القرص الصلب  
*Il faut **installer** les fichiers du système sur le disque dur*
- [4] *nuEyd <iqlAE AljihAz ly>xi\* Al<iEdAdAt Aljadydap*  
 نعيد إقلاع الجهاز ليأخذ الإعدادات الجديدة  
*Nous **redémarrons** l'appareil pour qu'il prenne en compte les nouveaux paramètres*

## Critère 2

Le contexte choisi reflète le domaine, ou encore le sous-domaine, auquel appartient l'UTP. Dans la section traitant le sujet de notre corpus (5.1.2), nous nous sommes attardé sur la question de la diversité de nos sujets. Nous avons dit que nous nous intéressons à la technologie informatique et à ses différentes branches allant des composants logiciels et matériels des ordinateurs aux systèmes des cellulaires et tablettes. Dans cette perspective, nous jugeons utile que le contexte donne une idée claire de ces sous-domaines. Considérons les exemples suivants :

---

<sup>54</sup> Dans les deux derniers exemples, nous donnons deux UTN qui sont traduites par des verbes en français. Donc, nous parlons de « installation » et de « redémarrage ».

- [1] *hunAk alkavyr min Al> \$ya' Al~aty tuHad~id qudrapAlmuEAlij ELY tanfy\* almahAm bisurEp >kbar*  
 هناك الكثير من الأشياء التي تحدد قدرة المعالج على تنفيذ المهام بصورة أكبر  
*Il existe plusieurs facteurs qui déterminent la capacité du processeur dans son exécution des tâches rapidement*
- [2] *mwzyl~A tuTaw~iru namw\*aj li>Aw~al HASib lawHiy~ muzaw~ad biniZAm fayarfws*  
 موزيلا تطور نموذجا لأول حاسب لوحي مزود بنظام فايرفوكس  
*Mozilla développe le prototype d'une première tablette équipée du système Firefox*
- [3] *AlHASib yat~aSil bil>intirnit*  
 الحاسب يتصل بالإنترنت  
*L'ordinateur se connecte à l'Internet*
- [4] *ha\*A AljihAz yaTbaE bisurEap taSil >IY 11 SafHap fy Aldaqyqap*  
 هذا الجهاز يطبع بسرعة تصل إلى 11 صفحة في الدقيقة  
*Cet appareil imprime à une vitesse de 11 pages par minute*

Les exemples proviennent de plusieurs sous-domaines. Le premier exemple parle de la performance dans le domaine des composants matériels de l'ordinateur. Nous observons les éléments relatifs à ce sujet, comme *muEAlij* (معالج) (processeur) et *mahAm* (مهام) (tâches). Dans le deuxième exemple, les éléments observés indiquent que l'on parle d'un système d'exploitation conçu pour une nouvelle tablette. Donc nous trouvons le nom de la compagnie *Mozilla*, connu pour la production des solutions relatives au navigateur (donc logiciel et internet), *HASib lawHiy* (حاسب لوحي) (tablette), et *niZAm* (نظام) (système). Dans le troisième exemple, le verbe *yat~aSil* (يتصل) (se connecter) désigne le processus selon lequel un composant informatique établit une connexion avec l'Internet, *HASib* (حاسب) (ordinateur) et *>intirnit* (إنترنت) (Internet). Dans le quatrième exemple, le verbe *TabaEa* (طبع) (imprimer) est relatif au sous-domaine de l'impression en informatique. Dans un tel contexte, nous remarquons la présence des termes comme *TabiEap* (طابعة) (imprimante), *SafHap* (صفحة) (page) et la notion de la vitesse d'impression.

### Critère 3

Dans le cadre de notre thèse, nous nous intéressons à l'analyse et à la description des UTP et de leur structure actancielle. Il importe donc que, dans les contextes choisis, le maximum de participants de l'UTP soient présents. De plus, la présence de ces participants

nous aide à mieux identifier le comportement linguistique de l'unité même. Cependant, l'absence de certains éléments (ou encore participants) ne signifie pas que le contexte ne fera pas l'objet d'une sélection de notre part. Dans certains cas, le contexte ne comprend pas suffisamment de participants pour toutes sortes de raison. Nous présentons dans les exemples ci-dessous des contextes retenus pour l'analyse.

[1] >**uDbuT** xayArAt mutaSaf~iH Al<Aintarnit  
أضبط خيارات متصفح الإنترنت  
**Réglez les options du navigateur**

[2] >wtwkAd **yun\$}** Tabaqap <iftrADy~ap  
أوتوكاد ينشئ طبقة افتراضية  
**AutoCAD crée une couche virtuelle**

[3] **baEda Antiha' Eamaly~ap Altahy}ap, tabdA Eamaly~ap nasx Almalaf~At**  
بعد إنتهاء عملية التهيئة تبدأ عملية نسخ الملفات  
**À la fin du formatage, le processus de la copie des fichiers commence**

[4] **lilrAgibyn bitaHmyl** >y min ha\*ihi AltawzyEAt maj~AnAN min Al>intirnit mur.AjaEap  
AlrAbiT AltAly  
للمرغبيين بتحميل أيا من هذه التوزيعات مجانا من الانترنت مراجعة الرابط التالي  
**Pour ceux qui désirent télécharger l'une de ces distributions de l'Internet, veuillez voir le lien suivant**

Dans l'exemple [1], l'UTV *DabaTa* (ضبط) (régler) admet dans sa structure actancielle l'Agent, en charge de l'action, et le Patient, le composant logiciel qui subit l'action. Étant donné que le verbe dans ce contexte est conjugué à l'impératif, son sujet syntaxique est compris dans le verbe au moyen d'un pronom. Ce cas est fréquent dans les manuels, surtout lorsqu'il s'agit des verbes comme *naqara* (نقر) (cliquer), *DagaTa* (ضغط) (appuyer) et >n\$> (أنشأ) (créer). L'exemple [2] présente un cas typique où tous les éléments de la structure actancielle sont présents, à savoir l'Agent, l'entité responsable de l'action, et le Patient, le composant logiciel créé. Dans l'exemple [3], nous avons une UTN. Dans ce contexte, l'unité désigne le nom de l'action, et par conséquent, l'Agent est absent, mais le Patient est présent. Ce genre de structure est fréquente pour les UTN dans des contextes où l'unité désigne le processus (l'action decopier). Dans le dernier exemple, nous avons aussi une UTN, réalisée par un syntagme prépositionnel (littéralement cette phrase se traduit : *pour ceux qui désirent faire le*



*téléchargement de l'une de ces distributions de l'Internet*). Dans la phrase, nous avons l'Agent, *lilrAgibyn* (للراغبين) (*les personnes qui s'intéressent à*), le Patient, *tawzyEAt* (توزيعات) (distributions) et la Source, *Antirnit* (انترنت) (Internet).

#### **Critère 4**

Le dernier critère concerne la langue des contextes. Étant donné que nous travaillons sur la terminologie de l'informatique en ASM, il importe que nous présentions des contextes entièrement dans cette langue. Nous nous intéressons aux contextes où les termes sont traduits, c'est-à-dire que l'auteur emploie la terminologie en ASM, ou il fait la translittération en caractères arabes des termes qui n'ont pas d'équivalents. Comme ce n'est pas toujours le cas, nous retenons les contextes où des termes anglais sont présents, à condition que l'UTP y figure en ASM. Dans les exemples présentés ci-haut, on observe la manière dont les idées sont exprimées. À titre d'exemple, le terme *Bios* existe souvent dans cette forme. Parfois, nous le trouvons translittéré dans certains contextes, *byws* (بيوس).

### **5.5 Annotation des contextes des UTP dans XML**

Cette étape dans notre méthodologie de travail consiste à encoder les données recueillies dans une base de données. L'encodage se fait au moyen d'un éditeur XML qui s'appelle oXygen. Notre méthodologie d'encodage suit celle utilisée dans les différents dictionnaires de DiCoInfo qui s'inspire elle-même de celle utilisée dans FrameNet. Toutefois, nous avons adapté cette méthodologie pour convenir aux objectifs de notre recherche.

Le structure d'encodage (qui avait déjà été mis au point) comprend les éléments suivants : le vocable, la lexie, l'information grammaticale, la définition, le domaine, le régime, la structure actancielle, les variantes, les synonymes, les réalisations, les liens lexicaux, les contextes, les informations complémentaires et les équivalents, comme le montre la figure 77.

```

<vocable xmlns="http://olst.ling.umontreal.ca/dicoinfo/" xmlns:xi="http://www.w3.org/2003/XInclude"
  <lexie numero-acceptation="" statut="" projet="" redacteur="" mise-a-jour="" xml:id="_modele_">
    <information-grammaticale/>
    <definition/>
    <domaine/>
    <regimes/>
    <structure-actancielle/>
    <feminins/>
    <variantes/>
    <synonymes/>
    <realisations/>
    <liens-lexicaux/>
    <contextes>
      <contexte source="" statut="" mise-a-jour="" annotateur="">
        <contexte-texte></contexte-texte>
      </contexte>
    </contextes>
    <informations-complementaires/>
    <equivalences/>
  </lexie>
</vocable>

```

Figure 77 : Structure d'encodage dans le DiCoInfo

Tous ces éléments forment ce qu'on appelle une *fiche terminologique*. Nous avons créé une fiche terminologique pour chaque UTP retenue dans le cadre de notre travail. Pour la version en ASM de DiCoInfo, certains éléments de la fiche sont ignorés : définition, domaine, régimes, féminins, variantes, synonymes, liens lexicaux et informations complémentaires.

Étant donné que l'encodage des données se fait avec un éditeur XML, chaque élément prend la forme d'une balise. À l'intérieur de chaque balise, on trouve des attributs servant à donner des renseignements supplémentaires sur le contenu de la balise. Dans ce qui suit, nous nous attardons sur chacun des éléments que nous avons utilisés.

### 5.5.1 Vocabulaire

Le vocable est le premier élément dans la création de la fiche. Il comprend des attributs qui désignent l'adresse de la fiche en ligne, un identificateur du terme et la langue de travail (ar), figure 78.

```

<vocable xmlns="http://olst.ling.umontreal.ca/dicoinfo/" xmlns:xi="http://www.w3.org/2003/XInclude"
  xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:saxon="http://saxon.sf.net/" xmlns:xs="http://www.w3.org/2001/XMLSchema"
  identificateur="_téléchargement_taHmyl_" xml:lang="ar">

```

Figure 78 : Vocabulaire dans une fiche dans XML

La figure 78 reproduit une partie de la fiche de l'UTP *taHmyl* (تحميل) (téléchargement). Dans notre cas, nous avons ajouté la forme translittérée du terme dans l'attribut **identificateur=** suivi de l'équivalent français. De cette façon, l'utilisateur a le choix de chercher soit par forme translittérée (s'il ne dispose pas d'un clavier qui supporte la langue arabe), soit par équivalent français.

## 5.5.2 Lexie

Dans cette balise, on précise l'acception d'un terme dans la fiche. La balise `<lexie>` comprend les attributs suivants : numéro de l'acception, statut, projet, rédacteur, mise à jour et identificateur du terme dans le fichier XML, comme dans l'exemple.

```
<lexie numero-acception="1" statut="1" projet="" redacteur="NG" mise-a-jour="2014-02-08"
xml:id="_tourner_1_عمل_">
```

Dans cet exemple, nous présentons les informations relatives à l'UTV *Eamila* (عمل) (tourner). Pour cette unité, il y a deux acceptions, donc, deux entrées. Le numéro de statut est 1, ce qui veut dire que la rédaction de cette fiche est très avancée. Ensuite, on donne le nom de projet (nous parlons de cette balise à titre indicatif, car nous ne l'avons pas utilisée), les initiales de la personne qui a créé la fiche, la date à laquelle elle a été mise à jour et son identificateur dans XML. Dans l'attribut **xml:id=**, nous avons mis l'équivalent français suivi du numéro de l'acception et le terme en ASM. De cette manière, la fiche peut être facilement repérée dans le dictionnaire, soit par le terme en langue arabe, soit par son équivalent français.

La balise dédiée à la lexie comprend plusieurs sous-balises. Il s'agit des informations détaillées sur le terme. Dans ce qui suit, nous discutons de chacun de ces éléments.

## 5.5.3 Information grammaticale

Pour cette partie, le seul élément qui nous intéresse est celui de `<information-grammaticale>`. Pour notre recherche, nous nous limitons à deux parties du discours, à savoir *v.* pour le verbe et *n.* pour le nom. Nous ajoutons aussi des informations quant à la transitivité

du verbe, *v.t.* ou *v.in.* Pour les noms, nous ajoutons le genre, *f* pour féminin et *m* pour masculin (comme en français).

#### 5.5.4 Structure actancielle

Pour chaque UTP, nous définissons la structure actancielle. La structure actancielle d'un terme nous permet d'identifier ses différents actants (c'est-à-dire les participants qui sont obligatoires quant au sens du terme). Cette balise, elle aussi, comprend des sous-balises. Ces dernières englobent les rôles sémantiques de chaque actant et le terme typique (figure 79).

```
<structure-actancielle>دعم  
  <role nom="Destination">  
    <tt>1 لوحة أم</tt>  
  </role>  
  <role nom="Patient">  
    <tt>1 معالج</tt>  
  </role>  
</structure-actancielle>
```

Figure 79 : Structure actancielle dans XML

Dans la figure, nous donnons l'exemple de l'UTV *daEama* (دعم) (accepter). Nous commençons par inscrire le verbe au début de la balise et, par la suite, au moyen de la balise `<role nom="">` nous choisissons les rôles sémantiques pour chaque actant à partir d'un menu compris dans l'attribut `nom=""`. Enfin, le terme typique occupe la balise `<tt>`. Dans l'exemple, l'UTV admet deux actants, à savoir Destination et Patient, qui sont représentés par les termes typiques suivants : *lawHap Aum* (لوحة أم) (carte mère) et *muEAlij* (معالج) (processeur). Il faut noter que le numéro inscrit à côté de chaque terme typique correspond à son numéro d'acceptation dans le dictionnaire. Pour notre recherche, nous ne sommes pas concerné par les noms d'entité, donc, il n'y a pas d'entrées pour ces termes.

## 5.5.5 Réalisations

Dans cette balise, nous inscrivons les réalisations linguistiques de chaque actant. Cela nous permettra de les afficher dans l'interface web de notre dictionnaire (figure 80).

```
<realisations>
  <role nom="Destination">
    <realisation identificateur="لوحة أم" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="بطاقة رسوميات" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="نظام تشغيل" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="معالج ذكي" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="حاسب لوحي" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
  </role>
  <role nom="Patient">
    <realisation identificateur="معالج" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="فيديو" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="معالج" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="تشغيل" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="إنصال" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="خريطة" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="سيرة نقل" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
    <realisation identificateur="دفء مائية" numero-acceptio="1" xlink:type="simple" xlink:show="replace" xlink:actuate="onRequest" xlink:href="">1</realisation>
  </role>
</realisations>
```

Figure 80 : Réalisations linguistiques des actants dans XML

Dans la figure, nous présentons les réalisations linguistiques de l'UTV *daEama* (دعم) (accepter). Pour chaque balise `<realisation>`, on inscrit des informations relatives à l'identification de l'unité, faisant l'objet de la réalisation, dans le dictionnaire. Dans la figure, le rôle Destination est réalisé par des termes comme *lawHap Aum* (لوحة أم) (carte mère), *niZAm ta\$gyl* (نظام تشغيل) (système d'exploitation) et *biTAqap ruswmiy~at* (بطاقة رسوميات) (carte graphique). Pour le rôle Patient, il est réalisé par des termes comme *muEAlij* (معالج) (processeur), *fydyw* (فيديو) (vidéo) et *iEyar* (معالج) (module). Finalement, le rôle Destination est réalisé par *qurS* (قرص) (disque) et *tabwyb* (تبويب) (onglet). Bien entendu, comme nous travaillons sur les unités prédicatives, il n'existe pas d'entrées pour ces termes.

## 5.5.6 Contextes

Il s'agit de la balise où sont mis les contextes annotés d'un terme. Les sous-balises comprises dans cette balise sont : `<contexte>`, `<contexte-texte>`, `<participant>` et `<lexie-att>`. Suivant la méthodologie de FrameNet et du DiCoInfo, nous annotons tous les contextes que nous donnons pour les termes analysés.

Le schéma de l'annotation d'un contexte commence par la balise `<contexte>` où l'on trouve des informations relatives à la source du contexte, le statut de sa rédaction, l'annotateur

et la mise à jour. Ensuite, dans la balise <contexte-texte>, on donne le contexte même. Cette partie de notre travail était problématique. Nous avons des problèmes techniques quant à l'insertion des contextes en ASM (caractères arabes) contenant des mots anglais. Nous avons remarqué que les phrases ne s'affichaient pas correctement, comme dans l'exemple suivant :

<contexte-texte>على قرص للقراءة فقط XMP أن تكتب ملف Camera Raw لا يمكن لإضافة</contexte-texte>

L'ordre des mots de la phrase est incorrect. On remarque la présence de deux mots en anglais et la phrase commence à gauche, contrairement au sens de l'écriture en langue arabe. La solution que nous avons trouvée à ce problème consistait à insérer le contexte d'une manière verticale, faisant en sorte que les mots anglais soient placés sur une ligne séparée. Cependant, nous avons remarqué que ces contextes s'affichent correctement sur la page HTML.

Après avoir mis le contexte dans la balise, la prochaine étape consiste à faire l'annotation du contexte. Dans la figure 81, nous montrons l'annotation de l'UTV *daEama* (دعم) (accepter).

```
<contexte source="ITP" statut="1" annotateur="NG" mise-a-jour="2014-02-21">
<contexte-texte> أن استخدام هذا البرنامج يتيح تشغيل الفيديو حتى لو كانت بطاقة الرسومات لا تدعم الفيديو عالي الوضوح بشكل كامل </contexte-texte>
أن استخدام هذا البرنامج يتيح تشغيل الفيديو حتى لو كانت
<participant type="Act" role="Destination">
  <fonction-syntaxique nom="Sujet" cas="Agentif">
    <groupe-syntaxique nom="SN">
      <realisation>بطاقة الرسومات</realisation>
    </groupe-syntaxique>
  </fonction-syntaxique>
</participant>
لا
<lexie-att lemme="دعم">
  </lexie-att>
<participant type="Act" role="Patient">
  <fonction-syntaxique nom="Objet" cas="Accusatif">
    <groupe-syntaxique nom="SN">
      <realisation>الفيديو</realisation>
    </groupe-syntaxique>
  </fonction-syntaxique>
</participant> بشكل كامل
</contexte>
```

Figure 81 : Contexte annoté dans XML

Le contexte qui fait l'objet de notre exemple est :

- [1] *AstixdAm ha\*a AlbarnAmij yutyHu ta\$gyl Alfgydyw HtY law kAnat biTAqap Alruswmiy~at  
LA tadEam Alfgydyw EAlY AlwDwH bi\$akl kAmil*  
استخدام هذا البرنامج يتيح تشغيل الفيديو حتى لو كانت بطاقة الرسومات لا تدعم الفيديو عالي الوضوح بشكل كامل  
*L'utilisation de ce programme permet l'exécution de cette vidéo même si la carte  
graphique n'accepte pas les vidéos HD*

Dans un contexte, les seuls éléments annotés sont les actants et le terme cible. C'est pour cette raison qu'on trouve que la phrase *AstixdAm ha\*A AlbarnAmij yutyHu ta\$gyl Alfgydyw HtY law kAnat* (استخدام هذا البرنامج يتيح تشغيل الفيديو حتى لو كانت) (L'utilisation de ce programme permet l'exécution de cette vidéo même si), située en haut dans la figure, n'est pas annotée.

### 5.5.6.1 Participant

La balise `<participant>` définit le type d'un participant, actant ou circonstant, et son rôle. De plus, elle englobe trois sous-balises qui fournissent des renseignements sur les propriétés linguistiques du participant et sa réalisation linguistique.

Dans le cas de notre exemple, l'annotation commence par préciser le premier participant (nous prenons en compte tout le groupe syntaxique) qui est *biTAqap ruswmiy~At* (بطاقة رسومات) (*carte graphique*). Ce participant est de type Actant et son rôle sémantique est Destination, `<participant type="Act" role="Destination">`. En ce qui concerne ses propriétés syntaxiques, dans la balise `<fonction-syntaxique>`, nous précisons la fonction, Sujet, et le cas, Agentif, `nom="Sujet" cas="Agentif"`. Dans notre exemple, le participant est un syntagme nominal (SN), puisqu'il est le sujet du verbe *tadEam* (تدعم) (accepter). Finalement, dans la balise `<groupe-syntaxique>`, nous précisons le groupe syntaxique, SN pour syntagme nominal. L'annotation du participant s'arrête à sa réalisation linguistique qu'on trouve dans la balise `<realisation>` بطاقة الرسومات `</realisation>`.

### 5.5.6.2 Étiquettes des rôles

Étant donné que notre travail est basé sur le DiCoInfo, la notion d'« actant » dans la méthodologie d'étiquetage correspond à celle de « frame element » de FrameNet. En ce qui concerne les étiquettes mêmes, nous adoptons celles utilisées dans le DiCoInfo. Ce choix

repose sur deux raisons. Premièrement, comme nous souhaitons établir les équivalents avec les autres versions du DiCoInfo, nous trouvons important que les étiquettes soient les mêmes. Cela permet de mieux comprendre les correspondances entre les trois langues (ASM, anglais et français). Deuxièmement, étant donné que nous envisageons inclure des données anglaises et françaises dans une ressource basée sur les cadres en informatique (voir section 5.6), il importe que toutes les UTP aient la même structure, en ce qui concerne les actants.

Dans le tableau 13, nous donnons quelques exemples des étiquettes les plus utilisées dans notre travail. Ces définitions sont inspirées du DiCoInfo. Dans l'Annexe 6, nous donnons une liste complète des rôles que nous avons utilisés.

Rôle sémantique	Définition	Exemple
<b>Agent</b>	La notion d'Agent désigne le participant qui se trouve à l'origine de l'action.	<i>An\$A mustaxdim</i> [Agent] <i>HisAbN</i> [Patient] أنشأ مستخدم حسابا <i>L'utilisateur</i> [Agent] <i>a créé un</i> <i>compte</i> [Patient]
<b>Patient</b>	Cet actant désigne l'entité subissant l'action. Patient peut renvoyer soit à l'élément créé ou utilisé par l'Agent.	<i>wiHdap</i> <i>AlmuEAlajap</i> <i>Almarkazy~ap</i> [Agent ] <i>tunaf~i*</i> <i>Al&gt;wAmir</i> [Patient] وحدة المعالجة المركزية تنفذ الأوامر <i>L'unité centrale</i> [Agent] <i>exécute les</i> <i>commandes</i> [Patient]
<b>Instrument</b>	L'actant Instrument renvoie à l'entité qui désigne le moyen par lequel l'action est effectuée.	<i>&gt;unqur bizir AlmAws</i> [Instrument] <i>EIY</i> <i>Almalaf</i> [Patient] انقر بزر الماوس على الملف <i>Cliquez avec le bouton de la</i> <i>souris</i> [Instrument] <i>sur le dossier</i> [Patient]
<b>Destination</b>	L'actant Destination désigne l'objectif de l'action effectuée par l'Agent. De plus, il renvoie à la finalité typique d'une action.	<i>yrafaE Almustxdim</i> [Agent] <i>AlSwrap</i> [Patient] <i>EIY Flickr</i> [Destination] يرفع المستخدم الصورة على فليكر <i>L'utilisateur</i> [Agent] <i>téléverse la</i> <i>photo</i> [Patient] <i>sur Flickr</i> [Destination]

Tableau 13 : Partie des rôles sémantiques utilisés

### 5.5.6.3 Lexie

Le terme cible occupe la balise `<lexie-att>` et dans l'attribut `lemme=""` on met la forme lemmatisée du terme. La difficulté rencontrée ici concerne la confusion que crée l'insertion du lemme dans la partie attribut, comme dans : `<lexie-att lemme="كتب">تكتب</lexie-`



att>. Pour cette raison, nous avons mis le verbe sur une ligne séparée comme c'est le cas des contextes comportant des mots anglais.

Comme montré dans la figure 81 dans la section 5.5.6, la structure actancielle de cette phrase est composée de deux participants, Patient et Destination. Ces participants suivent le même schéma d'annotation, en ce qui concerne la présentation des informations relatives aux types et rôle sémantique, propriétés syntaxiques et réalisation.

### 5.5.7 Équivalent

La dernière balise dans le modèle de la fiche terminologique est dédiée à l'équivalent, <equivalences>. Dans le cadre de notre travail, nous inscrivons les équivalents français et anglais, tout en respectant le numéro d'acceptation (figure 82).

```
xlink:href="../../../dicoInfo-fr/accepter.xml#_accepter2" xlink:show="replace" xlink:actuate="onRequest">accepter Z</equivalence>  
xlink:href="../../../dicoInfo-en/support.xml#_support2" xlink:show="replace" xlink:actuate="onRequest">support Z</equivalence>
```

Figure 82 : Balise <equivalences> dans XML

Comme montré dans la figure, cette balise regroupe un certain nombre d'attributs qui renvoient aux fiches des termes équivalents dans le DiCoInfo. Nous signalons que la façon dont nous avons associé les UTP de l'ASM aux UTP françaises et anglaises repose sur les deux éléments suivants :

1. le sens de l'UTP : pour qu'une UTP corresponde à une autre (française ou anglaise), il faut que l'unité ait le même sens. Ceci est déterminé par la définition des unités françaises et le sens des unités de l'ASM constaté dans le corpus. Par exemple, l'UTN française *traitement* est définie comme : « Activité réalisée par un Agent, un ordinateur ou un processeur qui manipule des données au moyen d'opérations, souvent dans le but de transformer ces données. » Cette définition correspond au sens de l'UTN *muEAlajap* (معالجة) (traitement) dans le corpus;

2. la structure actancielle des UTP de l'ASM par rapport à celle des autres UTP dans les deux autres langues. Cet élément est particulièrement important. Pour retenir une correspondance, il faut que les unités en question aient la même structure actancielle. Par exemple, l'UTN *muEAlaja<sub>1p</sub>* (معالجة) (traitement) a la structure actancielle : ~ de Patient par Agent. Cette structure correspond à la structure actancielle de l'UTN française *traitement* dans le DiCoInfo.

### 5.5.8 Particularités de l'annotation des contextes en langue arabe

Comme montré dans les sections 5.5.6 et 5.5.6.3, nous avons à faire face à un certain nombre de difficultés lors de l'insertion des contextes arabes dans l'éditeur XML, du fait que la langue arabe n'est pas complètement compatible avec cet éditeur. Pour contourner les difficultés que représentent les langues avec le système d'écriture allant de droite à gauche, l'éditeur propose deux options. Premièrement, il détecte automatiquement le sens de l'écriture et essaie d'adapter les balises en conséquence. Cette option ne s'est pas avérée utile, puisque nous n'avons remarqué aucun changement réel. Une autre option que l'éditeur propose est d'inverser les balises (dans ce cas, les balises se positionnent à droite). Nous n'avons pas utilisé cette solution, puisque nos fiches comportent des informations en caractères latins (les noms de balises, les actants, etc.). Pour cette raison, nous avons à nous adapter. Il convient de signaler que cette situation n'est pas présente sur toutes les plateformes informatiques (en effet, la langue arabe est compatible avec tous les environnements informatiques et avec un grand nombre d'applications et de programmes).

En plus de ces difficultés techniques, nous avons observé des cas particuliers lors de l'annotation des contextes. Ces cas relèvent du système morphologique riche et complexe de la langue arabe. Nous rappelons qu'à la section 2.3, nous avons parlé du système morphologique et avons donné l'exemple de la forme *wasayaktubwnhA* (وسيكتبونها) qui signifie *et ils vont l'écrire*. De plus, à la section 3.2.1.2, nous avons présenté les difficultés de l'annotation des clitiques des verbes dans Arabic PropBank (APB). Dans notre cas, les particularités observées concernent l'annotation des contextes où le sujet syntaxique de la

phrase et les compléments d'objet direct sont réalisés par un suffixe pronominal. Considérons les deux exemples suivants<sup>55</sup>.

[1] *i\*A AdxalnA qimatayn, Alqymap AlAftirADiy~ap AlvAnyap sawf tulgY*

إذا أدخلنا قيمتين، القيمة الافتراضية الثانية سوف تلغى

*Si nous entrons deux valeurs, la deuxième valeur virtuelle sera annulée.*

[2] *tastaqbil ALTAbiEap AlbayAnAt wa taHfaZhA fy \*Akirap Ea\$wA}y~ap*

تستقبل الطابعة البيانات وتحفظها في ذاكرة عشوائية

*L'imprimante reçoit les données et les sauvegarde dans une mémoire vive.*

D'après la morphologie de l'ASM, le sujet du verbe (lorsqu'il est un pronom) est réalisé par un suffixe faisant partie du verbe. Dans l'exemple [1], le verbe *AdxalnA* (أدخلنا) (nous entrons) peut être analysé comme suit : *Adxal* (entrer) et *nA* (un suffixe pronominal qui désigne *nous*). Dans l'exemple [2], le terme *AlbayAnAt* (بيانات) (les données) est réalisé par le suffixe pronominal *hA* (ها), comme dans *taHfaZhA* (تحفظها) (les sauvegarder) qui s'analyse comme suit : *taHfaZ* (تحفظ) (sauvegarde) et *hA* (ها) (les). Afin de mettre en évidence cette particularité, et comme ce suffixe est la réalisation linguistique du Patient, dans la balise <participant>, nous avons mis le suffixe pronominal entre parenthèses. Considérons l'exemple suivant pour l'annotation de l'exemple [2].

```
<antecedentxml:id="sauvegarder_1_حفظ_ref5">
  <valeur-antecedent>البيانات</valeur-antecedent>
</antecedent> و
<lexie-attlemme="حفظ">
  تحفظها
</lexie-att>
<participanttype="Act"role="Patient">
  <fonction-syntaxiquenom="Objet">
    <groupe-syntaxiquenom="Pro"particule="ها">
      <realisationlemme="ه"ref="sauvegarder_1_حفظ_ref5">
        (ها)
      </realisation>
    </groupe-syntaxique>
  </fonction-syntaxique>
</participant>
```

<sup>55</sup> Nous signalons que la traduction française des contextes ne montre pas ces cas particuliers, puisque cela relève d'une particularité de la langue arabe.

Pour montrer la correspondance entre le suffixe et l'unité (*bayanAt*, données) à laquelle il fait référence dans la version web, nous avons ajouté l'attribut `ref="sauvegarder_1_حفظ_ref5"`.

Dans le cas du premier exemple, nous n'avons pas suivi cette méthode. Nous avons considéré que même si la réalisation linguistique de l'actant est présente (le suffixe pronominal attaché au verbe), il n'existe pas une unité de référence. Donc, nous ne mettons rien.

## 5.6 Identification des cadres

Notre travail exigeait une découverte de cadres pour le domaine de l'informatique. Donc, en plus de proposer une méthodologie de compilation d'une ressource terminologique pour une langue comme l'ASM, nous proposons également une modélisation des termes de l'informatique selon la théorie de la Sémantique des cadres. Pour cette raison, nous croyons que notre méthodologie se distingue par un aspect multidimensionnel, c'est-à-dire qu'elle combine une description des UTP basée sur leurs propriétés syntaxico-sémantiques et une modélisation conceptuelle de ces unités en fonction des cadres sémantiques.

Nous rappelons que la Sémantique des cadres est une théorie qui repose sur l'idée des cadres sémantiques. Le cadre est un système conceptuel selon lequel un concept peut être compris en fonction du système entier. Par conséquent, le sens des mots est décrit par rapport aux cadres sémantiques. Cette théorie a fait l'objet d'une application en ligne, FrameNet (Ruppenhofer et al. 2010).

Dans l'identification des cadres, nous nous inspirons de la méthodologie suivie dans L'Homme et Robichaud (2014) et L'Homme et al. (2014). Cette méthodologie consiste à : 1) extraire les données du DiCoInfo; et 2) utiliser les données de FrameNet comme une référence pour identifier les cadres que les termes de DiCoInfo peuvent évoquer. Étant donné que les données dans FrameNet sont en anglais, nous nous intéressons à comparer les structures actanciennes qu'admettent les termes cibles dans les deux bases de données.

Dans un premier temps, une page web a été créée à des fins de comparaison. Cette page comporte les termes du DiCoInfo version anglaise et les unités lexicales correspondantes de FrameNet (figure 83).

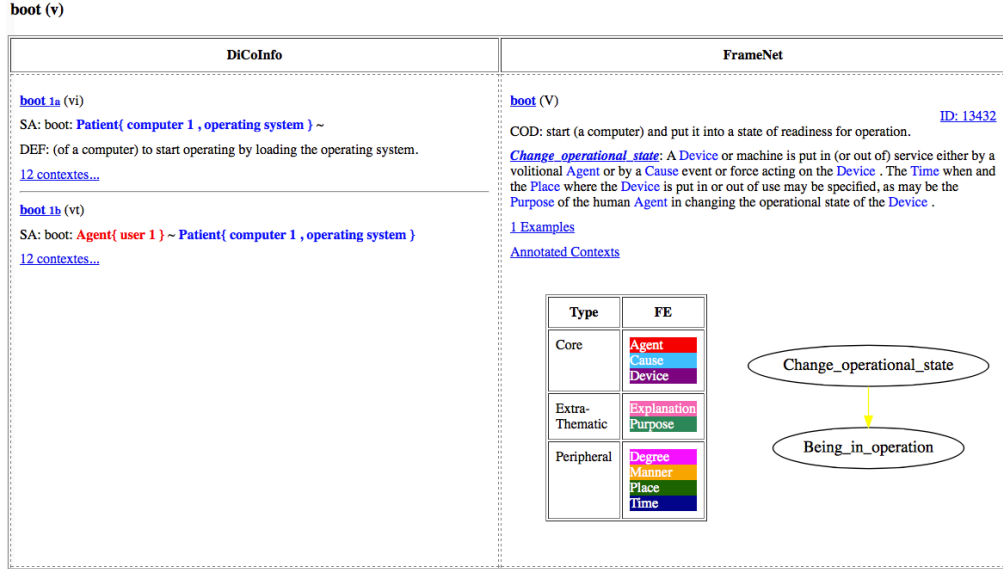


Figure 83 : Comparaison des termes dans DiCoInfo avec les UL dans FrameNet

La comparaison relie les éléments suivants : en ce qui a trait à FrameNet, l'UL (*boot* dans la figure) correspondant à l'UT du DiCoInfo, sa définition, le cadre auquel elle appartient (*Change\_operational\_state* dans la figure), la définition du cadre, un lien vers les contextes annotés, les core-FE, les non-core FE et les relations avec d'autres cadres. En ce qui concerne le DiCoInfo, on y trouve le terme, sa structure actancielle, sa définition et les contextes.

Comme la comparaison se fait avec la version anglaise du DiCoInfo, une deuxième comparaison doit être faite avec les données de la version en ASM. Bien entendu, comme discuté dans la section 1.4, le présent travail constitue une élaboration d'une version en ASM de DiCoInfo, et, par conséquent, nous précisons les correspondances entre les termes ASM et les termes anglais et français du dictionnaire.

### 5.6.1 Différences avec FrameNet

Comme L'Homme et Robichaud (2014 : 191-192) l'ont constaté pour le domaine de l'environnement, il existe un certain nombre de différences entre les deux ressources, différences dont il faut tenir en compte.

Premièrement, sur le plan méthodologique, dans FrameNet, les FE sont définis au niveau de cadres, alors que les actants dans le DiCoInfo sont établis au niveau des UT. Par exemple, un cadre comme Supporting dans FrameNet comprend les core-FE Supported et Supporter. Chacun de ces core-FE exprime des propriétés du cadre : le core-FE Supporter renvoie à une entité qui agit afin de soutenir le FE Supported en fournissant de stabilité et de force. Cependant, certains core-FE sont utilisés de façon générique, puisque l'idée qu'ils expriment se répète dans tous les cadres, comme le core-FE Agent qui désigne la personne à l'origine de l'action.

Deuxièmement, une différence peut être constatée en ce qui concerne le nombre de FE par rapport au nombre d'actants. Cette situation est souvent rencontrée lorsque nous associons les UT du DiCoInfo aux cadres de FrameNet. Un cadre comme Processing\_materials dans FrameNet comprend trois core-FE : Agent, Alterant et Material. Les UTP, *Ealaja* (معالجة) (traiter) et *muEAlajap* (معالجة) (traitement) qui coresspondent à ce cadre dans DiCoInfo ont une structure actancielle composée de deux actants : Agent et Patient.

Finalement, le DiCoInfo présente une particularité étant donné qu'il est conçu pour la terminologie de l'informatique. Dans la comparaison entre les données, il faut tenir en compte le phénomène de la terminologisation, surtout en ce qui concerne les unités prédicatives. Par exemple, nous retenons des unités comme *saj~ala* (سجل) (enregistrer), *tasjyl* (تسجيل) (enregistrement), *enregistrer<sub>1</sub>*, *écrire<sub>1</sub>*, *save<sub>1</sub>* et *load<sub>1b</sub>*. Ces unités existent dans la langue générale avec un sens différent que celui utilisé dans la langue de spécialité.

### 5.6.2 Association des cadres aux UTP

En associant les cadres dans FrameNet, nous nous sommes inspiré du travail discuté dans L'Homme et al. (2014). Nous avons formé quatre groupes :

#### *Correspondance parfaite*

Les données dans les deux ressources se correspondent parfaitement en ce qui concerne le nombre d'actants et leur rôle sémantique. À titre d'exemple, les UTP  $\langle n \rangle$  (أنشأ) (créer) et  $\langle in \rangle A'$  (إنشاء) (création), *créer* et *création*, *create* et *creation* ont la structure Agent ~ Patient.

Elles appartiennent au cadre *Intentionally\_create* (The Creator creates a new entity, the Created\_entity), qui a deux core-FE, Creator et Created\_entity.

#### *Correspondance semi-parfaite*

C'est la situation qui se produit lorsqu'une UTP dans le DiCoInfo a deux entrées ou plus et c'est l'une de ces entrées qui correspond au cadre. Par exemple, les UTP *iqLAE<sub>1a</sub>* (إقلاع) (démarrage), *start<sub>1a</sub>* et *démarrer<sub>1a</sub>* correspondent au cadre *Being\_in\_operation* (A Device or machine is in (or out of) service.). Il faut noter que *démarrer* fait l'objet de deux entrées dans les différentes versions de DiCoInfo (ASM, anglais et français). L'entrée qui appartient au cadre en question est *démarrer<sub>1a</sub>*, alors que *démarrer<sub>1b</sub>* fait partie d'un autre cadre, *Change\_operational\_state*, puisqu'elle n'a pas la même structure actancielle (Agent ~ Patient).

#### *Correspondance partielle*

Dans certaines situations, la correspondance entre les données dans les deux ressources n'est pas tout à fait parfaite, en ce qui concerne le nombre d'actants par rapport aux nombre de core-FE des cadres. À titre d'exemple, le cadre *Sending* (A Sender plans the Path (along with Source and Goal) of a Theme and places it in circumstances such that it travels along this Path under the power of some entity other than the Sender) comporte cinq FE, à savoir Goal, Recipient, Sender, Theme et Transport\_means. Les UTP correspondantes sont *Arsala<sub>1</sub>* (ارسال) (envoyer) et *irsAl<sub>1b</sub>* (إرسال) (envoi) et *send<sub>1</sub>*. Ces unités ont trois actants : Agent, Patient et Destination.

#### *Nouveau cadre*

Lorsqu'aucun des cadres dans FrameNet ne correspond aux termes dans le DiCoInfo, nous créons nos propres cadres. Par exemple, nous avons trouvé que les UTP *laSq* (لصق) (collage), *paste* et *coller et collage* ne correspondent à aucun cadre. Dans ce cas, nous avons créé un nouveau cadre qui correspond à ces unités, *Moving\_text\_into*.

### 5.6.3 Critères de sélection des cadres

Après avoir analysé les correspondances entre les données des deux ressources, nous procédons à la sélection des cadres. Pour cela, nous prenons en compte les dix critères établis dans Ruppenhofer et al. (2010 : 9-13). Il convient de signaler que nous n'avons pas pris en compte tous les critères, et ce pour les raisons suivantes. Premièrement, puisque nous avons affaire à une langue de spécialité, notre sélection de critères est basée sur les particularités des UTP en informatique. Deuxièmement, étant donné que notre travail s'appuie sur FrameNet et la théorie de la Sémantique des cadres, nous avons adapté certains aspects de ces critères à nos besoins. De plus, puisque notre méthodologie de découverte des cadres est fondée sur la comparaison entre les données de FrameNet et de DiCoInfo, nous avons trouvé nécessaire d'ajouter un premier critère qui n'est pas dans FrameNet. Troisièmement, nous avons écarté certains critères présentés par les auteurs. Ces critères semblaient reprendre certaines considérations couvertes par d'autres critères. Finalement, certains critères ne s'appliquent pas à notre recherche. Dans la présentation des critères, nous commençons par donner une petite citation des auteurs servant à le présenter, puis nous en discutons.

#### *Critère 1 : correspondance entre FrameNet et DiCoInfo*

Ce critère concerne la correspondance entre les cadres du FrameNet et les UTP dans le DiCoInfo. Pour qu'un cadre soit considéré comme candidat, celui-ci doit avoir une définition sémantiquement proche d'une ou d'un ensemble d'UTP qui lui appartiendraient. Nous avons constaté qu'une UTP peut appartenir à plusieurs cadres dans FrameNet. Pour contourner ce problème, nous analysons chaque cadre, en terme de définition, et nous décidons lequel peut être, potentiellement, le plus approprié. À titre d'exemple, l'UTV *vab~ata* (ثبت) (installer) dans le DiCoInfo correspond à deux cadres dans FrameNet, *Installing* et *Change\_of\_leadership*. Le cadre le plus approprié est celui de *Installing*, puisque sa définition comprend le sens le plus proche de l'UTV : « An Agent places a Component in a Fixed\_location so that the Component is attached and interconnected and thereby functional » (FrameNet 2016), contre « un utilisateur place un logiciel sur un ordinateur et définit les paramètres pour que le logiciel puisse bien fonctionner » (DiCoInfo 2015).



### *Critère 2 : accord en type et en nombre d'actants*

« All LUs in a frame must have the same number and types of frame elements in both explicit and implicit (NI) contexts » (Ruppenhofer et al. 2010 : 9).

Les UTP qui appartiennent au même cadre doivent avoir des actants qui concordent en nombre et en type. En ce qui concerne le nombre, les actants (ici dans le sens des core-FE) des UTP doivent être de même nombre. Si une UTP a plus d'actants que les autres UTP, alors elle doit être mise dans un autre cadre. De plus, puisque certains cadres renferment des UTP anglaises et françaises, ce critère s'applique aux UTP dans les trois langues, dans le sens où si une UTP d'une langue se distingue par un nombre d'actants plus élevé que les autres langues, nous plaçons cette UTP dans un autre cadre. En ce qui a trait au type, toutes les UTP d'un même cadre doivent avoir le même type d'actants. Par exemple, si la structure actancielle d'une unité fait appel à un Agent et un Patient, il faut que toutes les autres unités aient le même type de structure. Par exemple, si nous admettons que *Had~ava* (حدث) (mettre à jour) et *taHdyv* (تحديث) (mise à jour) font partie du même cadre, il faut que ces deux unités aient la structure Agent ~ Patient, où le premier actant est un Agent et le deuxième est un Patient.

### *Critère 3 : Étapes et transition*

« In aspectually complex frames, the lexical units should all entail the same set of stages and transitions » (Ruppenhofer et al. 2010 : 10).

Les UTP d'un même cadre doivent toutes impliquer les étapes et les transitions dénotées par les UTP. Par exemple, pour des UTV comme *Taw~ara* (طور) (développer) et *barmajap* (برمجة) (programmer), nous constatons une différence par rapport au processus que ces deux unités dénotent. L'unité *Taw~ara* (طور) (développer) désigne un processus selon lequel un logiciel est conçu à partir de composants logiciels. Cela implique l'Agent en charge de processus et le produit. En ce qui concerne *barmajap* (برمجة) (programmer), l'unité dénote également un processus de création de logiciels et d'applications. La différence qui existe entre cette unité et la précédente est le fait qu'elle fait appel à un Matériau utilisé afin de

concevoir le produit informatique. Par conséquent, dans le cas de ces deux unités, nous les mettons dans deux cadres différents.

#### *Critère 4 : Perspective*

« The same frame elements will be profiled across all lexical units of a frame » (Ruppenhofer et al. 2010 : 10).

Pour un cadre donné, les actants doivent refléter la même perspective. Par exemple, l'UTN *iqlAE* (إقلاع) (démarrage) exprime le fait qu'un composant informatique (typiquement un ordinateur) commence à fonctionner. Ici, une distinction doit être faite entre *iqlAE<sub>1a</sub>* (إقلاع) (démarrer<sub>1a</sub>) qui dénote le système qui commence à fonctionner, et *iqlAE<sub>1b</sub>* (إقلاع) (démarrer<sub>1b</sub>) qui dénote la personne qui prend en charge le processus du démarrage du système. Bien entendu, ces deux unités disposent de deux structures actanciennes différentes, ce qui ajoute une autre distinction.

#### *Critère 5 : Relations entre actants*

« We want the interrelations between frame elements to be the same for all the LUs in a frame » (Ruppenhofer et al. 2010 : 11).

Dans un cadre, les relations entre les actants doivent être identiques. Par exemple, l'actant Patient dans *An\$A* (أنشأ) (créer) et *TabaEa* (طبع) (imprimer) expriment deux réalités différentes. Dans le cas de *An\$A* (أنشأ) (créer), le Patient désigne l'entité créée. Dans le cas de *TabaEa* (طبع) (imprimer), le Patient désigne l'entité reproduite par l'Agent ou l'Instrument. Considérons les deux exemples suivants :

[1] *nun\$} mural~ad nasal~yh funwn wa naDaE fyhi kAf~ap AlbarAmij Alfaniy~ap*  
ننشئ مجلد نسيمه فنون و نضع فيه كافة البرامج الفنية  
Nous créons un fichier et nous l'appelons Art et nous y mettons tous les programmes artistiques

[2] *taxtalif TAbiEAt All~yzar Ean gayrihA fy An~hA taTbaE AlSafHap kAmilap wa lay saTr I El| AlAqal*  
تختلف طابعات الليز عن غيرها في أنها تطبع الصفحة كاملة و ليس سطرا 1 على الأقل  
Les imprimantes laser se distinguent des autres imprimantes par le fait qu'elles impriment la page au complet et non pas une ligne.

#### *Critère 6 : La scène évoquée*

« The presuppositions, expectations, and concomitants of the targets within a frame will be shared » (Ruppenhofer et al. 2010 : 11).

Toutes les UTP faisant partie d'un même cadre doivent évoquer la même situation. Ce critère est présenté différemment par les auteurs. Nous l'adaptions ici à nos données. Dans un cadre, si une UTP comme *Ham~ala* (حمل) (télécharger) évoque la scène où un composant informatique est déplacé à partir d'un ordinateur sur un autre ordinateur, les autres UTP du même cadre, comme *naqala* (نقل) (transférer<sub>1</sub>) et *taHmyl* (تحميل) (téléchargement<sub>1</sub>), doivent évoquer la même scène.

#### *Critère 7 : Regroupement des constructions générales de la langue*

« At all times, we lump together those groups whose semantic differences are due to general constructions of the language » (Ruppenhofer et al. 2010 : 12).

Par exemple, nous mettons *An\$A* (أنشأ) (il a créé) et *Aun\$IA* (أنشئ) (il a été créé) dans le même cadre.

#### *Critère 8 : Regroupement des antonymes*

« We group together words that are antonyms of each other » (Ruppenhofer et al. 2010 : 12).

Pour certaines UTP, nous regroupons des antonymes dans un même cadre. Par exemple, des antonymes comme *fataHa* (فتح) (ouvrir<sub>1</sub>) et *Aglaqa* (أغلق) (fermer<sub>1</sub>).

### **5.6.4 Relations entre cadres**

Nous avons utilisé les mêmes relations qui existent dans FrameNet. Il importe dans notre travail de relier les cadres découverts qui sont sémantiquement proches. Ce regroupement peut révéler certains aspects du domaine de spécialité, comme les processus et les actions. De plus, nous souhaitons créer des scénarios conceptuels. Ces scénarios ont pour but la structuration des cadres qui sont susceptibles de rendre compte d'un ensemble d'activités au sein du domaine de l'informatique.

Nous avons un total de sept relations que nous détaillons dans le tableau suivant.

Nom de cadre	Définition	Exemple
<b>Inheritance</b>	Cette relation correspond à « is-a relation », comme dans les ontologies. Elle relie deux cadres, un cadre Parent et un cadre Enfant. Le cadre Enfant hérite des propriétés sémantiques du cadre Parent, en ce qui concerne les FE, leur type et les relations qu'ils partagent. Inheritance comprend deux relations inverses : Is inherited by et Inherits from.	Activity_start Is inherited by Intentionally_create
<b>See_also</b>	Cette relation est utilisée lorsque les cadres sont très similaires et ils doivent être examinés et comparés.	Hardware_setup See_also Installing
<b>Subframe</b>	Certains cadres sont considérés comme complexes, car ils désignent une séquence d'états et de transitions. Ces derniers sont des sous-cadres du cadre général et ils lui sont reliés via une relation de Subframe. Subframe comprend deux relations inverses : Is subframe et Has subframe.	Computer_setup Has subframe hardware_setup, Installing et Placing
<b>Uses</b>	Cette relation est utilisée si une partie d'un cadre réfère à un autre cadre (plus général).	Information_display Uses Transmitting
<b>Perspectivized in</b>	Cette relation renferme deux points de vue d'un cadre plus général (ou neutre). La relation inverse est Is perspectivized on.	Display_process Is perspectivized on Information_display Duplication_procees Is perspectivized in Duplication
<b>Precedes</b>	Cette relation spécifie la séquence des états ou des événements qui définissent une situation donnée. La relation inverse est Preceded by.	Placing et Removing
<b>Is causative of</b>	Cette relation désigne le lien que deux cadres peuvent avoir si l'un est la cause de l'autre. La relation inverse est celle de Inchoative.	Cause_modification Is causative of Enhancing

Tableau 14 : Relations entre cadres utilisés

Une relation que nous n'avons pas mise dans le tableau est celle de Is opposed to. Cette relation ne fait pas partie des relations utilisées dans FrameNet, mais elle est proposée dans DiCoEnviro. La relation introduit une opposition entre deux cadres. Par exemple, le cadre Placing entretient une relation d'opposition avec le cadre Removing, Placing Is opposed to Removing.

Pour donner un exemple concret, nous analysons le cadre Transfer. Ce cadre renferme les UTP suivantes : *naqil* (نقل) (transfert), *Ham~ala* (حمل) (télécharger) et *taHmyl* (تحميل) (téléchargement), *download<sub>1</sub>*, *download<sub>1,1</sub>* et *upload<sub>1</sub>*, *télécharger<sub>1</sub>* et *téléchargement<sub>1</sub>*. Ce cadre est relié à deux cadres, le cadre Receiving et le cadre Transfer\_scenario. Avec

Receiving, la relation constatée est celle de Perspectivized in, tandis qu’avec Transfer\_scenario, la relation est Is Subframe of. Transfer et Receiving sont reliés par cette relation, car nous estimons que Receiving exprime une perspective du processus de transfert. En ce qui concerne la relation Is Subframe of, le cadre Transfer est relié à un cadre plus général et plus abstrait. Ce cadre est non lexical, puisqu’il ne renferme pas d’UT et il sert uniquement à relier sémantiquement des cadres.

### 5.6.5 Encodage des cadres

Une fois l’identification des cadres réalisée, nous avons procédé à l’encodage des renseignements portant sur eux. Tout comme les contextes, les cadres sont encodés au moyen du même éditeur XML. La structure d’encodage des cadres comprend les éléments suivants : le cadre, les core-FE, les actants, les lexies, la définition, les exemples, une note et les relations (figure 84).

```

<Frame identificateur="" note="" statut="" mise-a-jour="">
  <Core-FEs></Core-FEs>
  <Actants></Actants>
  <Lexies></Lexies>
  <Définitions-DiCoInfo>
    <Définition></Définition>
    <Exemple></Exemple>
    <Note langue=""></Note>
  </Définitions-DiCoInfo>
  <Relations></Relations>
</Frame>

```

Figure 84 : Structure d’encodage pour les cadres

Nous rappelons que, puisque notre méthodologie de repérage des cadres se base sur les cadres existants dans FrameNet, nous avons ajouté une section (core-FE) afin de comparer les core-FE de FrameNet aux actants de DiCoInfo. Bien entendu, pour les cadres que nous créons, cette section est laissée vide.

#### 5.6.5.1 Frame

La première balise est <Frame> regroupe toutes les autres balises dans le fichier XML dédié aux cadres. Cette balise comprend les attributs suivants : `identificateur=`, `note=`, `statut=`

et **mise-a-jour=**. Le premier donne le nom du cadre. Le deuxième attribut précise si le cadre en question est pris du FrameNet ou il s'agit d'un nouveau cadre. En effet, cet attribut peut accueillir jusqu'à trois mentions, selon le cadre analysé, à savoir "**nombre d'actants différent**", "**FN tel quel**" et "**nouveau**". Tout comme dans l'annotation des contextes discutée plus haut, l'attribut **statut=** indique l'état d'avancement de la rédaction et **mise-a-jour=** indique la date à laquelle la fiche a été mise à jour.

### 5.6.5.2 Core Frame Element (FE)

Dans la balise `<Core-FEs>`, nous inscrivons les core-FE dans FrameNet du cadre sélectionné. Dans cette balise on énumère les core-FE dans les sous-balises. Donc, un cadre comme `Intentionally_create` peut avoir les sous-balises : `<Core-FE no="1">Creator</Core-FE>` et `<Core-FE no="2">Created_entity</Core-FE>`. Les chiffres marqués dans les balises servent dans la comparaison avec les actants du DiCoInfo. Quand le cadre est inspiré de FrameNet, nous montrons à quels actants correspondent les core-FE. Au début de cette section, nous avons mentionné qu'une page à des fins de comparaison a été créée. Après avoir fait la correspondance entre les actants et les core-FE, les résultats s'affichent sur cette page (figure 85).


DiCoInfo	FrameNet																				
<p><a href="#">create_1</a> (vt)</p> <p>SA: create: <b>Agent</b>(<i>user 1</i>) ~ <b>Patient</b>(<i>file 1, site 1</i>)</p> <p>DEF: A USER causes a FILE or a SITE to exist by applying rules to allow a computer or a program to recognize the FILE or the SITE.</p> <p><b>Intentionally_create</b> (FN tel quel) :</p> <p>Definition(s):</p> <p>: Agent (user) causes that <b>Patient</b> (file) exists</p> <table border="1"> <thead> <tr> <th>DiCoEnviro</th> <th>FrameNet</th> </tr> </thead> <tbody> <tr> <td><b>Agent</b></td> <td>↔ Creator</td> </tr> <tr> <td><b>Patient</b></td> <td>↔ Created_entity</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>ar</th> <th>en</th> <th>fr</th> </tr> </thead> <tbody> <tr> <td><a href="#">الإنشاء</a>   <a href="#">الإنشاء</a>   <a href="#">الإنشاء</a></td> <td><a href="#">create_1</a>   <a href="#">creation_1</a></td> <td><a href="#">création_1</a>   <a href="#">créer_1</a></td> </tr> </tbody> </table> <p><a href="#">20 contextes...</a></p>	DiCoEnviro	FrameNet	<b>Agent</b>	↔ Creator	<b>Patient</b>	↔ Created_entity	ar	en	fr	<a href="#">الإنشاء</a>   <a href="#">الإنشاء</a>   <a href="#">الإنشاء</a>	<a href="#">create_1</a>   <a href="#">creation_1</a>	<a href="#">création_1</a>   <a href="#">créer_1</a>	<p><a href="#">create</a> (V) <span style="float: right;"><a href="#">ID: 5437</a></span></p> <p>COD: bring into existence.</p> <p><b>Intentionally_create</b>: The <b>Creator</b> creates a new entity, the <b>Created_entity</b>, possibly out of <b>Components</b>.</p> <p><a href="#">Annotated Contexts</a></p> <table border="1"> <thead> <tr> <th>Type</th> <th>FE</th> </tr> </thead> <tbody> <tr> <td>Core</td> <td><b>Created_entity</b> <b>Creator</b></td> </tr> <tr> <td>Extra-Thematic</td> <td><b>Components</b> <b>Co_participant</b> <b>Depictive</b> <b>role</b></td> </tr> <tr> <td>Peripheral</td> <td><b>Instrument</b> <b>Manner</b> <b>Means</b> <b>Place</b> <b>Purpose</b> <b>Time</b></td> </tr> </tbody> </table> 	Type	FE	Core	<b>Created_entity</b> <b>Creator</b>	Extra-Thematic	<b>Components</b> <b>Co_participant</b> <b>Depictive</b> <b>role</b>	Peripheral	<b>Instrument</b> <b>Manner</b> <b>Means</b> <b>Place</b> <b>Purpose</b> <b>Time</b>
DiCoEnviro	FrameNet																				
<b>Agent</b>	↔ Creator																				
<b>Patient</b>	↔ Created_entity																				
ar	en	fr																			
<a href="#">الإنشاء</a>   <a href="#">الإنشاء</a>   <a href="#">الإنشاء</a>	<a href="#">create_1</a>   <a href="#">creation_1</a>	<a href="#">création_1</a>   <a href="#">créer_1</a>																			
Type	FE																				
Core	<b>Created_entity</b> <b>Creator</b>																				
Extra-Thematic	<b>Components</b> <b>Co_participant</b> <b>Depictive</b> <b>role</b>																				
Peripheral	<b>Instrument</b> <b>Manner</b> <b>Means</b> <b>Place</b> <b>Purpose</b> <b>Time</b>																				

Figure 85 : Correspondance entre actants et core-FE sur la page de comparaison

### 5.6.5.3 Actants

Dans cette partie, la balise `<Actants>` introduit les actants qui correspondent aux core-FE énumérés dans la balise précédente. Comme dans la balise `<Core-FEs>`, les actants des UTP sont énumérés dans cette balise. Par exemple, les UTP *EAlaja* (عالج) (traiter) et *muEAlajap* (معالجة) (traitement) ont les actants Agent ~ Patient :

```
<Actants>
  <Actantno="1">Agent</Actant>
  <Actantno="2">Patient</Actant>
</Actants>
```

Le cadre correspondant est `Processing_materials`. Dans ce cas-ci, la correspondance constatée est partielle (nombre d'actants différents), les deux balises prennent la forme suivante :

```
<Core-FEs>
  <Core-FEno="1">Agent</Core-FE>
  <Core-FEno="2">Alterant</Core-FE>
  <Core-FEno="3">Material</Core-FE>
</Core-FEs>
<Actants>
  <Actantno="1">Agent</Actant>
  <Actantno="3">Patient</Actant>
</Actants>
```

### 5.6.5.4 Lexies

Dans la balise `<Lexies>`, nous inscrivons les UTP faisant partie du cadre. Les sous-balises qui découlent de la balise `<Lexies>` englobent quatre attributs. Le premier attribut définit le dictionnaire source. Bien entendu, dans le cas de notre travail, il s'agit du DiCoInfo. Le deuxième attribut définit la langue, ASM (dans l'attribut, « arabe » est inscrit), anglais et français. Le troisième attribut (`identificateur=`) identifie l'UTP en question tel qu'elle est présentée dans l'attribut `xml:id=`, déjà expliqué dans la section 5.5.2. De cette façon, on établit un lien avec l'entrée du terme dans le DiCoInfo. Enfin, le quatrième attribut précise le numéro d'acception de l'UTP. Dans ce qui suit, nous donnons un exemple du schéma du cadre `Processing_materials`.

```

<Lexies>
<Lexieidentificateur="_traiter_1_عالج_"dico="dicoinfo"langue="ar"no_acception="1"/>
<Lexieidentificateur="_traitement_1_معالجة_"dico="dicoinfo"langue="ar"no_acception="1"/>
<Lexieidentificateur="process"dico="dicoinfo"langue="en"no_acception="1"/>
<Lexieidentificateur="traiter"dico="dicoinfo"langue="fr"no_acception="1"/>
<Lexieidentificateur="traitement"dico="dicoinfo"langue="fr"no_acception="1"/>
</Lexies>

```

### 5.6.5.5 Définition

Dans cette balise, nous donnons une définition du cadre. Pour notre travail, nous donnons les définitions des cadres en langue anglaise<sup>56</sup>. La définition comprend deux éléments essentiels, les actants et les termes typiques, comme dans la définition du cadre Processing\_materials : <Définition langue="en">Agent (system or program) manipulates Patient (data) in order to modify them</Définition>.

### 5.6.5.6 Exemple

Après avoir défini le cadre, ses éléments et ses UTP, nous donnons trois exemples annotés, un exemple pour chaque langue. Les éléments annotés dans les exemples sont l'UTP et ses actants (figure 86).

```

<Exemple langue="ar"><Actant-cont nom="Agent">يحتوي كرت الصوت على</Actant-cont><lexie-att>معالجة</lexie-att>
<Actant-cont nom="Patient">الصوت</Actant-cont>او الحاسب</Exemple>
<Exemple langue="en">
<Actant-cont nom="Agent">A computer</Actant-cont>is an equipment used to analyze and<lexie-att>PROCESS</lexie-att>
<Actant-cont nom="Patient">data</Actant-cont>and display the results on any medium it is capable of writing on</Exemple>
<Exemple langue="fr">En général, plus<Actant-cont nom="Agent">un microprocesseur</Actant-cont>peut<lexie-att>traiter</lexie-att>
<Actant-cont nom="Patient">de bits</Actant-cont>, plus l'ordinateur est puissant et rapide.</Exemple>

```

Figure 86 : Encodage d'exemple pour le cadre Processing\_materials

<sup>56</sup> Pour la partie de notre travail qui concerne les cadres, la langue que nous utilisons est l'anglais, puisque cette partie comprend trois langues.



### 5.6.5.7 Relations

La dernière balise `<Relations>` est dédiée aux relations qu'entretient le cadre précédemment défini avec d'autres cadres. La structure de cette balise est construite comme suit :

```
<Relations>  
<Relationnom="Uses">Cause_change</Relation>  
</Relations>
```

## **6. Résultats**

Dans ce chapitre, nous discutons les résultats obtenus à la suite de l'application de la méthodologie décrite dans le chapitre précédent. Nous rappelons que notre méthodologie concerne le repérage et la validation des UTP verbales et nominales du domaine de l'informatique. En outre, nous proposons la compilation de ces UTP et des informations les accompagnant dans une ressource terminologique, qui formera, à terme, la version en ASM du DiCoInfo. Nous nous concentrons donc sur trois points : les résultats obtenus lors de l'extraction et de la validation des UTP (section 6.1), les particularités de la version en ASM du DiCoInfo (section 6.2), les cadres sémantiques obtenus (section 6.3) et la ressource dans laquelle sont compilés les cadres découverts (section 6.4).

### **6.1 Les unités terminologiques prédicatives (UTP) de l'ASM**

Dans cette sous-section, nous nous attardons sur les UTP obtenues lors de l'extraction. Nous commençons d'abord par analyser nos résultats pour les UTV. Par la suite, nous discutons des résultats obtenus pour les UTN. Nous faisons une comparaison entre les deux types d'unités et nous les relierons – lorsque c'est possible – à leurs équivalents français et anglais dans le DiCoInfo.

#### **6.1.1 Les UTV**

Comme montré dans la section 5.2.4.10, l'extraction des UTV de l'ASM au moyen de TermoStat constituait un vrai défi. Nous avons évoqué les difficultés techniques que nous avons à surmonter dans le traitement des UTV candidates et le traitement des documents de notre corpus. En outre, nous avons montré le très faible nombre d'unités validées ayant une spécificité positive et expliqué les raisons pour lesquelles nous devons prendre en considération les unités à spécificité négative et banale (35 unités à spécificité négative, 5 unités à spécificité banale et 18 unités à spécificité positive).

De plus, nous avons abordé dans la section 5.2.4.10 les raisons pour lesquelles de tels résultats sont obtenus. Nous avons expliqué le grand nombre d'erreurs d'étiquetage entraînés par l'ambiguïté morphologique des verbes et le conflit constaté entre AraComLex (l'analyseur

morphologique) et Stanford (l'étiqueteur). Nous avons également évoqué la spécificité sémantique des verbes, et comment cette spécificité constituait un obstacle important lors de la reconnaissance de ces unités par le système d'extraction.

Pour ces raisons, un grand nombre d'unités ont été écartées et nous n'avons retenu que celles que nous avons jugées être relatives au domaine de l'informatique. Le sens terminologique de ces unités a été validé par les critères discutés dans la section 5.3. Après cette étape de validation terminologique, nous avons procédé à la distinction des sens spécialisés des unités obtenues. Cette étape est particulièrement importante dans notre travail, car elle nous permet de vérifier les unités qui sont susceptibles d'avoir plusieurs sens spécialisés, ou qui renvoient à plusieurs réalités au sein de domaine. Au total, à partir des 45 unités validées (voir tableau 11), nous avons obtenu 47 UTV spécifiques au domaine de l'informatique.

Dans les sections qui suivent, nous donnons plus de détails sur les unités validées et nous montrons les particularités que nous avons constatées.

### 6.1.1.1 Les UTV validées

Comme mentionné dans la section 6.1.1, nous avons obtenu un total de 47 unités à la suite de la validation terminologique et de la distinction sémantique. Ces unités correspondent, pour la plupart, aux critères de validation discutés dans la section 5.3. Cependant, nous avons observé quelques cas particuliers, que nous discutons dans la présente section.

#### *Critère 1 : Appartenance au domaine*

En qui concerne le premier critère, tous les verbes retenus ont été validés en l'appliquant. Pour certains verbes, leur appartenance au domaine de l'informatique allait de soi. Nous mentionnons des verbes comme *naqara* (نقر) (cliquer) et *TabaEa* (طبع) (imprimer).

Pour d'autres verbes, il fallait vérifier une grande quantité de contextes pour nous assurer de leur appartenance au domaine. Ces verbes sont très polysémiques et ils sont morphologiquement très ambigus. Par exemple, nous avons des verbes comme *tawaq~afa* (توقف) (s'arrêter), *taEar~afa* (تعرف) (reconnaître), *qarA* (قرأ) (lire) et *Ham~ala* (حمل)

(télécharger). L'appartenance au domaine de ce dernier verbe n'est pas contestée. Cependant, sa forme est ambiguë. Dans la section 2.3, nous avons montré un cas similaire (le verbe *rafaEa* (رفع) (téléverser)). Pour le verbe en question, plusieurs analyses sont proposées par l'analyseur morphologique :

- Ham~ala (حمل) (télécharger)
- Hamala (حمل) (porter)
- Himl (حمل) (charge)

Nous rappelons que l'absence de diacritiques (voyelles courtes) augmente le taux d'ambiguïté morphologique pour une langue comme l'ASM.

#### *Critère 2 : La nature des actants*

En ce qui concerne le deuxième critère, la nature des termes constituant les actants de l'unité nous servait à valider le statut terminologique. Nous avons des cas typiques comme *daEama* (دعم) (accepter), *vab~ata* (ثبت) (installer), *xaz~ana* (خزن) (stocker) et *HafiZa* (حفظ) (sauvegarder). Considérons les exemples suivants pour les verbes *daEama* (دعم) (accepter) et *EAlaja* (عالج) (traiter).

[1] *yadEam AlhAtif Al\*aky |yfwN[Patient] wa alHAsib AllawHy |ybad [Patient] miEyAr Airbrint[Destination]*

يدعم الهاتف الذكي آيفون و الحاسب اللوحي آيباد معيار إيربرنت

*Le téléphone intelligent iPhone et la tablette iPad **acceptent** le module AirPrint.*

[2] *Anta taErif musbaqaN An AlHaswb[Agent] yEAlij AlArqAm[Patient] bidiq~a*

أنت تعرف مسبقاً أن الحاسوب يعالج الأرقام بدقة

*Tu sais déjà que l'ordinateur **traite** les chiffres avec précision.*

Dans les contextes présentés, nous remarquons que les UTV admettent des actants qui sont eux-mêmes des termes. Lorsque le critère 2 est combiné avec le critère 1, ces verbes véhiculent un sens spécialisé relatif au domaine de l'informatique. Toutes les UTV sans

exception correspondent à ce critère. Cependant, ce critère doit être toujours combiné au critère 1, puisque certains verbes admettent des termes dans leur structure actancielle mais ils ne véhiculent pas de sens spécialisé. Par exemple, le verbe *Aistaxdama* (استخدم) (utiliser) était proposé par TermoStat comme candidat. Ce verbe existe dans le corpus sous différentes formes, comme celle présentée dans l'exemple suivant :

[3] *i\*A kAn AlmAsiH alDaw}y [Ageni] yastavdim Taryqap Al Single Pass [Patiem], fasykwn ladyna valAv Sufwf wa lys wahid*

إذا كان الماسح الضوئي يستخدم طريقه ال Single Pass فسيكون لدينا ثلاث صفوف وليس صف واحد

*Si le numériseur utilise la méthode Single Pass, on aurait trois tableaux, et non pas un.*

Dans cet exemple, le verbe *Aistaxdama* (استخدم) (utiliser) admet les termes *masiH Daw}y* (ماسح ضوئي) (numériseur) et Single Pass. Toutefois, le sens du verbe n'est pas spécialisé puisqu'il ne dénote pas une activité propre au domaine et les autres critères de validation ne s'appliquent pas.

### *Critère 3 : La parenté morphologique*

Le troisième critère, la parenté morphologique, s'est révélé particulièrement utile pour notre travail. Toutes les UTV correspondent à ce critère. Nous avons montré, dans la section 5.2.4.11, les phénomènes qui affectent les verbes dans le discours spécialisé du point de vue de l'extraction automatique. Nous avons également montré comment l'ambiguïté de la forme verbale et de la construction syntaxique issue de la traduction contribuent à faire en sorte que la forme nominale soit favorisée. Par conséquent, tous les verbes retenus ont une parenté morphologique avec une forme nominale. Certains cas idéaux sont observés dans le corpus. Nous mentionnons des UTV comme :

[1] *TabaEa* (طبع) (imprimer) : *TibAEap* (طباعة) (impression), *maTbuEap* (مطبوعة) (imprimé);

[2] *nasaxa* (نسخ) (copier) : *nasx* (نسخ) (copie);

[3] *Ham~ala* (حمل) (télécharger) : *taHmyl* (تحميل) (téléchargement); et

[4] *Had~ava* (حدث) (mettre à jour) : *taHdyv* (تحديث) (mise à jour).

Toutefois, certaines formes verbales sont dominantes, dans le sens où ces verbes sont plus fréquents que leurs formes nominales. Par exemple, les verbes *naqara* (نقر) (cliquer), *DagaTa* (ضغط) (appuyer), *Eamila* (عمل) (tourner) et *daEama* (دعم) (accepter) sont plus fréquents que leur forme nominale *daEm* (دعم) (support), *DagT* (ضغط) (appui), *Eamal* (عمل) (le fait de tourner) et *naqr* (نقر) (clic). Dans le tableau 15, nous présentons la fréquence de chacune de ces unités dans le corpus.

UTV	Fréquence	UTN	Fréquence
<i>naqara</i> (نقر) (cliquer)	435	<i>naqr</i> (نقر) (clic)	X
<i>DagaTa</i> (ضغط) (appuyer)	523	<i>DagT</i> (ضغط) (appui)	352
<i>daEama</i> (دعم) (accepter)	378	<i>daEm</i> (دعم) (support)	92
<i>Eamila</i> (عمل) (tourner)	1480	<i>Eamal</i> (عمل) (le fait de tourner)	669

Tableau 15 : Comparaison entre les UTV les plus fréquentes et leurs dérivés nominaux

#### *Critère 4 : Les liens paradigmatisques*

Le quatrième critère, celui du lien paradigmatisque, est le critère le plus problématique. Le nombre d'UTV validées au moyen de ce critère est très limité. Nous avons quelques cas d'antonymie et de quasi-synonymie et le reste des verbes ne partagent aucun lien paradigmatisque avec d'autres. Dans le tableau 16, nous présentons les UTV qui partagent des liens paradigmatisques avec d'autres UTV.

UTV	Type de lien	UTV correspondant
fataHa (فتح) (ouvrir)	Antonymie	Aglaqa (أغلق) (fermer)
Arsala (أرسل) (envoyer)	Antonymie	Aistaqbala (استقبل) (recevoir)
AqlaEa (أقلع) (démarrer)	Antonymie	tawaqafa (توقف) (arrêter)
HaZafa (حذف) (supprimer)	Antonymie	vab~ata (ثبت) (installer)
naqara (نقر) (cliquer)	Quasi-synonymie	DagaTa (ضغط) (appuyer)
AstajAb (استجاب) (répondre)	Quasi-synonymie	taEar~afa (تعرف) (reconnaître)

Tableau 16 : Liens paradigmatiques observés entre les UTV

Les autres verbes sont, toutefois, terminologiques puisqu'ils ont été validés au moyen des autres critères. Bien entendu l'absence de certaines relations paradigmatiques peut s'expliquer par le fait que notre corpus n'atteste pas de certaines unités susceptibles de partager ce type de liens. Pour des unités comme *nasaxa* (نسخ) (copier), un antonyme existe, *qaS~a* (قص) (couper), sauf que cette unité n'est pas proposée dans la liste des candidats à cause de sa faible fréquence dans le corpus. L'autre explication possible est que certaines unités dans le domaine de l'informatique ne partagent pas de relations paradigmatiques avec d'autres termes. Par exemple, l'UTV *EaraDa* (عرض) (afficher) en informatique n'a probablement pas d'antonyme admis comme terme, comme *xab~A* (خبأ) (cacher). De même, l'UTV *AdAra* (أدار) (gérer) n'a probablement pas de synonyme<sup>57</sup>.

### 6.1.2 Les UTN

En ce qui concerne les UTN, à partir des 45 unités validées, nous en avons obtenu 59. Une partie de ces unités correspond aux formes nominales des UTV déjà retenues. Dans la section 6.1.4, nous faisons une comparaison entre ces unités, tout en mettant en lumière leurs particularités. En ce qui concerne les critères que nous avons appliqués pour la validation terminologique, d'après nos observations, nous avons constaté que les unités retenues

<sup>57</sup> Nous signalons que ces deux unités ont un antonyme et un synonyme en français, *cacher* et *administrer*.

correspondent à ces critères. Nous n'avons pas constaté de cas particuliers. Dans ce qui suit, nous donnons une description complète de l'application que nous avons faite de ces critères.

### *Critère 1*

Toutes les unités ont été validées au moyen du critère 1. Tout comme les UTV, l'appartenance au domaine de l'informatique de certaines UTN allait de soi. Nous mentionnons, à titre, d'exemple, *AiqlAE* (إقلاع) (démarrage), *barmajap* (برمجة) (programmation), *tavbyt* (تنصيب) (installation) et *taHdyv* (تحديث) (mise à jour). Nous n'avons pas rencontré de difficultés à valider le statut terminologique de ces unités. Une unité comme *AiqlAE* (إقلاع) (démarrage) dans le domaine de l'informatique ne désigne que le processus de démarrage d'un composant informatique. Dans la langue générale, le nom signifie le fait d'arrêter de fumer et dans certains contextes il signifie le fait d'amorcer un processus (par exemple, le fait de démarrer une voiture).

Par contre, pour d'autres unités, le processus de validation a soulevé des difficultés. Il fallait vérifier une grande quantité de contextes afin de vérifier le sens et le distinguer d'autres sens qu'une unité pouvait avoir. Parmi les unités les plus problématiques, nous mentionnons *ta\$gyl* (تشغيل) (actionnement ou amorçage), *AinhA'* (إنهاء) (mise à fin, dans le sens de *terminer*), *kitabap* (كتابة) (écriture) et *wuSwl* (وصول) (accès). Il semble que ces unités soient affectées par le phénomène de la terminologisation, dans le sens où ces unités sont employées dans un discours non spécialisé, mais, en informatique, elles ont acquis un sens propre au domaine.

### *Critère 2*

Toutes les UTN sans exception ont été validées par ce critère. Des cas typiques peuvent être observés dans le corpus comme *DagT* (ضغط) (appui), *ta\$gyl* (تشغيل) (actionnement), *AiqlAE* (إقلاع) (démarrage), *muEAlajap* (معالجة) (traitement) et *barmajap* (برمجة) (programmation). Ces cas sont présentés dans les exemples suivants.



- [1] *AlbiTAqap tuEad~u maksabaN kabyraN liman*<sub>[Agent]</sub> *yuHib~ muEAlajap EurwD Alfidyw*  
<sub>[Patient]</sub> *lijihAzih*  
 البطاقة تعد مكسبًا كبيرًا لمن يحب معالجة عروض الفيديو جهازه  
*La carte est très utile pour ceux qui souhaitent effectuer le **traitement** des présentations vidéo de son appareil*
- [2] *fy HAL Alragbap bifatH HisAb muHad~ad, yaktafy Almustaxdim biAIDagT Eiy Ayqwnap*  
<sub>[Patient]</sub> *AliDAfap w AxyAr AlHisAb min AlqA}imap*  
 في حال الرغبة بفتح حساب محدد يكتفي المستخدم بالضغط على أيقونة الإضافة واختيار الحساب من القائمة  
*Si on souhaite ouvrir un compte, il suffit que l'utilisateur applique un **appui** sur l'icône de l'ajout et de choisir le compte du menu*
- [3] *EindmA yastaxdim AlHASib niZAm, fAn risAlap saw taZhar Eind AiqlAE AlHASib* <sub>[Patient]</sub>  
*tuxbir bi\*Alik*  
 عندما يستخدم الحاسب نظام فإن رسالة سوف تظهر عند اقلاع الحاسب تخبر بذلك  
*Quant l'ordinateur utilise un système, un message s'affiche lors du **démarrage***
- [4] *Almustaxdimwn ystaTyEwn Almu\$Arakap fy ta\$gyl altaTbyqAt* <sub>[Patient]</sub> *w AlbarAmij* <sub>[Patient]</sub>  
*Almuxtalifap Eiy Ajhizap muxtalifap* <sub>[Environnement]</sub>  
 المستخدمون يستطيعون المشاركة في تشغيل التطبيقات و البرامج المختلفة على أجهزة مختلفة  
*Les utilisateurs peuvent faire le **partage** d'exécution des applications et programmes sur différents appareils.*

### Critère 3

La plupart des unités retenues partagent une parenté morphologique avec des unités déjà admises comme termes. Une partie de ces termes sont des verbes. Dans la section 6.1.1.1, nous avons discuté de ce critère en ce qui concerne les verbes, et nous avons montré que pour les verbes aussi, toutes les UTV ont des parentés avec des unités, qui sont principalement des noms déverbaux. Des cas exemplaires sont *taSaf~uH* (تصفح) (navigation) et *taSaf~aHa* (تصفح) (naviguer), *AiqlAE* (اقلاع) (démarrage) et *AqlaEa* (أقلع) (démarrer).

Cependant, certaines UTN n'ont pas de forme verbale correspondante attestée dans le corpus. À titre d'exemple, *Himayap* (حماية) (protection), *tafEyl* (تفعيل) (activation), *wSwl* (وصول) (accès) et *laSq* (لصق) (collage) n'ont aucune parenté avec d'autres termes dans notre corpus. Le problème que soulèvent ces unités est dans la nature même des UTV en discours spécialisé. Comme nous l'avons expliqué dans la section 5.3.1, nous avons écarté les unités qui n'ont pas

suffisamment d'occurrences dans le corpus et certaines unités ambiguës. Les verbes correspondants aux UTN présentées plus haut font partie des unités écartées. Par exemple, le verbe *laSaqa* (لصق) (coller) n'a que deux occurrences dans le corpus. Le verbe *HamY* (حمى) (protéger) n'a pas de sens spécialisé dans le corpus. De plus, comme discuté dans la section 5.2.4.11 certains procédés en informatique ne sont pas exprimés au moyen d'une UTV composée d'une seule lexie : ces unités ne sont pas retenues dans notre analyse puisque nous n'avons extrait que les unités simples (à la section 6.1.3, nous discutons d'une exception qui concerne les UTN).

#### Critère 4

Très peu d'UTN partagent des relations paradigmatiques. Les relations que nous avons observées sont la synonymie et l'antonymie, c'est le cas des paires montrées dans le tableau 17.

UTN	Type de lien	UTN correspondant
<i>Ainha'</i> (إنهاء) (le fait de terminer)	Antonymie	<i>ta\$gyl</i> (تشغيل) (actionnement ou amorçage)
<i>laSq</i> (لصق) (collage)	Antonymie	<i>nasx</i> (نسخ) (copie)
<i>AirsAl</i> (إرسال) (envoi)	Antonymie	<i>AistiqbAl</i> (استقبال) (réception)
<i>ta\$gyl<sub>2b</sub></i> (تشغيل) (actionnement ou amorçage)	Synonymie	<i>tanfi*<sub>1</sub></i> (تنفيذ) (exécution)
<i>AistiEAdap</i> (استعادة) (récupération)	Synonymie	<i>AistirdAd</i> (إسترداد) (récupération) et <i>AistrijaE</i> (إسترجاع) (restauration)

Tableau 17 : Relations paradigmatiques pour les UTN

Le quatrième exemple dans le tableau est un cas de synonymie parfaite ou véritable (L'Homme 2004 : 94). Dans le corpus, ces deux unités sont employées de façon interchangeable puisqu'elles ont la même structure actancielle (Agent ~ Patient) et les mêmes composantes sémantiques. Le dernier exemple dans le tableau est un cas différent du

quatrième. Les trois unités, elles aussi, s'emploient de la même façon dans le corpus. Il nous semble que les auteurs ne distinguent pas *AistirdAd* (إسترداد) (récupération) et *AistrijaE* (إسترجاع) (restauration). Ces unités renvoient à deux procédés : *récupération* dénote le fait de retrouver les données perdues ou endommagées, tandis que *restauration* dénote le fait de remettre les données (normalement le système d'exploitation) dans leur forme originale. Donc, pour les unités en question, nous croyons avoir un cas de quasi-synonymie.

### 6.1.3 Particularités

Dans notre travail, les unités visées sont les unités simples. Nous n'avons pas extrait les unités complexes. Cependant, dans la liste des candidats termes, nous avons remarqué l'existence de certaines unités qui font partie de la langue générale. Par exemple, la liste proposait des candidats comme *AiEAdap* (إعادة) (refaire) avec une spécificité de 17,02 et 425 occurrences. Après avoir vérifié les contextes de cette unité, nous avons trouvé qu'elle est souvent liée à des constructions où elle est suivie d'un terme, notamment une UTN, comme dans l'exemple suivant :

[1] *yumkinuka AiEAdap tavbyt Huzmap Alxidmap AlvAnyap Aw niZAm Alta\$gyl bilkAmil, i\*A kAn DarwryaN*

يمكنك إعادة تثبيت حزمة الخدمة ثانية أو نظام التشغيل بالكامل، إذا كان ذلك ضروريا

*Tu peux réinstaller (ou refaire l'installation du) le paquet du deuxième service ou le système d'exploitation entièrement, si c'est nécessaire*

[2] *A\*a ragibta biAiEAdap tasmyp Almalaf AlsAbiq il| Aism shahada, fanaktub AlAmr AltAly*

إذا رغبت بإعادة تسمية الملف السابق إلى Shehada فنكتب الأمر التالي

*Si tu désires renommer (ou refaire le renommage du) le précédent fichier, nous écrivons la commande suivante.*

Dans les exemples, les unités *AiEAdap tavbyt* (إعادة تثبيت) (refaire l'installation) et *AiEAdap tasmyp* (إعادة تسمية) (renommage) correspondent aux critères de validation terminologique, puisque :

1. toutes les deux appartiennent au domaine de l'informatique; et
2. leurs actants sont réalisés par des termes, comme *Huzmap* (حزمة) (paquet), *niZAm ta\$gyl* (نظام تشغيل) (système d'exploitation) et *malaf* (ملف) (fichier).

La décision de prendre en compte ces unités est fondée sur le raisonnement suivant. Nous ne pouvons pas considérer ces unités comme complexes, puisqu'elles sont issues d'une construction linguistique propre à l'ASM (en raison de l'absence des préfixes comme *re-* en ASM) et c'est la seule et unique façon d'exprimer ces réalités dans la langue technique<sup>58</sup>. Pour cette raison, nous croyons important de créer une catégorie à part pour ces unités. Une proposition possible est que ces unités tombent dans la catégorie des verbes phraséologiques de Lorente (2007). Cette proposition est fondée sur le fait que *AiEAdap* (إعادة) (refaire) appartient à la langue générale. Sa combinaison avec un terme le rend spécialisé. Une autre proposition possible est que ces unités correspondent à des collocations en discours spécialisé. Pour nous, ces unités relèvent d'une construction phraséologique. Cette construction, issue d'une traduction, est une collocation.

Un des problèmes soulevés par ces unités est leur repérage par le système d'extraction. Dans notre liste d'UTN candidates, la forme *AiEAdap* (إعادة) (refaire) faisait partie des unités à écarter, puisqu'elle relève de la langue générale. De plus, les UTN qui sont associées à cette forme figurent parmi les unités validées terminologiquement. Ceci est problématique, car la spécificité et la fréquence de chacune de ces unités (c'est-à-dire *AiEAdap* (إعادة) (refaire) et l'UTN qui lui succède) changent si la forme combinée est prise en compte. Pour cette raison, nous croyons que le système d'extraction doit être adapté pour ce genre d'unités de l'ASM de spécialité. Leur prise en compte est une nécessité puisqu'elles dénotent des procédés dans le domaine. Par conséquent, nous avons trouvé nécessaire de les inclure comme termes à part entière.

---

<sup>58</sup> Nous pouvons discuter d'autres cas comme la suffixe *-able*, dans *cliquable* dont l'équivalent ASM est *qAbil lillaqr* (قابل للنقر), ou (littéralement) *capable de se faire cliquer*.

## 6.1.4 Correspondance entre les UTV et les UTN

Notre travail sur les UTP nominales et verbales de l'ASM nous a amené à considérer la question des correspondances possibles entre les UTN et les UTV extraites. Dans les sous-sections suivantes, nous abordons cette question sous différents angles. D'abord, nous discutons la méthodologie d'extraction adoptée pour les deux unités, par la suite, nous examinons les correspondances entre les noms et les verbes extraits (les mêmes formes extraites). Enfin, nous nous attardons sur les unités françaises et anglaises qui correspondent aux UTP en ASM.

### 6.1.4.1 Extraction

Un certain nombre de différences a été observé en ce qui concerne l'extraction des verbes et des noms. Les résultats montrent clairement que la méthodologie adoptée pour l'extraction était plus efficace pour les noms que pour les verbes. Comme montré dans la section 5.2.4.10, la liste obtenue pour les verbes comportait un nombre important d'erreurs, que ce soit des erreurs d'étiquetage ou des erreurs causées par l'ambiguïté de la forme. Ainsi, le nombre des candidats verbaux était considérablement inférieur à celui des noms. Nous rappelons que nous avons obtenu 35 unités de spécificité négative, 5 unités de spécificité banale et 18 unités de spécificité positive.

Par contre, en ce qui concerne les UTN, toutes les unités retenues ont une spécificité positive (seuil de spécificité fixé à 1,96). Nous ajoutons à cela l'existence des collocations, ce qui explique la présence d'UTN complexes. Pour cette raison, nous croyons que la méthodologie d'extraction doit être modifiée et adaptée pour les verbes de l'ASM. Une méthodologie idéale d'extraction des verbes en discours spécialisé devrait tenir compte de l'ambiguïté morphologique et de la spécificité sémantique des verbes. De plus, étant donné que l'extraction s'appuie sur une méthode de comparaison, le corpus d'entraînement, utilisé dans l'analyseur morphe-syntaxique, doit être assez riche pour pouvoir distinguer les verbes spécialisés.

Nous rappelons que le corpus d'entraînement de l'analyseur utilisé dans notre recherche (AraComLex) est basé sur un corpus moderne de l'ASM (Attia et al. 2011 : 99). Dans la conception de l'analyseur, les auteurs se basent sur SAMA (voir section 5.2.4.3), mais

afin d'en exclure un certain nombre de mots obsolètes (et de mots qui relèvent du CA), ils font appel à un large corpus annoté de l'ASM sur le web. Les auteurs soutiennent que : « A lexical database intended for NLP purposes differs from traditional dictionaries in that information on inflection and derivation in the former needs to be represented in a formal and fully explicit way » (Attia et al. 2011 : 89). C'est pour cette raison qu'ils trouvent que SAMA ne dissocie pas les mots obsolètes et les mots du CA de sa ressource. De plus, l'analyseur AraComLex est de nature *open source*, c'est-à-dire qu'il nous sera possible de l'enrichir en ajoutant de nouvelles unités lexicales et le rendre plus adapté à nos besoins.

En ce qui concerne les noms, même si la méthode d'extraction a produit de meilleurs résultats, le nombre d'unités obtenues (45) était inférieur à nos attentes. Les unités dominantes dans la liste des noms étaient des noms d'entité. Dans le tableau 18, nous montrons les dix premières formes proposées par TermoStat.

UTN	Spécificité	Occurrences	Équivalent Fr
<i>niZAm</i> (نظام)	233,561221	3557	système
<i>ta\$gyl</i> (تشغيل)	143,450261	2053	actionnement
<i>\$akl</i> (شكل)	133,836875	1100	forme
<i>AistixdAm</i> (استخدام)	133,3585442	2029	utilisation
<i>malaf</i> (ملف)	129,4173799	2957	fichier
<i>mustaxdim</i> (مستخدم)	115,5246684	1766	utilisateur
<i>xidmap</i> (خدمة)	108,7008289	687	service (informatique)
<i>jihAz</i> (جهاز)	108,1107183	2368	appareil
<i>nusxap</i> (نسخة)	103,5044998	664	copie
<i>AdAp</i> (أداة)	99,65451437	626	outil

Tableau 18 : Dix premiers candidats retenus par TermoStat

Dans la liste, seulement deux formes sont des noms prédicatifs, *ta\$gyl* (تشغيل) (actionnement) et *AistixdAm* (استخدام) (utilisation)<sup>59</sup>. La méthodologie d'extraction pour ces unités est sûrement plus efficace (par rapport aux verbes).

Par curiosité, nous avons examiné le reste de la liste des candidats UTN (les unités dont la spécificité est inférieure à 1,96) afin de voir si nous pouvions trouver des unités ayant un statut terminologique. Nous avons effectivement trouvé des noms prédicatifs spécialisés. Nous mentionnons, par exemple, *tanSyb* (تنصيب) (installation) (qui est un synonyme de *tavbyt* (تنبيت) (installation)), *AstyrAd* (استيراد) (importation), *taEryf* (تعريف) (configuration) et *tazAmun* (تزامن) (synchronisation). Bien entendu, nous avons également trouvé des noms d'entités comme *zir* (زر) (boutons) et *Sadafap* (صدفة) (shell). Bien que ces unités ne soient pas de nombre important, leur position dans la spécificité négative illustre quelques limites de notre méthodologie d'extraction (du moins, lorsqu'elle est appliquée à l'ASM).

#### 6.1.4.2 Proportion des UTN par rapport aux UTV

Les deux listes des UTV et des UTN nous ont amené à considérer les correspondances possibles entre ces unités. Nous avons voulu déterminer combien de verbes présents dans la liste des UTV correspondent à un nom présent dans la liste des UTN et vice versa.

Cette question est particulièrement importante pour les raisons suivantes. Premièrement, une correspondance équilibrée entre ces unités confirme théoriquement le troisième critère de validation terminologique (la parenté morphologique). Deuxièmement, cette correspondance nous renseigne plus sur les unités les plus utilisées en ASM (les UTN versus les UTV). Dans la section 6.1.1.1, nous avons montré que certaines UTV (tableau 15) possèdent une fréquence élevée par rapport à leur forme nominale, donc les UTN. Nous avons dit que ces unités sont particulièrement intéressantes, car leur fréquence montre qu'elles sont plus utilisées dans le discours que les unités nominales.

À partir de deux listes des unités validées, nous avons réorganisé les unités, faisant en sorte de placer les verbes et leur nominalisation correspondante. Dans l'annexe 7, nous

---

<sup>59</sup> L'item *nusxap* (نسخة) (copie) désigne le résultat de l'action. C'est pour cette raison que nous ne pouvons pas le considérer comme nom prédicatif. Nous signalons qu'en ASM, la forme *nusxap* (نسخة) désigne le résultat de l'action (une copie de), tandis que *nasx* (نسخ) désigne l'action même (la copie des fichiers).

donnons la liste complète des verbes et des noms. Les résultats obtenus montrent qu'il y a 27 paires nom/verbe, c'est-à-dire qu'il existe un total de 27 correspondances entre les verbes et les noms. Dans le tableau 19, nous donnons quelques exemples.

UTN	Fréquence	Équivalent Fr	UTV	Fréquence	Équivalent Fr
إدارة <i>IdArap</i>	369	gestion <sub>1</sub>	أدار <i>AdAra</i>	30	gérer <sub>1</sub>
إدخال <i>IdxAI</i>	254	entrée <sub>1</sub>	أدخل <i>Adxala</i>	126	entrer <sub>1</sub>
إرسال <i>Ia irsAI</i>	296	envoi <sub>1a</sub>	أرسل <i>Arsala</i>	88	envoyer <sub>1</sub>
إغلاق <i>IglAq</i>	276	fermeture <sub>1</sub>	أغلق <i>Aglaqa</i>	53	fermer <sub>1</sub>
إقلاع <i>Ia iqlAE</i>	246	démarrage <sub>1a</sub>	أفلق <i>Ia AqlaEa</i>	35	démarrer <sub>1a</sub>

Tableau 19 : Exemples d'UTV et d'UTN correspondantes

En ce qui concerne les verbes et les noms qui n'ont pas de correspondances nominales et verbales, nous avons obtenu un total de 32 noms et de 19 verbes. Nous remarquons que le nombre de noms est supérieur à celui des verbes. Ce nombre pourrait suggérer que certaines activités en informatique sont surtout exprimées en ASM par des noms. Par exemple, le corpus n'atteste pas de fréquence importante de formes verbales pour *wuSwl* (وصول) (accès), *tafEyl* (تفعيل) (activation), *mu\$Arakap* (مشاركة) (partage) et *barmajap* (برمجة) (programmation). Finalement, le tableau montre clairement que la fréquence des UTN est considérablement plus élevée que celle des UTV.

#### 6.1.4.3 Correspondance avec les UTP françaises et anglaises

Dans notre travail, nous avons établi les correspondances entre les UTP que nous avons retenues et les UTP françaises et anglaises déjà décrites dans le DiCoInfo. La raison pour laquelle nous souhaitons établir une telle correspondance est que nous envisageons une portée multilingue de notre travail (voir Conclusion). De plus, puisque notre travail concerne une version en ASM du DiCoInfo, nous souhaitons montrer comment les UTP de l'ASM correspondent aux UTP dans d'autres langues. Comme montré dans la section 5.5.7, ces unités



ont été inscrites dans nos fiches terminologiques. Compte tenu de la taille de notre corpus et des objectifs de notre travail, nous avons observé un certain nombre de différences entre certaines UTP de l'ASM et leurs équivalents français et anglais. Ces différences concernent les unités qui n'ont pas de correspondances dans le DiCoInfo.

Nous avons trouvé un total de 14 UTV et de 19 UTN qui n'ont pas de correspondances dans le DiCoInfo anglais ou français. Dans les tableaux 20 et 21, nous présentons ces unités.

UTV	Signification	Équivalent Fr	Équivalent En
أدار1	gérer	gérer <sub>1</sub>	X
أوقف2	arrêter	X	X
اختار1	sélectionner	sélectionner <sub>1</sub>	X
استجاب1	répondre	X	X
بحث1	recherche	rechercher <sub>1</sub>	X
تحكم1	contrôler	X	X
تعرف1	reconnaitre	X	X
تفحص1	inspecter	analyser <sub>1a</sub>	X
توافق1	être compatible	X (compatible <sub>1</sub> )	X (compatible <sub>1</sub> )
توقف1	s'arrêter	X	X
حسن1	optimiser	optimiser <sub>1a</sub>	X
شغل1	actionner	X	run <sub>2b</sub>
عدل1	modifier	modifier <sub>1</sub>	X
فحص1	inspecter	X	X

Tableau 20 : UTV sans correspondance dans DiCoInfo

UTN	Signification	Équivalent Fr	Équivalent En
إدارة.1	Envoi	gestion <sub>1</sub>	X
إرسال.1a	Renommage	envoi <sub>1a</sub>	X
إرسال.1b	Contrôle	envoi <sub>1b</sub>	X
تسمية-إعادة.1	Accélération	renommage <sub>1</sub>	X
إغلاق.1	Fermeture	fermeture <sub>1</sub>	X
إنهاء.1	Terminer	terminer <sub>1</sub>	X
اختيار.1	Sélection	sélection <sub>1</sub>	X
1.1 بحث	Recherche	recherche <sub>1</sub>	X
1.1 تحكم	Contrôle	X	X
1. تسجيل	enregistrement	enregistrement <sub>1</sub>	X
1. تسريع	Accélération	X	X
1. تعديل	Modification	modification <sub>1</sub>	X
1. تفعيل	Activation	activation <sub>1</sub>	X
2.1 توصيل	Connexion	connexion <sub>2,1</sub>	X
1. حماية	Protection	protection <sub>1a</sub>	X
1.1 ضغط	Appui	appui <sub>1</sub>	X
1. قراءة	Lecture	lecture <sub>2</sub>	X
1. كتابة	Écriture	écriture <sub>1</sub>	X
2. كتابة	Écriture	écriture <sub>2</sub>	X
1. لصق	Collage	collage <sub>1</sub>	X

Tableau 21 : UTN sans correspondance dans DiCoInfo

Dans le tableau, nous remarquons l'UTV *tawAfaqa* (توافق) (être compatible). Cette UTV est propre à l'ASM : elle correspond à « être compatible » en français. Ce sens ne peut pas être exprimé par un verbe en anglais ou en français.

Plusieurs explications peuvent être proposées pour ce nombre d'unités sans équivalents. Premièrement, ces unités relèvent du style employé dans le discours spécialisé. Ce style est propre à l'ASM de spécialité. Cela peut nous renseigner sur les particularités de cette langue par rapport aux autres pour un domaine comme l'informatique. Deuxièmement, puisque l'extraction pour les trois langues a été faite au moyen du même extracteur (TermoStat), les corpus utilisés pour l'anglais et le français présentent des différences. Troisièmement, dans le cas des UTN, nous remarquons que les UTN anglaises équivalentes sont plus touchées que les UTN françaises. Cela peut constituer une possible similitude entre l'ASM et le français, en ce qui concerne l'utilisation de la forme nominale.

## 6.2 DiCoInfo : la version ASM

Nous rappelons qu'à la suite de l'application de notre méthodologie de repérage et de sélection des UTP en ASM de l'informatique, nous avons obtenu un total de 106 unités. Ces unités font désormais partie d'une ressource terminologique en ligne appelée le DiCoInfo version ASM<sup>60</sup>.

Comme nous l'avons dit à plusieurs reprises dans notre travail, cette version du DiCoInfo se distingue par ce qui suit. D'abord, il s'agit d'une version en langue arabe standard moderne (ASM) du DiCoInfo. Ensuite, cette version s'appuie sur une méthodologie que nous avons mise en place faisant appel à un corpus spécialisé et une extraction automatique de termes. Finalement, cette version est dédiée à la description des unités terminologiques verbales et nominales.

Le DiCoInfo de l'ASM a une interface semblable à celle du DiCoInfo (figure 87).

---

<sup>60</sup> Notre ressource est accessible gratuitement en ligne à l'adresse : <http://olst.ling.umontreal.ca/dicoinfo/arabe/index.php?ui=fr>. Nous signalons que la mise en forme de l'interface a été réalisée par Benoît Robichaud.



Figure 87 : Interface du DiCoInfo version ASM

Cette interface permet de naviguer dans les différentes sections de la ressource. D'abord, en haut, les boutons permettent d'accéder aux informations relatives à la ressource, à l'équipe, à la documentation, à la version en cadres (voir section 6.4), à l'interface en anglais et en français et, finalement, à une liste alphabétique des termes.

La liste alphabétique des termes donne accès à toutes les UTP décrites dans la ressource. Cette section comporte trois sous-listes : une liste des UTP en caractères arabes, une liste des UTP translittérées en caractères latins et une liste des termes français<sup>61</sup> correspondants (figures 88 et 89).

<sup>61</sup> Dans le cas où un équivalent français est inexistant, nous mettons tout simplement la traduction.

## Liste alphabétique des termes

[Termes arabes](#) | [Termes anglais](#) | [Termes français](#) | [Formes translittérées](#)

### Termes arabes

1. نسخ	2. عمل	1.1. ضغط	1. حماية	1. تفحص	1. تسجيل	1. برمجة	1. اختبار	2. إرسال	1. أدار
1.1. نسخ	1. فتح	1a. طباعة	1. حمل	1. تفعيل	1. تسريع	2. برمجة	1. اختيار	1. إرسال	1. أدخل
1. نفذ	1. فحص	1b. طباعة	1. خزن	1. تنفيذ	1a. تشغيل	1. تثبيت	1. استجاب	1. إعادة-تثبيت	1. أرسل
1. نقل	1. قرأ	1a. طبع	1. دخل	1. توافق	1b. تشغيل	2. تثبيت	1. استرجاع	1. إعادة-تسمية	1. أغلق
1. نقل	1. قراءة	1. طور	1. دعم	2.1. توصيل	2a. تشغيل	1. تحديث	1. استرداد	1a. إعادة-تشغيل	1a. أقم
1.1. وصول	1. كتابة	1a. ظهر	1.1. دعم	1. توقف	2b. تشغيل	1. تحرير	1. استعادة	1b. إعادة-تشغيل	1. أنشأ
1.2. وصول	2. كتابة	1. عالج	1. سجل	2. ثبت	1.1. تصفح	1. تحسين	1. استقبال	1. إغلاق	2. أوقف
	1. كتب	1. عدل	1. شغل	1. حدث	1. تصفح	1.1. تحكم	2. استقبال	1a. إغلاق	1.1. اتصال
	1. لصق	1b. عرض	1. ضبط	1. حذف	1. تطوير	1. تحكم	1. استقبال	1b. إغلاق	1. اتصال
	1. مشاركة	1b.1. عرض	1. ضغط	1. حسن	1. تعديل	1. تحميل	1. بحث	1. إنشاء	1. إدارة
	1. معالجة	1. عمل	2. ضغط	1. حفظ	1. تعرف	1. تحويل	1.1. بحث	1. إنهاء	1. إدخال

Figure 88 : Liste alphabétique des UTP en caractères arabes

### Termes français

accepter.1	afficher.1b	connecter.1	démarrage.1b	exécuter.1	installation.2	optimiser.1	redémarrage.1a	stocker.1	téléchargement.1
accès.1.1	afficher.1a	connexion.1.1	démarrer.1a	exécution.1	installer.2	ouvrir.1	redémarrage.1b	support.1.1	télécharger.1
accès.1.2	analyser.1	connexion.2.1	développement.1	fermer.1	lecture.1	paramétrer.1	renommage.1	supprimer.1	écrire.1
accélération.1	appui.1.1	contrôle.1	développer.1	fermeture.1	lire.1	partage.1	restauration.1	sélection.1	écriture.1
actionnement.1a	appuyer.1	contrôler.1.1	enregistrement.1	gestion.1	mettre à jour.1	programmation.1	réception.1	sélectionner.1	écriture.2
actionnement.1b	arrêter.1	conversion.1	enregistrer.1	gérer.1	mise à jour.1	programmation.2	réception.2	terminer.1	édition.1
actionnement.2a	arrêter.2	copie.1.1	entrer.1	impression.1a	modification.1	protection.1	récupération.1	tourner.1	
actionner.1	cliquer.1	copier.1	entrée.1	impression.1b	modifier.1	recevoir.1	réinstallation.1	tourner.2	
activation.1	collage.1	création.1	envoi.1	imprimer.1a	navigation.1	recherche.1.1	répondre.1	traitement.1	
affichage.1b.1	compatible.1	créer.1	envoi.2	inspecter.1	naviguer.1.1	rechercher.1	saisir.1	traiter.1	
	compression.2	démarrage.1a	envoyer.1	installation.1	optimisation.1	reconnaitre.1	sauvegarder.1	transférer.1	

### Formes translittérées

AdAra.1	AinhA>.1	AixtAra.1	DagT.1.1	Hassana.1	barmajap.2	muEAlajap.1	taEdyl.1	tafEvl.1	tawAfaqa.1
Adxala.1	AiqIAE.1a	Anchaa.1	Dagata.1	Hazafa.1	chaggala.1	mushArakap.1	taHakkama.1.1	tafaHHaSa.1	tawSvl.2.1
Aglaqa.1	AiqIAE.1b	Aqlaa.1a	EAlaja.1	HimAyap.1	daEama.1	naffaza.1	taHakkum.1	tanfy*.1	tawaqqafa.1
AiEAdap-tashgyl.1a	AirsAl.2	Arsala.1	Eaddala.1	TabaEa.1a	daEm.1.1	naqala.1	taHdyv.1	tashgyl.1a	yabbata.2
AiEAdap-tashgyl.1b	AirsAl.1	Astaqbala.1	Eamila.1	Tawwara.1	daxala.1	naqara.1	taHmyl.1	tashgyl.1b	wuSwl.1.1
AiEAdap-tasmyap.1	AistajAba.1	AstiEAdap.1	Eamila.2	TibAEap.1a	faHaSa.1	nasaxa.1	taHnyr.1	tashgyl.2a	wuSwl.1.2
AiEAdap-tavbyt.1	AistiqbAl.1	AstirdAd.1	EarD.1b.1	TibAEap.1b	fataHa.1	nasx.1.1	taHsyn.1	tashgyl.2b	xazzana.1
AidArap.1	AistiqbAl.2	Awqafa.2	EaraDa.1b	Zahara.1a	kataba.1	qaraa.1	taHwyl.1	tasyl.1	
AidxAl.1	AistirAE.1	AxtvAr.1	Haddava.1	baHava.1	kitAbap.1	qirA>p.1	taSaffaHa.1.1	tasvyE.1	
AiglAq.1	AittaSala.1	DabaTa.1	HafZa.1	baHv.1.1	kitAbap.2	sajjala.1	taSaffuH.1	tavbyt.1	
AinchA'.1	AittiSAI.1.1	DagT.2	Hammala.1	barmajap.1	laSq.1	taEarrafa.1	taTwyrl.1	tavbyt.2	

Figure 89 : Listes alphabétiques des équivalents français et des formes translittérées

La recherche dans le DiCoInfo de l'ASM se fait à partir de l'interface. L'utilisateur a le choix de chercher par UTP ASM, par l'équivalent français ou par la forme translittérée. De plus, l'interface permet de définir la précision de la recherche : terme exact, débutant par ou

contenant. Conscient du fait que les caractères arabes ne sont pas disponibles sur certaines plateformes, nous avons ajouté un clavier virtuel avec des caractères arabes.

Le résultat d'une recherche dans le DiCoInfo version ASM consiste à proposer à l'utilisateur le terme recherché, les équivalents français et anglais, la structure actancielle et trois contextes (figure 90).

Recherche dans les fiches du DiCoInfo

Mode : terme  Afficher les équivalences

Langue : arabe

Précision : exacte

Rechercher : طور

**Résultats de la recherche :**  
Apparaît comme entrées dans les articles :  
( fr : [développer\\_1](#) , en : [develop\\_1](#) ↔ ) طور\_1

طور\_1 , v. tr .

طور: شبكة 1 + , بواجهة 1 +

**Contextes**  
**Contextes annotés**

يعتبر نظام " اي او اس " الذي تطور ه ابل مغلق المصدر (Source : Aljazeera)  
يذكر ان الطابعة المسماة " غوغل كلاود برنت " قدمت ها غوغل اول مرة في ابريل / نيسان 2010 ك حل مستقبلي ل الطابعة من نظام التشغيل كروم الذي تطوره الشركة (Source : Aljazeera)  
موزيلا تطور نموذج لأول حاسب لوحي مزود بنظام " فايفوكس (Source : AitNews)

français : [développer\\_1](#)  
anglais : [develop\\_1](#)  
Rédacteur(s) : NG  
Date de mise à jour : 2014-01-01

Figure 90 : Résultat d'une requête dans le DiCoInfo, version ASM

Un certain nombre d'informations s'affiche dans les résultats. Par exemple, on peut y voir la date de mise à jour, un code de rédacteur et le statut. Dans notre exemple, le statut est à 1, ce qui veut dire que la rédaction de cette fiche est très avancée. La structure actancielle peut afficher les termes réalisés pour chaque rôle sémantique. Pour ce faire, il suffit de placer le curseur de la souris sur le signe (+) situé à côté du terme typique. Dans la figure, le terme en question est l'UTV *Taw~ara* (طور) (développer) : la structure actancielle de cette unité est **Agent** (compagnie) ~ **Patient** (logiciel).

Un aspect important de la fiche réside dans l'accès aux contextes annotés. Ces contextes sont accessibles en cliquant sur le lien « Contextes annotés » situé au-dessus des trois contextes affichés (figure 91).

يعتبر نظام " اي او اس " الذي تطوره ابل مغلق المصدر (Source : Aljazeera)

يذكر ان الطباعة السحابية " غوغل كلاود برنت " قدمت ها غوغل اول مرة في ابريل / نيسان 2010 ك حل مستقبلي ل الطباعة من نظام التشغيل كروم الذي تطوره الشركة (Source : Aljazeera)

موزيلا تطور نموذج لأول حاسب لوحي مزود بنظام " فايرفوكس (Source : AitNews)

لم تطور انظمة تشغيل يوما بهذا التوزع و الامتداد (Source : Linux)

تطور مائديفا برمجياتها الخاصة والعديد من المعالجات للمساعدة في إدارة وتثبيت معظم المهام الشائعة (Source : Linux)

اما لينوكس فلم يطور ليكون نظام تشغيل تجاري ، بل على العكس فقد طور وفقا لسياسة تطوير الأنظمة المفتوحة (Source : Taibahu)

من العوامل التي ساهمت في نجاح لينوكس هو استخدامه لنظام و الذي طور عام 1984 (Source : Taibahu)

طورت شركة هولندية سبرانغ SPRANQ مجموعة خطوط تقدمها مجانا لتقليص استهلاك حبر الطباعة من خلال تصميم خطوط فيها نقاط فارغة (Source : ITP)

واستغل من يطور الفيروسات هذه الميزة للتحايل على الناس لتجنب كشف الملفات التنفيذية ليعتقدوا أنها مجرد ملف نصي (Source : ITP)

عادة ما تطور الشركات المصنعة لتلك الطابعات تطبيقات خاصة مجانية استخدامها على الأجهزة الذكية مع طابعاتهم (Source : Al-Jazeera)

لغة بيثون ما تزال تطور (Source : Python-ar)

عند ما تطور تطبيق ، ف من المهم ان تفكر في اعادة تحجيم الصفحة (Source : Python-ar)

سوف تفهم ان نا لن نشرح في هذه الصفحات كيفية صنع برنامج خادم . و هذا من عمل المتخصصين ( على سبيل المثال ، كانتك ستطور لغة ب برمجة جديدة (Source : Python-ar)

بالإضافة الى لب النظام فإنه يحتوي على العديد من برامج و تطبيقات التي طورها مستخدمون من كل مكان في العالم (Source : Taibahu)

طور سولاريس ليعمل فقط على اجهزة و عتاد شركة (Source : Taibahu)

Figure 91 : Contextes annotés pour l'UTV *Taw~ara* (طور) (développer)

Les contextes apparaissant dans la figure sont donnés pour l'UTV *Taw~ara* (طور) (développer). Les rôles sémantiques de la structure actancielle peuvent être identifiés au moyen des couleurs associées à chaque rôle. Dans le cas de notre exemple, le rôle Agent est en rouge, tandis que Patient est en bleu :

[1] *yuEtabar niZAm « Ay Aw As » Al~a\*y tuTaw~iruhu Ab~il muglaq almaSdar*  
يعتبر نظام "أي او اس" الذي تطوره ابل مغلق المصدر  
*Le système iOS, développé par Apple est considéré une source fermée*

De plus, ces rôles peuvent être affichés lorsque l'utilisateur place le curseur de la souris sur le terme annoté.

Finalement, pour l'ensemble des contextes annotés, nous ajoutons un tableau récapitulatif qui résume le groupe ainsi que la fonction syntaxiques observés pour chaque rôle sémantique (figure 92).

Actants		
Patient (15)	Objet [SN] (8) Objet [Pro] (4) Sujet [SN] (3)	الذي نموذج لينوكس لغة ب رمجة لغة سولاريس تطبيق التي خطوط تطبيقات برمجياتها انظمة تشغيل الفيروسات
Agent (8)	Sujet [SN] (7) Sujet [Pro] (1)	موزيلا من ماندريفا شركة الشركة ابل مستخدمون الشركات

Figure 92 : Tableau récapitulatif des réalisations linguistiques des actants

Pour l'UTV *Taw~ara* (طور) (développer), le Patient est réalisé par trois groupes syntaxiques : un objet réalisé par un syntagme nominal (SN), un objet réalisé par un pronom (Pro) et un sujet<sup>62</sup> (syntagme nominal). Pour l'Agent, nous avons deux groupes syntaxiques, SN et Pro, comme sujet.

### 6.3 Cadres identifiés

Nous rappelons que notre travail sur les UTP de l'ASM s'appuie sur la théorie de la Sémantique des cadres. De plus, un de nos objectifs est la modélisation des UTP dans des cadres sémantiques, suivant le modèle de projet FrameNet (Ruppenhofer et al. 2010).

Pour cette raison, après avoir analysé, décrit et encodé les UTP de l'ASM, nous avons procédé à l'identification des cadres sémantiques qui regroupent les UTP décrites. Dans la section 5.6.3, nous avons expliqué la méthodologie qui nous a permis de faire cette

<sup>62</sup> Il s'agit du sujet syntaxique d'un verbe à la voix passive.



identification. Nous avons également décrit notre méthode d'encodage de ces derniers. Dans la présente section, nous abordons nos résultats. Nous parlons des cadres que nous avons pu identifier, les unités associées à ces derniers et, finalement, la ressource dans laquelle nous présentons ces cadres.

### **6.3.1 Les UTP et les cadres correspondants**

Pour pouvoir identifier des cadres sémantiques, nous avons utilisé toutes les UTP validées dans notre travail, soit un total de 106 termes. Ces unités ont été intégrées dans des cadres en fonction des critères discutés dans la section 5.6.3. Il convient de signaler que les cadres ont été identifiés à partir des UTP en ASM. Comme certains équivalents anglais et français ont été identifiés (voir section 6.1.4.3), ils ont été ajoutés aux cadres à titre indicatif lorsque des contextes annotés étaient disponibles dans les versions anglaise et française du DiCoInfo. De plus, nous signalons que l'établissement des liens d'équivalence a fait l'objet d'un travail préliminaire (Ghazzawi et al. à paraître).

À la suite de cette intégration, nous avons obtenu 57 cadres. Ces cadres appartiennent aux quatre groupes, discutés dans la section 5.6.2. Dans le tableau 21, nous présentons nos résultats pour chaque groupe. Nous signalons que, dans les tableaux qui suivent, nous ne comptabilisons pas certains cadres, puisque ces cadres sont de nature non lexicale, c'est-à-dire qu'ils ne renferment pas d'UTP. L'explication pour la création de tels cadres sera donnée dans la section section 6.3.3.

<b>Groupe</b>	<b>Cadre</b>	<b>Total</b>
<b>Nouveau</b>	Being_off_operation, Cause_modification, Connecting, Display_process, Duplication_process, Enhancing, Hardware_setup, Installing_resume, Manipulating_components, Moving_text_into, Program_creation, Programming, Reforming, Reproduction, Running_on_system, Running_a_program, Text_editing, Transmitting	18
<b>Correspondance parfaite</b>	Activity_start, Activity_resume, Being_in_operation, Change_operational_state, Choosing, Having_or_lacking_access, Having_access_to, Intentionally_create, Operating_a_system, Product_development, Placing, Process_start, Process_resume, Protecting, Rejuvenation, Removing, Response, Sharing, Storing, Supporting	20
<b>Correspondance semi-parfaite</b>	Duplication, Installing, Process_start, Placing, Reading_perception, Receiving,	6
<b>Correspondance partielle</b>	Attaching, Becoming_aware, Cause_change, Cause_impact, Compatibility, Control, Information_display, Inspecting, Name_conferral, Processing_materials, Scrutiny, Seeking, Sending, Text_creation, Transfer	15

Tableau 22 : Quatre groupes de cadres

Nos résultats montrent que la plupart des cadres dont nous avons eu besoin pour nos UTP correspondent à des cadres se trouvant déjà dans FrameNet. Le nombre le plus élevé est celui des cadres qui sont en correspondance parfaite avec nos données.

En ce qui concerne les cadres mêmes et les UTP qui leur appartiennent, nous présentons les résultats dans deux tableaux. Le premier (tableau 23) comprend les résultats pour l'ASM et le deuxième (tableau 24) comprend les résultats pour les trois langues.

<b>Cadres à 1 terme</b>	Activity_resume, Attaching, Becoming_aware, Being_off_operation, Cause_change, Compatibility, Display_process, Duplication, Hardware_setup, Having_or_lacking_access, Installing_resume, Moving_text_into, Name_conferral, Process_resume, Protecting, Removing, Response, Running_a_program, Sending, Sharing, Text_editing	<b>Total : 21</b>
<b>Cadres à 2 termes</b>	Being_in_operation, Cause_modification, Choosing, Connecting, Control, Duplication_process, Enhancing, Having_access_to, Information_display, Inspecting, Installing, Intentionally_create, Operating_a_system, Process_start, Program_creation, Programming, Product_development, Processing_materials, Reading_perception, Receiving, Reforming, Reproduction, Running_on_system, Scrutiny, Seeking, Supporting, Transmitting	<b>Total : 27</b>
<b>Cadres à 3 termes</b>	Activity_start, Rejuvenation, Storing, Transfer	<b>Total : 4</b>
<b>Cadres à 4 termes</b>	Cause_impact, Manipulating_components	<b>Total : 3</b>
<b>Cadres à 5 termes</b>	Change_operational_state	<b>Total : 2</b>
<b>Cadres à 6 termes</b>	Placing	<b>Total : 1</b>

Tableau 23 : Répartition des UTP de l'ASM dans les cadres correspondants

<b>Cadres à 1 terme</b>	Activity_resume, Attaching, Becoming_aware, Being_off_operation, Display_process, Name_conferral, Process_resume, Protecting, Response, Sharing	<b>Total : 9</b>
<b>Cadres à 2 termes</b>	Cause_change, Choosing, Control, Enhancing, Information_display, Inspecting, Process_start, Programming, Receiving, Seeking, Sending, Storing	<b>Total : 10</b>
<b>Cadres à 3 termes</b>	Cause_modification, Compatibility, Hardware_setup, Having_or_lacking_access, Installing, Process_start, Reading_perception, Rejuvenation, Transmitting	<b>Total : 7</b>
<b>Cadres à 4 termes</b>	Being_in_operation, Connecting, Having_access_to, Moving_text_into, Operating_a_system, Reproduction, Running_on_system, Text_creation, Text_editing	<b>Total : 9</b>
<b>Cadres à 5 termes</b>	Duplication, Duplication_process, Processing_materials, Reforming, Removing, Supporting	<b>Total : 4</b>
<b>Cadres à 6 termes</b>	Change_operational_state, Intentionally_create, Product_development, Running_a_program, Scrutiny	<b>Total : 5</b>
<b>Cadres à 7 termes</b>	Activity_start	<b>Total : 1</b>
<b>Cadres à 8 termes</b>	Cause_impact, Transfer	<b>Total : 2</b>
<b>Cadres à 9 termes</b>	Manipulating_components	<b>Total : 1</b>
<b>Cadres à 10 termes</b>	Program_creation	<b>Total : 1</b>
<b>Cadres à 12 termes</b>	Placing	<b>Total : 1</b>

Tableau 24 : Répartition des UTP des trois langues dans les cadres correspondants

D'après le tableau 23, pour l'ASM, le nombre maximal d'unités qu'on peut regrouper dans un seul cadre est six unités. De plus, la plupart des cadres renferment une à deux unités. La situation ne change pas radicalement lorsqu'il s'agit des trois langues. La particularité constatée dans le tableau 24 est que nous avons des cas où un cadre renferme jusqu'à douze UTP. Il convient de signaler que les chiffres présentés dans le tableau 24 sont de nature indicative, puisque de nouvelles données peuvent toujours être ajoutées. De plus, nous rappelons qu'en ce qui concerne les UTP anglaises et françaises intégrées dans ce travail, nous

n'avons inclus que les données disponibles (les UTP déjà annotées dans le DiCoInfo anglais et français).

Comme discuté dans la section 5.6.4, un cadre comportant des UTP fournit les renseignements suivants : le nom du cadre, sa définition, des exemples de contextes, les actants et les UTP qui l'évoquent. Dans ce qui suit, nous donnons une description complète de deux cadres différents. Le premier cadre renferme un minimum d'UTP (une seule unité), tandis que le deuxième cadre renferme un maximum d'UTP (10 unités).

Le premier cadre est Compatibility<sup>63</sup>. La particularité que présente ce cadre est qu'il est propre à l'ASM. L'UTP qu'il renferme n'a pas d'équivalent direct dans les autres versions du DiCoInfo. Ce cadre décrit l'état de compatibilité entre deux composants informatiques. Il renferme deux actants, Patient-a, qui désigne le premier composant, et Patient-b, qui désigne le deuxième composant informatique.

<b>Cadre</b>	Compatibility	
<b>Définition</b>	Patient-a (hardware or software) can work with Patient-b (hardware or software) in a system.	
<b>UTV</b>	<i>tawAfaqa</i> <sub>1</sub> (توافق) (être compatible)	
<b>Note</b>	This frame is based on Compatibility in FrameNet.	
<b>Contexte</b>	<b>ASM</b>	laEal i\$Arap min mAykrwswift tuxbiruka biAn Alwaqt qad HAn min Ajil \$irA' jihAz jadyd yaEmal EIY windwz 7 Aw Hat~Y \$ira' nusxap windwz jadydap tatwAfaq maE Almasinjar Aljadyd ولعل إشارة من مايكروسوفت لتخبرك بأن الوقت قد حان من أجل شراء جهاز جديد يعمل على ويندوز 7 أو حتى شراء نسخة ويندوز جديدة تتوافق مع الماسنجر الجديد Peut-être un signe de Micro\$oft t'informe qu'il est temps d'acheter un nouvel appareil tournant sous Windows 7, ou même acheter une nouvelle version de Windows qui est compatible avec le nouveau Messenger.
<b>Actant</b>	<b>Patient-a</b>	kart \$A\$ap (كرت شاشة) (carte graphique)
	<b>Patient-b</b>	lawHap Aum (لوحة أم) carte mère

Tableau 25 : Description du cadre Compatibility

<sup>63</sup> Nous rappelons que les descriptions des cadres sont données en anglais.

Le deuxième cadre que nous décrivons est celui qui renferme un nombre plus élevé d'UTP. C'est le cadre Program\_creation. Ce cadre décrit le procédé selon lequel un Agent (utilisateur, concepteur, développeur, informaticien, etc.) crée un Patient (composant informatique : logiciel, programme, application, etc.) au moyen d'un Matériau (langage de programmation).

<b>Cadre</b>	Program_creation	
<b>Définition</b>	Agent (user) creates Patient (program) using Material (language) intended for this purpose.	
<b>UTP</b>	<b>ASM</b>	kitAbap2 (كتابة) (écriture), barmajap1 (برمجة) (programmation)
	<b>Anglais</b>	write2, program3, programming1, scripting1
	<b>Français</b>	écrire2, écriture2, programmer1, programmation1
<b>Note</b>	This frame is defined for DiCoInfo.	
<b>Contexte</b>	<b>ASM</b>	in~a kitAbap barnAmij bilugap rafyEap Almustaw  Ashal kavyraN wa ystagriq waqtaN Aqal likitabatihi w AHtimAl AlAxTA' fyhi qalylap إن كتابة برنامج بلغة رفيعة المستوى اسهل كثيرا و يستغرق وقتا اقل لكتابته و احتمال الاخطاء فيه قليلة Écrire un programme dans un langage sophistiqué est plus facile et prend moins de temps. La possibilité de faire des erreurs est minime.
	<b>Anglais</b>	Java can be used to write applications for both Web and non-Web use.
	<b>Français</b>	Le programmeur écrit son programme dans un langage de programmation l'aide d'un éditeur.
<b>Actant</b>	<b>Agent</b>	mubarmij (مبرمج) (programmeur), mutaw~ir (مطور) (développeur),
	<b>Patient</b>	barnAmij (برنامج) (programme), naS barmajy (نص برمجي) (script), Eamaly~ap (عملية) (processus), niZAM ta\$gyl (نظام تشغيل) (système d'exploitation),
	<b>Matériau</b>	lugap (لغة) (langage),
<b>Circonstant</b> <sup>64</sup>	But, Degree, Destinataire, Environnement, Instrument, Lieu, Location, Manière, Manner, Method, Moyen, Méthode, Purpose, Reason, Recipient, Reference, Temps, Time.	

Tableau 26 : Description du cadre Program\_creation

<sup>64</sup> Il faut signaler que la dernière section, où nous donnons les circonstants, concerne les UTP anglaises et françaises et non pas les UTP de l'ASM, puisque les circonstants ne font pas partie de notre analyse.

Toutes les UTP de ce cadre partagent les mêmes propriétés sémantiques et elles correspondent aux critères discutés dans la section 5.6.3.

### 6.3.2 Relations entre UTP dans les cadres

Dans cette section, nous décrivons les relations entre les UTP lorsqu'un cadre en renferme plusieurs. Comme montré dans le tableau 22, certains cadres ont plus d'une UTP. Ces unités possèdent des propriétés sémantiques communes et c'est pour cette raison qu'elles sont regroupées dans un même cadre. Dans ce qui suit, nous abordons cette question sous trois angles : la relation entre forme nominale et forme verbale, la relation de synonymie et d'antonymie et la relation de correspondance interlinguistique.

#### 6.3.2.1 Relation entre forme verbale et nominale

Une partie de nos cadres renferme des verbes et des noms déverbaux (UTV et leur nominalisation). Bien entendu, la forme nominale d'un verbe n'est retenue que lorsque celle-ci ne présente aucune différence par rapport au verbe. Dans le tableau 27, nous présentons un cadre renfermant une UTV et une UTN.

<b>Cadre</b>	Seeking	
<b>Définition</b>	Agent (user) looks for Patient (data, web site or files) in Location (Internet or storage unit).	
<b>UTP</b>	baHava <sub>1.1</sub> (بحث) (rechercher), baHv <sub>1</sub> (بحث) (recherche)	
<b>Note</b>	This frame is based on Seeking in FrameNet. The number of actants vs. FE is different.	
<b>Contexte</b>	<b>ASM</b>	Aw~alaN, iftaH Almaktabap Almawjwdap EIY mawqiEak. vum~ ibHav Ean AlmustanadAt Almawjwdap Ela  Alkumbywtar wa Al~aty turyd taHmylhA أولاً، افتح المكتبة الموجودة على موقعك. ثم إبحث عن المستندات الموجودة على الكمبيوتر والتي تريد تحميلها D'abord, ouvre la bibliothèque qui se trouve sur ton site. Ensuite, recherche les documents situés dans ton ordinateur que tu veux télécharger
<b>Actant</b>	<b>Agent</b>	mustaxdim (مستخدم) (utilisateur)
	<b>Patient</b>	mustanadAt (مستند) (documents)
	<b>Lieu</b>	kumbywtar (كمبيوتر) (ordinateur)

Tableau 27 : UTV et UTN dans cadre Seeking

Dans le cadre Seeking, les unités qui lui appartiennent sont reliées par une parenté morphologique et sémantique (critère 3 de la validation terminologique). Considérons les exemples suivants pour ces deux unités.

[1] *ibHav Ean ha\*A Almalaf biAstixdAm qA}map Start vum~ Find*

Find ثم Start إبحث عن هذا الملف باستخدام قائمة

*Recherche ce fichier en utilisant le menu Start, puis Find*

[2] *yumkin lilmustaxdim Eind AlbaHv Ean Swrap fy Ay min AltaTbyqayn mu\$Ahadap Ay min nata}j AlbaHv fy waDE tamil Al\$A\$ap*

يمكن للمستخدم عند البحث عن صورة في أي من التطبيقين مشاهدة أي من نتائج البحث في وضع كامل الشاشة

*Lors de la recherche d'une photo dans une des applications, l'utilisateur peut visionner les résultats de la recherche en mode plein écran.*

D'après les contextes, les deux unités en question ont les mêmes actants de même nature, à savoir Agent, Patient et Lieu. Nous n'avons pas constaté une différence de sens dans les différentes réalisations linguistiques. Le regroupement des UTP par forme morphologique apparentée est assez courant. Toutes les UTV et UTN équivalentes partagent le même cadre, pour un total de 27 cadres.

### 6.3.2.2 Relations de synonymie et d'antonymie

La deuxième relation est celle de synonymie et d'antonymie entre les UTP. D'après les critères présentés dans Ruppenhofer et al. (2010), un cadre peut renfermer des antonymes. Dans notre travail, nous avons trouvé deux cadres contenant des antonymes, à savoir Manipulating\_components (tableau 28) et Change\_operational\_state.



<b>Cadre</b>	Manipulating_components	
<b>Définition</b>	Agent (user) shows or hides the contents of Patient (files) in order to use it or shut it.	
<b>UTP</b>	<b>ASM</b>	Aglaqa <sub>1</sub> (أغلق) (fermer), iglaq <sub>1</sub> (إغلاق) (fermeture) fataHa <sub>1</sub> (فتح) (ouvrir), DagT <sub>1</sub> (ضغط) (compression)
	<b>Anglais</b>	open <sub>1</sub>
	<b>Français</b>	ouvrir <sub>1</sub> , ouverture <sub>1</sub> , fermer <sub>1</sub>
<b>Note</b>	This frame is defined for DiCoInfo.	
<b>Contexte</b>	<b>ASM</b>	biAimkanika An taftaH fy kul saTH maktab mA tawad~ min barAmij aw nawAfi* wa tantaqil bikul shwlap byn h*ihi AlAsTuH mimA ysAhim fy Eamdah AzdiHAM saTH Almaktab بإمكانك أن تفتح في كل سطح مكتب ما تود من برامج أو نوافذ و تنتقل بكل سهولة بين هذه الأسطح مما يساهم في عدم ازدحام سطح المكتب Tu peux ouvrir dans chaque bureau autant de programmes et de fenêtre que tu veux et de te déplacer facilement entre ces bureaux. Ceci permet d'éviter d'avoir trop d'applications sur le bureau.
	<b>Anglais</b>	For example, suppose you want to OPEN a file and write the numbers 1 to 10 in it
	<b>Français</b>	Vous OUVREZ le sous-menu de la commande POSITION (POSITIONNER)
<b>Actant</b>	<b>Agent</b>	mustaxdim (مستخدم) (utilisateur)
	<b>Patient</b>	barnAmij (برنامج) (programme), nAfi*ap (نافذة) (fenêtre), malaf (ملف) (fichier)

Tableau 28 : Relation d'antonymie dans le cadre Manipulating\_components

Comme montré dans le tableau, ce cadre a été défini pour nos données (ASM). Dans le domaine de l'informatique, les UTP faisant partie de ce cadre dénotent un processus qui consiste à ouvrir ou à fermer un composant informatique (fenêtre, logiciel, etc.) par un utilisateur. On peut noter que l'unité *DagT* (ضغط) (compression) en fait partie, puisque cette unité désigne l'activité de renfermer des composants (fichiers, photos, etc.).

En ce qui concerne les relations de synonymie, nous en avons découvert un certain nombre. L'exemple que nous donnons est le cadre Transfer (tableau 29).

<b>Cadre</b>	Transfer	
<b>Définition</b>	Agent (user) relocates or removes Patient (data) from Source (computer or server) to Destination (computer).	
<b>UTP</b>	<b>ASM</b>	naqala <sub>1</sub> (نقل) (transférer), ham~ala <sub>1</sub> (حمل) (télécharger), taHmyl <sub>1</sub> (تحميل) (téléchargement)
	<b>Anglais</b>	download <sub>1</sub> , download <sub>1.1</sub> , upload <sub>1</sub>
	<b>Français</b>	télécharger <sub>1</sub> , téléchargement <sub>1</sub>
<b>Note</b>	This frame is based on Transfer in FrameNet.	
<b>Contexte</b>	<b>ASM</b>	nusxap windwz tistA Al*y AxtarahA Almustaxdim lyHam~ilhA EIY jihAzihi yumkin An twad~y ilY xsArap baED AlwaZA}if AlxASap biAlniZam Alqadym نسخة ويندوز فيستا الذي اختارها المستخدم ليحملها على جهازه يمكن ان تؤدي الى خسارة بعض الوظائف الخاصة بالنظام القديم La version de Windows Vista que l'utilisateur a choisi de télécharger sur son ordinateur peut entraîner la perte de certaines fonctions de l'ancien système.
	<b>Anglais</b>	DOWNLOAD the file to your download directory
	<b>Français</b>	Décompressez ensuite tous les fichiers TÉLÉCHARGÉS dans ce répertoire Destination
<b>Actant</b>	<b>Agent</b>	mustaxdim (مستخدم) (utilisateur)
	<b>Patient</b>	barnAmij (برنامج) (programme), malaf (ملف) (fichier), taHdyv (تحديث) (mise à jour), luEbp (لعبة) (jeu)
	<b>Source</b>	mawqiE (موقع) (site), intarnit (إنترنت) (Internet)
	<b>Destination</b>	jihAz (جهاز) (appareil), qurS (قرص) (disque)

Tableau 29 : Relation de synonymie dans le cadre Transfer

Les deux unités des trois langues, *Ham~ala<sub>1</sub>* (حمل) (télécharger), *taHmyl<sub>1</sub>* (تحميل) (téléchargement), *download<sub>1.1</sub>* et *download<sub>1</sub>*, *télécharger<sub>1</sub>* et *téléchargement<sub>1</sub>* dénotent l'activité de déplacer un composant informatique d'un ordinateur source (Source) à un ordinateur cible (Destination). Dans les contextes analysés de ces unités, nous avons constaté la présence (implicite ou explicite) des actants faisant partie de la structure actancielle.

En ce qui concerne les UTP *naqala<sub>1</sub>* (نقل) (transférer) et *upload<sub>1</sub>*, nous les avons inclus dans ce cadre pour les raisons suivantes. L'UTV ASM *naqala<sub>1</sub>* (نقل) (transférer) exprime la même idée de déplacement avec la même structure actancielle. Bien que la structure actancielle ne soit pas la seule un indice quant au regroupement des unités dans un même cadre, elle permet d'identifier le schéma conceptuel dénoté par les unités. Une exception constatée dans ce cadre est l'UTV anglaise *upload<sub>1</sub>*. Comme discuté dans la section précédente, d'après les critères de sélection des cadres, un cadre peut renfermer des antonymes. Cependant, être un antonyme ne suffit pas pour l'inclure dans un cadre. Cette unité dénote un procédé selon lequel l'utilisateur déplace un composant d'une Source à une Destination. Par conséquent, nous estimons qu'elle peut aussi faire partie de ce cadre.

### **6.3.2.3 Relation de correspondance**

Finalement, la relation de correspondance concerne les cadres qui renferment des unités de différentes langues, puisque nous avons pris en compte certaines unités anglaises et françaises dans cette partie de notre travail. Comme signalé auparavant, nous avons inclus les unités qui sont disponibles dans les versions anglaise et française du DiCoInfo. Dans la section 5.5.7, nous avons parlé de l'annotation et nous avons présenté les deux facteurs pour la prise en compte des UTP anglaises et françaises, à savoir le sens et la structure actancielle.

Les cadres dont il est question dans cette sous-section représentent la presque totalité des cadres découverts. Nous avons remarqué que la prise en compte des relations interlinguistiques nous a permis d'éviter qu'un cadre ne renferme qu'une seule unité. Dans le tableau 30, nous présentons un exemple typique de cette relation.

<b>Cadre</b>	Installing_resume	
<b>Définition</b>	Agent (user) reintegrates Patient (system or program) in Destination (hardware)	
<b>UTP</b>	<b>ASM</b>	<i>iEAdap tavbyt<sub>1b</sub></i> (إعادة تثبيت) (réinstallation),
	<b>Français</b>	réinstaller <sub>1</sub>
<b>Note</b>	This frame is defined for the DiCoInfo.	
<b>Contexte</b>	<b>ASM</b>	Ealyka biAlHal AlADman w hwa iEAdap tavbyt baramijika ELY Alkumbywtar Aljadyd عليك بالحل الأمثل وهو إعادة تثبيت برامجك على الكمبيوتر الجديد Il te faut la solution idéale : réinstaller tes logiciels sur le nouvel ordinateur
	<b>Français</b>	Si jamais vous voulez RÉINSTALLER le système par défaut dans le MBR, vous devez taper la commande FDISK MBR sous DOS.
<b>Actant</b>	<b>Agent</b>	mustaxdim (مستخدم) (utilisateur)
	<b>Patient</b>	niZAm (نظام) système, barnAmij (برنامج) logiciel,

Tableau 30 : Relation de correspondance entre UTP dans le cadre Installing\_resume

Le cadre Installing\_resume ne renferme qu'une seule unité, l'UTN ASM *iEAdap tavbyt<sub>1b</sub>* (إعادة تثبيت) (réinstallation). Dans le DiCoInfo, la seule unité susceptible de faire partie du cadre est l'UTV française *réinstaller<sub>1</sub>*. Bien entendu, ces deux UTP sont équivalentes.

### 6.3.3 Relations entre cadres : les scénarios conceptuels

Nous avons montré dans la section précédente comment les UTP sont reliées les unes aux autres dans les cadres. Ces derniers peuvent également partager des relations. Comme constaté par Ruppenhofer et al. (2010 : 73), le but de relier les cadres est de faciliter leur compréhension.

Dans le cas d'une application à un domaine de spécialité, le fait de relier les cadres permet de montrer les interrelations entre les différentes activités et comment ces activités et les processus s'enchaînent dans le domaine. Dans la section 5.6.4, nous avons présenté les relations que nous avons utilisées dans notre recherche. Dans la présente section, nous parlons de leur application à nos données et nous donnons des exemples pour chacune des relations

utilisées. Afin de mieux les présenter, nous les avons incluses dans les scénarios. Nous rappelons que les scénarios permettent de faire une organisation des cadres dans le domaine. De cette façon, les cadres présents dans un scénario donnent une image plus complète des activités relatives au domaine.

### **6.3.3.1 Scénarios conceptuels et relations entre cadres**

Dans notre travail, nous avons découvert sept scénarios conceptuels : Computer\_setup\_scenario, Activity\_scenario, Transfer\_scenario, Searching\_scenario, Placing\_scenario, Product\_development\_scenario, Change\_scenario et Access\_scenario. Ces scénarios décrivent un certain nombre d'opérations en informatique. Les cadres regroupés concernent des activités comme le téléchargement, l'envoi et la réception des informations, la mise à jour et le changement qu'on peut apporter au système et, enfin, l'accès aux informations. Dans ce qui suit, nous décrivons plus en détails trois scénarios. Nous signalons que la découverte des scénarios conceptuels s'appuie sur des propriétés partagées par certains cadres dans le domaine de l'informatique.

#### *Scénario 1 : Computer\_setup\_scenario*

Le premier scénario que nous présentons est celui de Computer\_setup\_scenario. Ce scénario conceptuel renferme des activités qui consistent à assembler et à configurer un ordinateur, sur les plans matériel et logiciel. Les cadres regroupés dans ce scénario et les relations qu'ils entretiennent avec le scénario sont énumérés dans le tableau 31.

<b>Computer_setup scenario</b>			
<b>Définition : A process through which an Agent (user) sets up a Patient (computer system).</b>			
<b>Cadre</b>	<b>UTP</b>	<b>Relation avec le scénario</b>	<b>Relation avec d'autres cadres</b>
<b>Hardware_setup</b>	<b>ASM :</b> tavbyt1 (تنبييت) (installer1) <b>Anglais :</b> install1 <b>Français :</b> installer1	Is subframe	<b>See also :</b> Installing
<b>Installing</b>	<b>ASM :</b> vab~ata2 (ثبيت) (installer), tavbyt2 (تنبييت) (installer) <b>Anglais :</b> install2 <b>Français :</b> installer2	Is subframe	<b>See also :</b> Hardware_setup <b>Is used by:</b> Placing
<b>Placing</b>	<b>ASM :</b> saj~ala1 (سجل) (enregistrer), tasjyl1 (تسجيل) (enregistrement), kitAbap (كتابة) (écriture), kataba1 (كتب) (écrire1) Adxala1 (أدخل) (entrer), AidxAl (إدخال) (entrée) <b>Anglais :</b> load1b, save1, write1 <b>Français :</b> changer2, enregistrer1, écrire1	Is subframe	<b>Uses:</b> Installing, Storing <b>Precedes:</b> Removing <b>Is opposed to :</b> Removing
<b>Activity_start</b>	<b>ASM :</b> naf~a*a1 (نفذ) (exécuter), tanfy*1 (تنفيذ) (exécution), ta\$gyl2b (تشغيل) (exécution), \$ag~ala1 (شغل) (actionner), <b>Anglais :</b> execution2b, run2b, <b>Français :</b> exécution2b, exécuter2b,	Is subframe	<b>Precedes :</b> Activity_resume <b>Is inherited by :</b> Name_conferral. Choosing, Intentionally_create
<b>Process_start</b>	<b>ASM :</b> ta\$gyl21 (تشغيل) (exécution), Eamal2	Is subframe	<b>Precedes :</b> Process_resume <b>Is prespectivized</b>

	(عمل) (tourner) <b>Anglais</b> : execute2a		in : Activity_start
<b>Information_display</b>	EaraDa1b (عرض) (afficher), EarD1b.1 (عرض) (affichage)	Is subframe	<b>Uses</b> : Transmitting

Tableau 31 : Scénario Computer\_setup

Comme montré dans le tableau, le scénario Computer\_setup\_scenario est composé de plusieurs cadres, à savoir Hardware\_setup, Installing, Placing, Activity\_start, Process\_start et Information\_display. Ces cadres sont tous reliés au scénario au moyen d'une relation Is subframe. Cette relation relie les cadres qui constituent des composants d'un cadre plus complexe. Autrement dit, ces cadres décrivent le processus d'installation d'un système informatique. Le schéma du scénario rend explicite le processus de :

- installer les composants matériels (carte mère, carte son, carte graphique, etc.);
- installer les composants logiciels (système d'exploitation, programmes et logiciels);
- entrer des données;
- commencer le fonctionnement du système; et finalement,
- afficher les informations.

En ce qui concerne les cadres eux-mêmes, ils entretiennent des relations avec d'autres cadres. Ces relations sont aussi le reflet d'autres aspects du domaine. Le cadre Hardware\_setup est relié au cadre Installing par une relation See also. Cette relation signifie que ces deux cadres sont très proches sémantiquement. La seule différence constatée est le fait que Hardware\_setup concerne l'installation des composants matériel, tandis que Installing concerne l'installation des composants logiciels.

Les cadres Installing et Placing sont reliés par une relation Uses et Is used by. Cette relation signifie que Placing utilise une partie de la scène évoquée par Installing. Il s'agit des données inscrites dans le système (le fait d'archiver, d'écrire ou d'entrer) par rapport au fait de placer des composants logiciels dans le système. Placing, quant à lui, est relié par une relation Is opposed to avec le cadre Removing. Ces deux cadres expriment deux scènes opposées : le fait d'inscrire des données et le fait de les supprimer. De plus, nous remarquons une autre relation qui relie ces deux cadres, à savoir Is preceded by. Nous considérons que Placing précède la suppression des données.

La même relation existe entre Activity\_start et Activity\_resume. Ces deux cadres renferment des unités comme *Sag~ala* (شغل) (actionner) et *iEAdap ta\$gyl<sub>1b</sub>* (إعادة تشغيل) (redémarrage). Le cadre Process\_start est analysé dans les sous-sections suivantes. En ce qui concerne le cadre Information\_display, il est relié au cadre Transmitting au moyen d'une relation Uses. Cette relation signifie que Information\_display, qui dénote l'activité d'affichage des informations sur un écran, fait appel à une partie du cadre Transmitting. Cette partie est le fait que l'affichage d'information implique la transmission des données à un dispositif d'affichage.

Dans le scénario analysé, nous avons deux cadres qui font partie d'autres scénarios. Ces cadres sont Placing et Activity\_start. Les deux scénarios en question sont Placing\_scenario et Activity\_scenario. Dans ce qui suit, nous décrivons ces deux scénarios en nous attardant sur les cadres qu'ils renferment.

### *Scénario 2 : Placing\_scenario*

Ce scénario est composé des cadres suivants : Installing, Placing et Removing. Il concerne le processus qui consiste à placer ou à supprimer un composant logiciel d'un ordinateur ou d'un système (tableau 32).



Placing_scenario	
<b>Définition</b>	Agent (user) places Patient (data) in Destination (disk, storage medium or system)
<b>Cadres membres</b>	<b>Relation avec le scénario</b>
<b>Placing</b>	Is perspectivized in
<b>Installing</b>	Is used by
<b>Removing</b>	Is perspectivized in

Tableau 32 : Placing\_scenario

Placing est relié au scénario via une relation Is perspectivized in. Cette relation signifie qu'un scénario peut être envisagé de deux points de vue différents. Le cadre représente un de ces points de vue. Nous remarquons le même type de relation entre le scénario est le cadre Removing. Placing est défini comme le procédé selon lequel un utilisateur (Agent) place des données (Patient) sur un support de stockage (Destination). Quant à Removing, il s'agit de supprimer des données (Patient) d'un support de stockage (Source). En outre, ces deux cadres sont reliés par une relation d'opposition. Finalement, le cadre Installing partage une relation Is used by avec le scénario. Le cadre Installing concerne l'activité de mettre un composant logiciel sur une plateforme. Donc, nous avons estimé que cette activité a un rapport avec une partie du schéma général spécifié par le scénario.

### *Scénario 3 : Activity\_scenario*

Le dernier exemple est Activity\_scenario. Ce scénario concerne les activités qui consistent à manipuler un composant informatique. Cette manipulation peut être effectuée par un utilisateur ou par un composant. De plus, le scénario englobe des états et des activités. Considérons tableau 33.

<b>Activity_scenario</b>	
<b>Définition</b>	Agent (user) or Instrument (système or computer) runs a Patient (operation).
<b>Cadres membres</b>	<b>Relation avec le scénario</b>
<b>Activity_start</b>	Is subframe
<b>Activity_resume</b>	Is subframe
<b>Being_in_operation</b>	Is used by
<b>Being_off_operation</b>	Is used by
<b>Process_start</b>	Is perspectivized in
<b>Process_resume</b>	Is perspectivized in

Tableau 33 : Activity\_scenario

Activity\_scenario est probablement le scénario le plus complexe que nous avons découvert. Ce scénario regroupe plusieurs cadres, à savoir Activity\_start, Activity\_resume, Being\_in\_operation, Being\_off\_operation, Process\_start et Process\_resume. Comme montré dans le tableau 31, trois relations dominent le scénario, à savoir Is subframe, Is used by et Is perspectivized in. La première relation concerne les cadres Activity\_start, Activity\_resume. Ces deux cadres décrivent une activité selon laquelle l'utilisateur fait tourner, ou refait tourner, un composant informatique, notamment un système informatique. Nous remarquons que les deux cadres expriment une activité précise où l'on peut tracer une partie de la scène générale.

En ce qui concerne la deuxième relation, Is used by, les cadres reliés par cette relation reflètent l'état d'une activité, comme illustré dans les deux exemples suivants.

[1] *fy Hal AqlaEa AlniZAm iLY AlwaDE Al|min wa kAn AliqlaE saryEaN, faha\*A yaEny An~a niZAm Alta\$gyl yaEmal bi\$akl salym*

في حال أفلح النظام إلى الوضع الآمن وكان الإقلاع سريعاً ، فهذا يعني أن نظام التشغيل يعمل بشكل سليم

*Si le système a démarré en mode sans échec et le démarrage était rapide, cela veut dire que le système d'exploitation fonctionne bien.*

[2] *i\*A kAn AlqurSu maly}aN fin~a Alwindwz sayatawaqaf Ean alEamal bikafa}ap*

إذا كان القرص مليئاً فإن الويندوز سيتوقف عن العمل بكفاءة

*Si le disque est plein, alors Windows arrêterait de fonctionner comme il faut*

Dans le cas de la relation *Is used by*, les cadres concernés sont reliés au scénario, puisqu'ils expriment l'état de fonctionnement (ou une activité) en cours. Cet état évoqué se trouve en partie dans le scénario.

La dernière relation est *Is perspectivized in*. Cette relation concerne les cadres *Process\_start* et *Process\_resume*. La raison pour laquelle ces deux cadres sont reliés au scénario par cette relation est que les deux expriment une autre perspective d'une activité, à savoir le Patient qui prend en charge l'action, contrairement aux cadres *Activity\_start* et *Activity\_resume*, où c'est l'Agent qui se trouve à l'origine de l'action.

## 6.4 DiCoInfo : A Framed Version

Les données présentées et décrites sont toutes encodées dans une ressource terminologique que nous avons appelée *DiCoInfo : A Framed Version*<sup>65</sup>. Cette ressource a pour but de présenter à la fois nos données et le fruit de notre réflexion sur le domaine étudié.

### 6.4.1 Page d'accueil

Dans la page d'accueil, l'utilisateur peut avoir une idée du contenu et de la nature de la ressource. Nous avons ajouté un tableau explicatif qui renseigne sur la nature du travail. De plus, l'interface permet un accès rapide à toutes les données comprises dans la ressource : les cadres, les scénarios, les UTP et un lien vers le DiCoInfo (figure 93).

---

<sup>65</sup> La ressource est accessible à l'adresse : <http://olst.ling.umontreal.ca/dicoinfo/framed/>. La programmation de l'interface a été réalisée par Benoît Robichaud.

## DiCoInfo: A Framed Version

DiCoInfo  
Le dictionnaire fondamental  
de l'informatique et de l'internet

Frames Participants TUs

Go

- 2016-10-04 : Change\_scenario
- 2016-10-04 : Activity\_scenario
- 2016-04-12 : Computer\_setup
- 2016-02-17 : Moving\_text\_into
- 2016-02-16 : Text\_editing

### Welcome

This is the framed version of the DiCoInfo. In the present resource, we propose a representation of the terminology of computing based on the FrameNet project (Ruppenhofer et al. 2010). Terminological units are organized and described in semantic frames.

For example, the frame *Program creation* describes a situation in which a programmer (**Agent**) develops an application (**Patient**) using a programming language (**Material**). To this frame, we relate a number of terminological units, such as *to write* and *to program*. These two units have the same number of obligatory participants (also called actants). Participants are also of the same nature.

Being a multilingual resource, you can find units in different languages such as *programmer*, *écrire*, *كتابة* and *برمجة*.

This project is carried out at the *Observatoire de Linguistique Sens-Texte* of the University of Montréal. Ideas, data, conception and programming are under the responsibility of Nizar Ghazzawi, Marie-Claude L'Homme and Benoît Robichaud.

Feel free to contact us for any comments or critiques at the following address: [nizar.ghazzawi@umontreal.ca](mailto:nizar.ghazzawi@umontreal.ca)

© OLST 2015

Figure 93 : Page d'accueil de DiCoInfo : A Framed Version

La ressource est en langue anglaise. Notre choix de l'anglais tient au fait que ce travail veut avoir une portée multilingue. Donc, nous avons trouvé essentiel que cette ressource soit présentée dans une langue qu'un grand nombre d'utilisateurs peut comprendre.

### 6.4.2 Recherche dans Framed Version

Dans la Framed Version, l'utilisateur a la possibilité de rechercher par cadre, par terme et même par participant (actant) (figure 94).

## DiCoInfo: A Framed Version

DiCoInfo  
Le dictionnaire fondamental  
de l'informatique et de l'internet

Frames Participants TUs

Placing

Go

- 2016-10-04 : Change\_scenario
- 2016-10-04 : Activity\_scenario
- 2016-04-12 : Computer\_setup
- 2016-02-17 : Moving\_text\_into
- 2016-02-16 : Text\_editing

Placing

الحريرية

د	1	2	3	4	5	6	7	8	9	0	=	bksp
Tab	ص	ض	ط	ق	ف	ع	غ	ه	خ	ج	ح	\
Caps	ش	س	ب	ل	ا	ب	ت	ث	ج	د	ط	Enter
Shift	ي	ة	و	ز	ة	ي	ك	ر	ة	ي	ظ	Shift

Alt v1.49

Figure 94 : Recherche dans DiCoInfo : A Framed Version

La figure montre une recherche par cadre, Placing. Tout comme dans le DiCoInfo version ASM, la Framed Version propose un clavier virtuel pour faciliter la saisie des caractères arabes. De plus, la recherche intuitive propose à l'utilisateur les options disponibles lorsque la recherche correspond à plusieurs entrées dans la ressource.

### 6.4.3 Affichage des résultats

Le résultat d'une requête typique prend la forme illustrée à la figure 95 pour le cadre Placing.

<b>Definition:</b> <b>Agent</b> (user) places <b>Patient</b> (data) on <b>Destination</b> (storing unit).		<b>Relations:</b> 
<b>Example(s):</b> • [ar] البيانات يتم تسجيلها على القرص الصلب من خلال توليد شحنات مغناطيسية • [en] <b>You</b> can <b>WRITE</b> and rewrite to a <b>CD-RW disc</b> over 1,000 times. • [fr] <b>Les messages</b> peuvent <b>ÊTRE ENREGISTRÉS</b> sur <b>disque dur</b>		
<b>Note(s):</b> • This frame is based on Placing in FrameNet.		
<b>Participants (1):</b> 1. <b>Agent</b> 2. <b>Patient</b> 3. <b>Destination</b>	<b>Participants (2):</b> 1. Departure_point : Méthode 2. Frequency : l. Result 3. Instrument : l. Resultat 4. Lieu : l. Source 5. Manière : Temps 6. Manner : Time	
<b>Arabic TUs:</b> • سجل.1 • تسجيل.1 • كتابة.1 • أدخل.1 • إدخال.1 • كتب.1	<b>English TUs:</b> • load.1b • save.1 • write.1	<b>French TUs:</b> • charger.2 • enregistrer.1 • écrire.1

Figure 95 : Vue d'ensemble du résultat de recherche dans Framed Version

Chacune des parties présentées dans la figure donne des renseignements différents. La première case présente la définition du cadre. La définition indique les actants, mis en évidence par des couleurs. Nous avons fait en sorte que les définitions dans notre ressource soient courtes et concises afin de faciliter la compréhension de l'idée générale du cadre. De plus, la présence des actants que le cadre renferme dans la définition permet de mieux

percevoir les rôles qu'ils jouent pour chacune des UTP. Dans la figure, nous donnons l'exemple du cadre Placing. La définition met en évidence les trois actants (figure 96).

**Definition:**

**Agent** (user) places **Patient** (data) on **Destination** (storing unit).

Figure 96 : Définition d'un cadre dans Framed Version

Après la définition, trois exemples en trois langues sont donnés. Bien entendu, les exemples comportent une UTP faisant partie du cadre et les actants sont soulignés au moyen des couleurs (figure 97).

**Example(s):**

- [ar] البيانات يتم تسجيلها على القرص الصلب من خلال توليد شحنات مغناطيسية
- [en] **You** can **WRITE** and rewrite **to a CD-RW disc** over 1,000 times.
- [fr] **Les messages** peuvent **ÊTRE ENREGISTRÉS sur disque dur**

Figure 97 : Exemples dans Framed Version

Le fait de donner des exemples dans les trois langues permet à l'utilisateur de mieux comprendre comment les UTP se réalisent dans chacune des langues. Dans la figure 97, le contexte en ASM est donné pour l'UTN *tasjyl*<sub>1</sub> (تسجيل) (enregistrer), tandis que les contextes anglais et français présentent les UTV *to write*<sub>1</sub> (en) et *enregistrer*<sub>1</sub> (fr). Dans le choix des exemples, nous optons pour les contextes où tous les actants sont présents, dans la mesure du possible. Finalement, pour les contextes, les UTP sont en caractères gras, et dans le cas de l'anglais et du français, elles sont en majuscule.

Sous les exemples, une section contient une note qui précise si le cadre est basé sur FrameNet ou non (dans ce cas, le cadre est précisé et il renvoie au cadre d'origine dans FrameNet). Dans le cas du cadre analysé, il s'agit d'un cadre basé sur un cadre existant dans FrameNet.

Ensuite, deux listes comportant les actants et les circonstants sont donnés. Bien entendu, la liste « Participant (2) » ne concerne pas les données de l'ASM, puisque les circonstants ne font pas partie de notre analyse (figure 98).

Participants (1):	Participants (2):
<ol style="list-style-type: none"><li>1. <b>Agent</b></li><li>2. <b>Patient</b></li><li>3. <b>Destination</b></li></ol>	<ol style="list-style-type: none"><li>1. <b>Departure_point</b></li><li>2. <b>Frequency</b></li><li>3. <b>Instrument</b></li><li>4. <b>Lieu</b></li><li>5. <b>Manière</b></li><li>6. <b>Manner</b></li><li>7. <b>Méthode</b></li><li>8. <b>Result</b></li><li>9. <b>Resultat</b></li><li>10. <b>Source</b></li><li>11. <b>Temps</b></li><li>12. <b>Time</b></li></ol>

Figure 98 : Participant 1 et Participant 2 dans la Framed Version

Finalement, dans l'avant-dernière section, nous énumérons toutes les UTP qui font partie du cadre (figure 99).



Figure 99 : Listes des UTP faisant partie d'un cadre

Toutes ces UTP sont des renvois. En cliquant, l'utilisateur est redirigé vers son entrée dans le DiCoInfo.

La dernière section décrit les relations que le cadre partage avec d'autres cadres (figure 100).

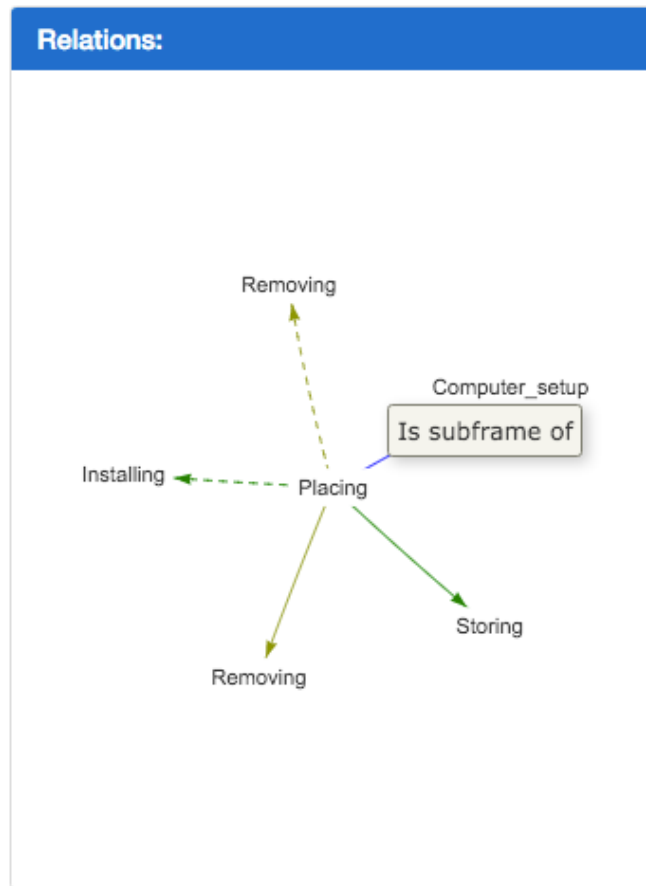


Figure 100 : Relations entre cadres dans la Framed Version



Les relations entre les cadres sont affichées au moyen des courbes. Pour distinguer les différentes relations, il suffit de placer le curseur sur la courbe pour faire afficher le nom de relation.

## Conclusion

La présente recherche avait pour objectif l'étude, l'analyse et la description des unités terminologiques prédicatives (UTP) verbales et déverbales du domaine de l'informatique en vue de la réalisation d'une ressource terminologique et la modélisation sémantique de ces unités. La langue visée par notre recherche est la langue arabe, plus particulièrement l'*arabe standard moderne* (ASM). Nous avons basé notre analyse sur la théorie de la Sémantique des cadres de Fillmore (1976, 1977, 1982, 1985).

Dans notre recherche, nous avons adapté et mis en œuvre une méthodologie de compilation des UTP dans une ressource terminologique. Cette méthodologie consiste à compiler un corpus en ASM du domaine de l'informatique. Les sources de notre corpus viennent principalement du web. Pour l'extraction des unités terminologiques verbales et nominales (UTV et UTN), nous avons intégré l'ASM à un extracteur automatique des termes, TermoStat (Drouin 2004). À la suite de l'extraction, nous avons validé le statut terminologique des unités obtenues avec les critères de validation terminologique de L'Homme (2004). Après cette étape, nous avons associé des contextes provenant de notre corpus à chaque unité validée. Finalement, nous avons encodé les données dans des fiches terminologiques avec un éditeur XML.

Cette ressource constitue la base d'une version ASM d'une ressource terminologique multilingue existante, à savoir le DiCoInfo (L'Homme 2008). Notre ressource se distingue par le fait qu'elle est conçue pour les UTP de l'ASM. Donc, nous y avons compilé les unités prédicatives verbales et déverbales. Les unités faisant partie de cette première ressource ont fait l'objet d'une modélisation plus complexe en fonction de cadres sémantiques, suivant le modèle de FrameNet (Ruppenhofer et al. 2010). Cette modélisation a fait l'objet d'une deuxième ressource que nous avons appelée DiCoInfo : A Framed Version. Elle est conçue pour les cadres sémantiques que nous avons découverts pour le domaine de l'informatique à partir des données que nous avons obtenues en ASM.

Au total, nous avons obtenu 106 UTP, soit 47 UTV et 59 UTN. Ces unités ont fait l'objet d'une description complète, en ce qui concerne leur structure actancielle et le type d'actants qu'elles admettent. Cette description préalable nous a permis d'aller à la découverte

de cadres sémantiques. Notre méthodologie pour la découverte des cadres sémantiques consiste, en partie, à trouver des cadres qui correspondent à nos données dans FrameNet. Toutefois, dans le cas où les cadres ne correspondent pas à nos données, nous procédons à la création de nos propres cadres. Nous avons ainsi obtenu 57 cadres sémantiques. Ces cadres sont regroupés au moyen de relations (un total de 11 relations). De plus, sept scénarios conceptuels ont été découverts réunissant des cadres qui portent sur des activités et des procédés dans le domaine de l'informatique.

Comme nous l'avons signalé à plusieurs reprises dans notre recherche, nous avons à faire face à un certain nombre de difficultés. Ces difficultés sont de deux natures : des difficultés liées à la morphologie de l'ASM et des difficultés liées aux outils de traitement automatique de l'ASM. La question des difficultés reste ouverte pour les raisons suivantes. D'abord, la morphologie de l'ASM est complexe et riche. Cela fait partie des défis qu'un linguiste s'intéressant à la langue arabe en général, et à l'ASM en particulier, doit relever. En ce qui concerne la terminologie, ces défis se multiplient, surtout lorsqu'il s'agit de travailler sur des unités comme les verbes. L'ambiguïté morphologique de la forme verbale dans les textes constitue un obstacle réel à son repérage et a fortiori à son repérage automatique. En ce qui concerne les outils de traitement automatique, même s'ils peuvent proposer des résultats intéressants pour la langue générale, nous estimons qu'il reste encore fort à faire en terminologie.

En ce qui a trait aux résultats de notre recherche, nous constatons que nos résultats confirment nos hypothèses. Nous avons trouvé que l'ASM peut s'adapter à une méthodologie de travail terminologique particulière. Toutefois, dans le traitement de l'ASM, un certain nombre de points doit être considéré. Premièrement, une méthodologie comprenant une extraction automatique des termes doit prendre en compte la particularité morphologique de l'ASM. Nous avons évoqué la question de la complexité et de la richesse du système morphologique de la langue arabe et les défis que cette complexité pose aux linguistes. Deuxièmement, lorsqu'il est question des unités morphologiquement ambiguës comme les verbes, une attention spéciale doit être portée à la taille du corpus. Un corpus plus large que celui auquel nous avons fait appel pourrait probablement assurer une plus grande présence de ces unités. Troisièmement, afin de faciliter le travail du terminologue, les plateformes

informatiques doivent être compatibles avec les caractères arabes, surtout en ce qui concerne le sens de l'écriture qui, parfois, n'est pas pris en compte.

En ce qui concerne notre deuxième hypothèse, nos données montrent que les UTP de l'ASM se distinguent par la présence de certaines unités qui n'ont pas d'équivalents dans les versions anglaise et française du DiCoInfo. Nous rappelons que nous avons eu un total de 14 verbes et de 19 noms sans équivalents dans le DiCoInfo. Ces unités expriment des activités différentes dans le domaine. De plus, nous avons trouvé des unités propres à l'ASM. Nous avons parlé de l'UTV *tawAfaqa* (توافق) (être compatible), qui n'a pas d'équivalent verbal direct en anglais ou en français. Son équivalent est une unité adjectivale dans le DiCoInfo. Enfin, nous avons remarqué que les UTN de l'ASM ont moins d'équivalents anglais, tandis qu'une similitude est constatée avec les unités françaises.

Notre troisième hypothèse est également confirmée par les résultats de notre recherche. Les UTP de l'informatique en ASM partagent des points communs avec leurs équivalents anglais et français. Pour ces équivalents, nous n'avons pas relevé de cas où une unité ASM a une structure actancielle différente de celle de ses équivalents dans les autres langues. De plus, les UTP de l'ASM expriment les mêmes réalités extralinguistiques. Cela est sans doute attribuable au fait que, dans un domaine technique, dominé en grande partie par la langue anglaise, la réalisation des concepts est partagée par toutes les langues. L'ASM de spécialité ne représente pas un cas exceptionnel dans ce sens. En outre, les critères de validation terminologique de L'Homme (2004) sont applicables aux unités prédicatives de l'ASM.

En ce qui concerne les cadres, la théorie de la Sémantique des cadres est une théorie qui peut être adaptée à l'étude d'un domaine de spécialité. Dans la section 6.3.1, nous avons présenté les résultats du regroupement des UTP dans des cadres. Nous avons montré qu'un certain nombre de cadres existants dans FrameNet sont exploitables pour notre recherche. Nous rappelons que nous avons découvert 41 cadres dans FrameNet qui correspondent au domaine de l'informatique (20 cadres avec correspondance parfaite, 6 cadres avec correspondance semi-parfaite et 15 cadres avec correspondance partielle). Bien entendu, puisque FrameNet est conçu pour la langue générale, ces cadres sont adaptés à la langue de spécialité. Quant aux relations, elles ont subi une adaptation pour convenir au domaine. Nous

avons fait en sorte que les relations reflètent les liens entre les différentes activités. Cela s'applique surtout aux nouveaux cadres, comme *Product\_development* et *Program\_creation*. Comme nous avons intégré certaines UTP anglaises et françaises, nous avons découvert que l'ASM peut aussi se distinguer sur certains points. Tout comme les UTP sans équivalents (section 6.1.4.3), nous avons aussi des cadres qui ne renferment que des UTP en ASM. Par exemple, un cadre comme *Inspecting* ne comprend que les UTV *faHaSa* (فحص) (inspecter) et *tafaH~aSa* (تفحص) (analyser).

Notre travail est essentiellement monolingue, c'est-à-dire que nous avons réalisé nos analyses à partir des termes en ASM. Toutefois, nous avons trouvé important d'inclure les équivalents anglais et français dans notre version du DiCoInfo et dans la Framed Version de la même ressource. Dans la section 5.5.8, nous avons abordé brièvement la question des équivalents, et dans la section 5.6.3, nous avons parlé de notre deuxième critère de choix de cadre, qui concerne l'inclusion des UTP d'autres langues. Nous estimons qu'une ressource terminologique conçue pour l'ASM doit impérativement comporter des équivalents. L'ASM est une langue importatrice de terminologie (au moins pour les domaines techniques). Pour cette raison, l'utilisateur qui consulte une ressource spécialisée de l'informatique s'attend à ce que les termes équivalents existent aux côtés des termes en ASM. De plus, dans le cas de l'informatique, les termes de référence sont les termes anglais. Cela signifie que l'utilisateur recherche le terme anglais d'abord. Dans la section 2.4.1, nous avons évoqué les problèmes et les défis du travail terminologique dans l'espace arabophone. Parmi les défis les plus importants se trouve la question de la standardisation et du transfert des termes. Par conséquent, et en ce qui concerne la terminologie technique, l'ASM s'intègre mieux dans un environnement multilingue.

Ce que nous avons présenté dans notre recherche constitue un premier pas vers un travail plus étendu. Les résultats obtenus reflètent notre corpus et ses propriétés. Nous sommes conscients des limites imposées par nos données. Travailler sur des termes prédicatifs de l'ASM nécessite un corpus plus conséquent, surtout lorsqu'il est question des termes verbaux. Bien entendu, un corpus plus important n'est pas toujours la solution idéale pour un travail terminologique (Bowker et Pearson 2002 : 45). Toutefois, il serait capital de régler certains problèmes que nous avons rencontrés lors de la compilation de corpus. Nous avons mentionné

à la section (5.2.4.2) que la conversion des documents PDF en format texte brut (.txt) n'était pas une tâche facile à réaliser. Certains documents ont dû être écartés en raison de problèmes de conversion. Cette question relève également des problèmes techniques et des outils de traitement automatique discutés plus tôt dans cette section.

En outre, notre travail s'inscrivait dans une démarche visant à combler les lacunes dans les ressources spécialisées de l'ASM de spécialité. Premièrement, nous souhaitons présenter une méthodologie adéquate pour la compilation des termes. Notre recherche met l'accent sur le fait qu'une simple traduction des termes n'est pas toujours une méthodologie efficace pour la conception d'une ressource spécialisée. Deuxièmement, étant donné que notre ressource est basée sur corpus, nous avons voulu qu'elle soit une sorte de reflet de l'usage réel des termes. Troisièmement, notre travail souhaitait contribuer à l'adaptation des outils de traitement automatique destinés à la terminologie. Pour cette raison, nos réalisations, en ce qui concerne le corpus de spécialité et le corpus général, constituent des exemples importants pour les futures recherches en terminologie et en terminologie computationnelle.

Il reste évidemment un certain nombre de pistes que nous souhaitons explorer. Tout d'abord, nous envisageons d'inclure un plus grand nombre d'UTP, afin de couvrir d'autres activités dans le domaine. Il sera important à ce stade d'inclure aussi les unités prédicatives autres que verbales et déverbales, à savoir les unités prédicatives adjectivales et adverbiales. Ensuite, la prise en compte des termes complexes constituerait une étape à considérer, surtout que ceux-ci sont susceptibles d'être majoritaires dans le domaine de l'informatique. Finalement, l'aspect multilingue reste un aspect d'une grande importance à enrichir dans un travail futur. Notre ressource en cadres, DiCoInfo : A Framed Version, recevra une attention particulière en ce qui concerne l'inclusion d'un plus grand nombre d'unités anglaises et françaises.

## Bibliographie

- Abed, A. M., S. Tiun and M. Albared (2013). "Arabic Term Extraction Using Combined Approach on Islamic Document." In *Journal of Theoretical & Applied Information Technology*, 58(3), p. 601-608.
- Abdelali, A., J. Cowie and H. S. Soliman (2005). "Building a Modern Standard Arabic Corpus." In *Workshop on computational modeling of lexical acquisition*, Croatia, p. 1-7.
- Adolphs, S. and P. M. S. Lin (2011). "Corpus Linguistics." In Simpson, J. (eds.) *The Routledge handbook of applied linguistics*. London/New York: Routledge, p. 597-610.
- Alansary, S., M. Nagi and N. Adly (2007). "Building an International Corpus of Arabic (ICA): Progress of Compilation Stage." In *7th International conference on language engineering*, Cairo, Egypt, p. 1-30.
- Alfaifi, A. Y. G., E. Atwell, and I. Hedaya (2014). "Arabic Learner Corpus (ALC) v2: a New Written and Spoken Corpus of Arabic Learner." In *Proceedings of Learner Corpus Studies in Asia and the World 2014*, 2, p. 77-89.
- Al Khatib, K. and A. Badarneh (2010). "Automatic Extraction of Arabic Multi-Word Terms." In *Proceedings of the International Multiconference on Computer Science and Information Technology (IMCSIT)*, Wisla, Poland, p. 411-418.
- Al-Safadi, L., M. Al-Badrani and M. Al-Junidey (2011). "Developing Ontology for Arabic Blogs Retrieval." In *International Journal of Computer Applications*, 19(4), p. 40-45.
- Alshehri, A. M. (2014). *The Frame Semantics of 'SelfMotion' Frame in Arabic and English*, Thèse de doctorat, San Francisco: San Francisco State University.
- Al-Sultani, L. and E. Atwell (2006). "The Design of a Corpus of Contemporary Arabic." In *International Journal of Corpus Linguistics*, 11(2), p. 135-171.
- Andor, J. (2010). "Discussing Frame Semantics: The State of the Art: An Interview with Charles J. Fillmore." In *Review of Cognitive Linguistics*, 8 (1), p. 157.
- Arabic Propbank (2015). Arabic PropBank Framefiles. Repéré à <https://verbs.colorado.edu/propbank/framesets-arabic/>

- Arabic VerbNet (2015). Home Page. Repéré à [http://ling.uni-konstanz.de/pages/home/mousser/files/Arabic\\_verbnet.php](http://ling.uni-konstanz.de/pages/home/mousser/files/Arabic_verbnet.php)
- ARABTERM. (2015). The Methodology in the Creation of Unified Dictionaries. Repéré à <http://www.arabterm.org/index.php?id=72&L=1>
- AraCorpus. (2015). Arabic Corpora Resource. Repéré à <http://aracorporus.e3rab.com/index.php?content=english>
- Araúz, L., Pilar, P. Faber and S. Montero (2012). “Specialized Language Semantics.” In Faber, P. (dir.) *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin/New York: Mouton de Gruyter, p. 95–176.
- Arts, T., Y. Belinkov, N. Habash, A. Kilgarriff and V. Suchomel (2014). “arTenTen: Arabic Corpus and Word Sketches.” In *Journal of King Saud University-Computer and Information Sciences*, 26(4), p. 357-371.
- Atkins, S., C. J. Fillmore and C. Johnson (2003). “Lexicographic Relevance: Selecting Information from Corpus Evidence.” In *International Journal of lexicography*, 16 (3), p. 251-280.
- Attia, M., P. Pecina, A. Toral, L. Tounsi and J. van Genabith (2011). “A Lexical Database for Modern Standard Arabic Interoperable with a Finite State Morphological Transducer.” In *Systems and Frameworks for Computational Morphology*. Berlin, Heidelberg: Springer, p. 98-118.
- Atwel, E. and A. Sharaf (2009). “Knowledge Representation of the Quran through Frame Semantics. A Corpus-based Approach.” In Mahlberg, M. V. González-Díaz and C. Smith (dir), *Proceedings of the corpus linguistics conference*, Great Britain: University of Liverpool. Repéré à <http://ucrel.lancs.ac.uk/publications/cl2009/>
- Barsalou, L.W. (2003). “Situated Simulation in the Human Conceptual System.” In *Language and Cognitive Processes*, 18, p. 513–62.
- Barsalou, L. W. (2008). “Grounded Cognition.” In *Annual Review of Psychology*, 59, p. 617–645.
- Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge university press.
- Bearman, P.J., Th. Bianquis, C.E. Bosworth, E. van Donzel and W.P. Heinrichs (dir.) (2002). *Encyclopædia of Islam*. Netherlands: Brill Academic Publishers.



- Belkredim, F. Z. and F. Meziane (2008). "DEAR-ONTO: A Derivational Arabic Ontology Based on Verbs." In *International Journal of Computer Processing of Languages*, 21(03), p. 279-291.
- Belkredim, F. Z. and A. El Sebai (2009). "An Ontology Based Formalism for the Arabic Language Using Verbs and Their Derivatives." In *Communications of the IBIMA*, 11(5), p. 44-52.
- Black, W., S. Elkateb, H. Rodriguez, M. Alkhalifa, P. Vossen, A. Pease and C. Fellbaum (2006). "Introducing the Arabic Wordnet Project." In *Proceedings of the Third International WordNet Conference*, p. 295-300.
- Blau, J. (1981). "The State of Research in the Field of the Linguistic Study of Middle Arabic." In *Arabica*, 28(Fasc. 2/3), p. 187-203.
- Boas, H. C. (2009). "Recent Trends in Multilingual Computational Lexicography." In Boas, H. C. (dir.) *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Berlin/New York: Mouton de Gruyter, p. 1-36.
- Boulaknadel, S., B. Daille and D. Aboutajdine (2008). "A Multi-Word Term Extraction Program for Arabic Language." In *Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Maroc, p. 1485-1488.
- Bowker, L. and J. Pearson (2002). *Working with Specialised Language: A Practical Guide to Using Corpora*. Great Britain: Routledge.
- Buckwalter, T. (2002). "Buckwalter Arabic Morphological Analyzer Version 1.0." Linguistic Data Consortium, University of Pennsylvania, 2002. LDC Catalog No.: LDC2002L49.
- Buckwalter, T. (2016). Arabic Transliteration. Repéré à <http://www.qamus.org/transliteration.htm>
- Buendia, M. (2013). *Phraseology in Specialized Language and its Representation in Environmental Knowledge Resources*, Thèse de doctorat. Granada: Universidad de Granada.
- Buendia, M., S. M. Martinez and P. Faber (2014). "Verb Collocations and Phraseology in EcoLexicon." In Colson, J-P. (dir.) *Yearbook of Phraseology*, (5)1, Berlin/Munich/Boston: De Gruyter Mouton, p. 57-94.
- Burchardt, A., K. Erk, A. Frank, A. Kowalski, S. Padó and M. Pinkal (2009). "Using FrameNet for the Semantic Analysis of German: Annotation, Representation, and

- Automation.” In Boas, Hans C. (dir) *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Berlin/NewYork: Mouton de Guyter, p. 209-244.
- Cabré, M. T. (1999). *Terminology: Theory, Methods, and Applications*. Amsterdam: John Benjamins.
- Cabré, M. T. (2003). “Theories of Terminology: Their Description, Prescription and Explanation.” *Terminology*, 9(2), 163-199.
- Chomsky, N. (2002). *Syntactic Structures*. Berlin/New York: Walter de Gruyter.
- Cienki, A. (2007). « Frames, Idealized Cognitive Models, and Domains.” In Geeraerts, D. and H. Cuyckens (dir.) *The Oxford handbook of cognitive linguistics*, Oxford University Press: New York, p. 171, 186.
- Claveau, V. and M.-C. L’Homme (2004). “Discovering Specific Semantic Relationships between Nouns and Verbs in a Specialized French Corpus.” In *Proceedings of the 3rd International Workshop on Computational Terminology, CompuTerm’04*, p. 39-46.
- Condamines, A. (1993). « Un exemple d’utilisation de connaissances de sémantique lexicale : acquisition semi automatique d’un vocabulaire de spécialité », *Cahiers de lexicologie*, 62(1), p. 25-65.
- Condamines, A. et D. Bourigault (1999). « Alternance nom/verbe : explorations en corpus spécialisés », dans B. Victorri et J. François (dir.) *Sémantique du lexique verbal, Actes de l’atelier de Caen, Cahiers de l’Elsap*, p. 41-48.
- Côté, R. A. (1996). *Répertoire d’anatomopathologie de la SNOMED internationale*. v3.4. Université de Sherbrooke, Sherbrooke, Québec.
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge University Press: Cambridge.
- Darwish, A. (1988). *Terminology and Translation: A Phonological-semantic Approach to Arabic Terminology*. Writescop Publishers: Australia.
- Diab, M., A. Mansouri, M. Palmer, O. Babko-Malaya, W. Zaghouani, A. Bies and M. Maamouri. (2008). “A Pilot Arabic Propbank.” In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)*, p. 3467-3472.
- DiCoInfo. (2015). Liste alphabétique des termes. Repéré à <http://olst.ling.umontreal.ca/dicoinfo/dicoinfo-bilingue-fr.html>.
- Dictionary of Information Technology. (2000). Damas, Syrie: Syrian Computer Society.

- Dolbey, A., M. Ellsworth and J. Scheffczyk (2006). "BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies." In *KR-MED 2006 Biomedical Ontology in Action* (222), Baltimore, Maryland, p. 87-94.
- Doran, C., D. Egedi, B. A. Hockey, B. Srinivas and M. Zaidel (1994). "XTAG System: A Wide Coverage Grammar for English." In *Proceedings of the 15th conference on Computational Linguistics-Volume 2*, Association for Computational Linguistics, Stroudsburg, PA, USA, p. 922-928.
- Drouin, P. (2002). *Acquisition automatique des termes : l'utilisation des pivots lexicaux spécialisés*, Thèse de Doctorat. Université de Montréal.
- Drouin, P. (2003). "Term Extraction Using Non-Technical Corpora as a Point of Leverage." *Terminology*, 9(1), p. 99-115.
- Drouin, P. (2004). "Detection of Domain Specific Terminology Using Corpora Comparison." In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal, p. 79-82.
- EcoLexicon (2016). Terminological Knowledge Base on the Environment. Repéré à <http://ecolexicon.ugr.es/en/index.htm>
- El-Haj, M. and R. Koulali (2013). "KALIMAT a Multipurpose Arabic Corpus." In Atwell, Eric and A. Hardie (dir.) *Second Workshop on Arabic Corpus Linguistics (WACL-2)*, Lancaster: Lancaster University, p. 1-9.
- Elkateb, S., W. Black, P. Vossen, D. Farwell, H. Rodríguez, A. Pease and M. Alkhalifa (2006a). "Arabic WordNet and the Challenges of Arabic." In *Proceedings of Arabic NLP/MT Conference*, London, UK, p. 15-24
- Elkateb, S., W. Black, H. Rodríguez, M. Alkhalifa, P. Vossen, A. Pease and C. Fellbaum (2006b). "Building a WordNet for Arabic." In *Proceedings of The fifth international conference on Language Resources and Evaluation (LREC)*, p. 29-34.
- El Mahdaouy, A., S. E. Ouatic and E. Gaussier (2013). "A Study of Association Measures and their Combination for Arabic MWT Extraction." In *Terminology and Artificial Intelligence*, Paris, France, p. 45-52.
- Elmgrab, R. A. (2011). "Methods of Creating and Introducing New Terms in Arabic Contributions from English-Arabic Translation." In *Proceedings from the International*

- Conference on Languages, Literature and Linguistics IPEDR vol 26*, Singapore, p. 491-500.
- Ellsworth, M., K. Erk, P. Kingsbury and S. Padó (2004). “PropBank, SALSA, and FrameNet: How Design Determines Product.” In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal, p. 17-23.
- Evans, V. and M. Green (2006). *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Faber, P. and R. Mairal (1999). *Constructing a Lexicon of English Verbs*. Berlin/New York: Mouton de Gruyter.
- Faber, P., C. Márquez Linares and M. Vega Expósito (2005). “Framing Terminology: A Process-Oriented Approach.” *Meta* 50(4). Repéré à <http://id.erudit.org/revue/meta/2005/v50/n4/019916ar.pdf>.
- Faber, P., P. L. Araúz, J. A. P. Velasco and A. Reimerink (2007). “Linking Images and Words: The Description of Specialized Concepts.” *International Journal of Lexicography* 20(1), p. 39-65.
- Faber, P. (2009a). “The Pragmatics of Specialized Communication.” *Entreculturas* 1, p. 61-84.
- Faber, P. (2009b). “The Cognitive Shift in Terminology and Specialized Translation.” *MonTI. Monografías de Traducción e Interpretación*, no. 1, p. 107-134.
- Faber, P. (2011). “The Dynamics of Specialized Knowledge Representation: Simulational Reconstruction or the Perception–Action Interface.” *Terminology* 17 (1) p. 9-29.
- Faber, P. (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin: De Gruyter Mouton.
- Faber, P. (2014). “Frames as a Framework for Terminology.” In Kockaert, H. et F. Steurs (dir.) *Handbook of Terminology*, New York/Amsterdam: John Benjamins, p.14-33.
- Faber, P. and M. Buendía Castro (2014). “EcoLexicon.” In Abel, A., C. Vettori and N. Ralli (dir.) *Proceedings of the XVI EURALEX International Congress*, Bolzano, Italie: Euralex, p. 601–607.
- Fauconnier, G. (1985). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. United States of America: The Massachusetts Institutes of Technology.

- Fellbaum, C. (1990). "English Verbs as a Semantic Net." *International Journal of Lexicography* 3(4), p. 278-301.
- Fellbaum, C and G. Miller (1990). "Folk Psychology or Semantic Entailment? A Reply to Rips and Conrad (1989)." *Psychological Review* 97, p. 565–570.
- Fellbaum, C. (dir.) (1998). *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.
- Fellbaum, C. (2006). "WordNet(s)." In Brown, K. (dir.) *Encyclopedia of Language & Linguistics*, Second Edition, volume 13, Oxford: Elsevier, p. 665-670.
- Fensel, D. (2001). *Ontologies: A Silver Bullet for Knowledge Management and Electronic-Commerce*. Berlin: Springer-Verlag.
- Farreres, J. (2005). *Creation of Wide-Coverage Domain-Independent Ontologies*, Thèse de doctorat. Universitat Politècnica de Catalunya.
- Fillmore, C. J. (1965). *Indirect Object Constructions in English and the Ordering of Transformations*. The Hague: Mouton.
- Fillmore, C. J. (1968). "The Case for Case." In Bach, E. and R. T. Harms (dir.) *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston, p. 1-25.
- Fillmore, J. C. (1975). "An Alternative to Checklist Theories of Meaning." In Cogen, C., H. Thompson, G. Thurgood, K. Whistler and J. Wright (dir.) *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, p. 123-131.
- Fillmore, C. J. (1976). "Frame Semantics and the Nature of Language." In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 280 (1), p. 20-32.
- Fillmore, C. J. (1977). "The Case for Case Reopened." *Syntax and Semantics* 8, p. 59-82.
- Fillmore, C. J. (1982). "Frame Semantics." In *The Linguistic Society in Korea* (dir.), *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co., p. 111-137.
- Fillmore, C. J. (1985). "Frames and the Semantics of Understanding." *Quaderni di Semántica* 6(2), p. 222–254.
- Fillmore, C. J. and B. T. Atkins. (1992). "Toward a Frame-Based Lexicon: The Semantics of RISK and its Neighbors." In Lehrer, A. and E. F. Kittay (dir.) *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*. New York: Routledge, p. 75-102.

- Fillmore, C. J., M. R. L. Petruck, J. Ruppenhofer and A. Wright (2003a). "Framenet in Action: The Case of Attaching." *International Journal of Lexicography* 16(3), p.298–332.
- Fillmore, C. J., C.R. Johnson and M. Petruck (2003b). "Background to FrameNet" *International Journal of Lexicography* 16(3), p. 235-250.
- Fillmore, C. J. (2006). "Frame Semantics." In Geeraerts, D. (dir.) *Cognitive Linguistics: Basic Readings*. Berlin/New York: Mouton de Gruyter, p. 373–400.
- FrameNet. (2015). FrameNet Data. Repéré à [https://framenet.icsi.berkeley.edu/fndrupal/framenet\\_data](https://framenet.icsi.berkeley.edu/fndrupal/framenet_data).
- Ghazzawi, N., M.-C. L'Homme et F. Sadat (à paraître). « Du terme au Frame : la terminologie de l'informatique du point de vue de la Sémantique des cadres », In *Actes de Conférence TOTH*, Chambéry : Institut Porphyre, Savoir et Connaissance.
- Ghneim, N., E. Karhely and W. Safi (2009). "First Step of Building an Arabic FrameNet (AFN)." In *13th IBIMA Conference*, Marrakech, Morocco. Repéré à <http://www.ibima.org/MA2009/papers/nada.html>.
- Gross, M. (1975). *Méthode en syntaxe, régime des constructions complétives*. Paris : Hermann.
- Gross, M. (1981). « Les bases empiriques de la notion de prédicat sémantique », *Langages* 63, p. 7-52.
- Gross, G. (1994). « Classes d'objets et description des verbes », *Langages* 28 (115), p. 15-30.
- Gross, G. et M. Mathieu-Colas (2001). « Description de la langue de la médecine », *Meta* 46(1), p. 68-81.
- Gross G. et R. Vivès (2001). « La description en termes de classes d'objets et l'enseignement des langues », *Langue Française* 131, p. 38–51.
- Gross, G. et F. Guenther (2002). « Comment décrire une langue de spécialité ? », *Cahiers de lexicologie*, 80, Paris : Champion, p. 179-199.
- Gross, G. (2003). "On the Description of Classes of Predicates." *Language Research*, special issue, p. 39-53.
- Gross, G. (2008). « Les classes d'objets », *Lalies* 28, p. 111-165.
- Gross, G. (2010). « Sur la notion de contexte », *Meta* 55(1), p. 187-197.

- Habash, N., A. Souidi and T. Buckwalter (2007). "On Arabic Transliteration." In Souidi, A., G. Neumann and A. V. D. Bosch (dir.), *Knowledge-based and Empirical Methods: Arabic Computational Morphology*. Nederland: Springer, p. 15-22.
- Habash, N., O. Rambow and R. Roth (2009). "MADA+ TOKAN: A Toolkit for Arabic Tokenization, Dacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization." In *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR)*, Cairo, Egypt, p. 102-109.
- Habash, N. Y. (2010). *Introduction to Arabic Natural Language Processing*. Synthesis Lectures on Human Language Technologies, 3(1), Toronto: Morgan and Claypool Publishers, p. 1-187.
- Hadni, M., A. Lachkar and S. A. Ouatik (2014). "Multi-Word Term Extraction Based on New Hybrid Approach for Arabic Language." In *Second International Conference on Computational Science and Engineering*, Dubai, UAE, p. 109-120.
- Hamzé, H. (2005). "An Example of Linguistic Submission: The Translation of Affixes and Greco-Latin Formants into Arabic." In Branchadell, A. and L. M. West (dir.) *Less translated languages* (Vol. 58). Amsterdam: John Benjamins Publishing, p. 49-66.
- Harabagiu, S. and C. A. Bejan (2010). "A knowledge Extraction Framework for Biomedical Pathways." In *Proceedings of Summit on Translational Bioinformatics*. Repéré à <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041549/>
- Imprimante. (2016). Dans UNTERM. Repéré à <https://unterm.un.org/UNTERM/search?urlQuery=imprimante>
- Internet World Stats. (2015). INTERNET WORLD USERS BY LANGUAGE. Repéré à <http://www.internetworldstats.com/stats7.htm>
- Jarrar, M. (2011). "Building a Formal Arabic Ontology." In *Proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks*. Aleco, Arab League: Tunis, p. 26-28.
- Jarrar, M. (2006). "Towards the Notion of Gloss, and the Adoption of Linguistic Resources in Formal Ontology Engineering." In *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, Scotland, p. 497-503.

- Joan-Casademont, A. (2008). *Syntaxe et sémantique verbales dans le discours de spécialité : éléments pour l'activation de la valeur terminologique*, Thèse de Doctorat. Barcelona : IULA.
- Johansson, R. and P. Nugues (2006). "A FrameNet-based Semantic Role Labeler for Swedish." In *Proceedings of the COLING/ACL on Main Conference Poster Sessions*, Association for Computational Linguistics, p. 436-443.
- Khuwaileh, A. A. (2010). "IT Terminology and Translation: Cultural, Lexicographic and Linguistic Problems." In *LSP Journal-Language for Special Purposes, Professional Communication, Knowledge Management and Cognition* 1(2), p. 15-22.
- Kilgarriff, A., P. Rychly, P. Smrz and D. Tugwell (2004). "Itri-04-08 the Sketch Engine." In *Proceedings of Euralex*, Lorient, France, p. 105-116.
- Kipper, K., H. T. Dang and M. Palmer (2000a). "Class-Based Construction of a Verb Lexicon." In *AAAI/IAAI*, p. 691-696.
- Kipper, K., H. T. Dang, W. Schuler and M. Palmer (2000b). "Building a Class-Based Verb Lexicon Using TAGs." In *TAG+ 5 Fifth International Workshop on Tree Adjoining Grammars and Related Formalisms*, p. 147-154.
- Korhonen, A. and T. Briscoe. (2004). "Extended Lexical-Semantic Classification of English Verbs." In *Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics*. Association for Computational Linguistics, p. 38-45.
- Kocourek, R. (1991). *La langue française de la technique et de la science. Vers une linguistique de la langue savante*. Wiesbaden : Brandstetter.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago press.
- Lakoff, G. (1992). "How Metaphor Structures Dreams: The Theory of Conceptual Metaphor Applied to Dream Analysis." *Dreaming* 3(2), p. 1-25.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites (Vol. 1)*. Stanford: Stanford university press.
- Langacker, R. W. (1991). *Foundations of Cognitive Grammar, Vol. 2: Descriptive Application*. Stanford: Stanford University Press.
- Lelubre, X. (2008). « La constitution de la terminologie arabe de la physique : aspects diachroniques », dans Maniez, F. et P. Dury (dir.), *Lexicographie et terminologie :*



- Histoire des mots. En hommage à Henri Béjoint* (Travaux CRTT), Lyon : Presses Universitaires de Lyon, p. 1-8.
- Lelubre, X. (2009) « La métaphore dans la formation des termes arabes de la physique : aspects diachroniques », In Dury, P., F. Maniez, N. Arlin et C. Rougement (dir.), *La métaphore en langues de spécialité*, Travaux du CRTT, Presses Universitaires de Grenoble, p. 143-159
- Lemay, C., M.-C. L'Homme and P. Drouin (2005). "Two Methods for Extracting "Specific" Single-Word Terms from Specialized Corpora: Experimentation and Evaluation." *International Journal of Corpus Linguistics* 10(2), p. 227-255.
- León Araúz, P. and P. Faber (2012). "Causality in the Specialized Domain of the Environment." In Mititelu, V. B., O. Popescu and V. Pekar (dir.) *Proceedings of the Workshop "Semantic Relations-II. Enhancing Resources and Applications"* (LREC'12). Istanbul: ELRA, p. 10-17.
- Le Pesant D. et M. Mathieu-Colas (1998). « Introduction aux classes d'objets », *Langages* 32 (131), p. 6-33.
- Lerat, P. (1995). *Les langues spécialisées*. Paris : Presses Universitaires de France.
- Lerat, P. (2000). « Les emplois juridiques dans le lexique français », In *Lexique, Syntaxe et Sémantique, Mélanges offerts à Gaston GROSS, Bulag*, n° hors série, Centre Tesnière, Besançon, p. 41-47.
- Lerat, P. (2002a). « Qu'est-ce qu'un verbe spécialisé ? Le cas du droit », *Cahiers de Lexicologie* 80, p. 201-211.
- Lerat, P. (2002b). « Vocabulaire juridique et schémas d'arguments juridiques », *Meta* 47(2) : p. 155-162.
- Lerat, P. (2005). « Le vocabulaire juridique entre langue et texte », In Gémar, J-C. et N. Kasirer (dir.) *Jurilinguistique : entre langues et droits*, Montréal et Bruxelles : Bruylant, Les éditions thémis, p. 59-70.
- Lerat, P. (2006). « Terme et micro-contexte. Les prédications spécialisées », In *Mots, termes et contextes*, Blampain, D., Ph. Thoiron et M. Van Campenhoudt (dir), Paris : AUF, p. 89-98.
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press: Chicago.

- L'Homme, M.-C. (1995). « Définition d'une méthode de recensement et de codage des verbes en langue technique : applications en traduction », *TTR* 8(2) : p. 67-88.
- L'Homme, M.-C. (1998). « Le statut du verbe en langue de spécialité et sa description lexicographique », *Cahiers de lexicologie* 73(2), p. 61-84.
- L'Homme, M.-C. et C. Bodson (1998). « Modèle de description des verbes spécialisés (anglais français) combinant base de connaissances et hypertexte », and *Proceedings du Congrès international de terminologie*. Repéré à <http://www.uzei.com/Modulos/UsuariosFtp/Conexion/archivos202A.pdf>.
- L'Homme, M.-C. (2001). « Combinaisons lexicales spécialisées : Regroupement des mots clés par classes conceptuelles », In *Journées d'étude de l'ATALA. La collocation*. Rapport de recherche, Nantes : Institut de recherche en informatique de Nantes, p. 19-22.
- L'Homme, M.-C. (2002). "What Can Verbs and Adjectives Tell us about Terms?." In *Terminology and Knowledge Engineering, TKE 2002. Proceedings*. Nancy: INRIA, p. 65-70.
- L'Homme, M.-C. (2003). "Capturing the Lexical Structure in Special Subject Fields with Verbs and Verbal Derivatives: A Model for Specialized Lexicography." *International Journal of Lexicography* 16(4), p. 403-422.
- L'Homme, M.-C. (2004). *La terminologie : Principes et Techniques*. Montréal : Les presses de l'université de Montréal.
- L'Homme, M.-C. (2008). « Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés », *Traduire*, 217, p. 78-103.
- L'Homme, M.-C. (2012). « Le verbe terminologique : un portrait de travaux récents », dans Neveu, F. (dir), *Actes du 3e Congrès mondial de linguistique française*, Lyon, France : EDP. Sciences, p. 93-107.
- L'Homme, M.-C. and B. Robichaud (2014). "Frames and Terminology: Representing Predicative Terms in the Field of the Environment." In *The 25th International Conference on Computational Linguistics*, Dublin, Ireland, p. 186-197.
- L'Homme, M.-C., B. Robichaud and C. Subirats. (2014). "Discovering Frames in Specialized Domains." In *Language Resources and Evaluation. LREC 2014*, Reykjavik, Iceland, p. 1364-1371.

- L'Homme, M.-C. (2015). "Predicative Lexical Units in Terminology." In Gala, N., R. Rapp and G. Bel-Enguix (dir.), *Recent Advances in Language Production, Cognition and the Lexicon*. Switzerland: Springer, p.75-93.
- Lightfoot, D. W. (2002). "Introduction." In Chomsky, N. *Syntactic Structure*. Berlin/New York: Mouton de Gruyter, p. iv-xviii.
- Lorente, M. (2000). « Tipología verbal y textos especializados », In González Pereira, M. et M. Souto Gómez (dir.) *Cuestiones conceptuales y metodológicas de la lingüística*, Santiago de Compostela: Universidade de Santiago de Compostela, p. 143-153.
- Lorente, M. et C. Bevilacqua (2000). « Los verbos en las aplicaciones terminográficas », In *Actas del VII Simposio Iberoamericano de Terminología RITerm 2000*. Lisboa: ILTEC.
- Lorente, M. (2002). « Verbos y discurso especializado. » Estudios de Lingüística Española (ELiEs), (16). Repéré à <http://elies.rediris.es/elies16/Lorente.html>
- Lorente, M. (2007). « Les unitats lèxiques verbals dels textos especialitzats. Redefinició d'una proposta de classificació. » In M. Lorente, R. Estopà, J. Freixa, J. Martí and C. Tebé (dir.) *Estudis de lingüístics i de lingüística aplicada en honor de M. Teresa Cabré Catellví. Volum II: De deixebles*. Barcelona : Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra, p. 365-380.
- Lorente, M., J. Freixa et A. Casademont (2008). « Análisis de la sinonimia en algunos verbos jurídicos ». In *Actas del X Simposio Iberoamericano de Terminología "Terminología, conocimientos, sociedad y poder"*. Montevideo, Montevideo: RITERM.
- Maamouri, M., D. Graff, B. Bouziri, S. Krouna and S. Kulick (2010). "LDC Standard Arabic Morphological Analyzer (SAMA) v. 3.1." LDC Catalog No. LDC2010L01. ISBN: 1-58563-555-3.
- Maamouri, M., A. Bies, T. Buckwalter and W. Mekki (2004). "The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus." In *NEMLAR conference on Arabic language resources and tools*, Cairo, Egypt, p. 102-109.
- Marcus, M. P., B. Santorini and M. A. Marcinkiewicz (1993). "Building a Large Annotated Corpus of English: The Penn Treebank." *Computational Linguistics* 19(2), p. 313-330.
- Mathieu-Colas, M. (2002). « La représentation des verbes dans un dictionnaire électronique : de la langue générale aux langues spécialisées », *Cahiers de lexicologie*, (81), p. 51-67.

- Mathieu-Colas, M. (2006). « Les classes de verbes : syntaxe et sémantique », In Mejri, S. et J. Bacha (dir.) *Actes de conférence Le traitement du lexique. Catégorisation et Actualisation*, Sousse : Tunisie, p. 10-24.
- McCarthy, J.A. (1981). “Prosodic Theory of Non-Concatenative Morphology.” *Linguistic Inquiry* 12, p. 373–418.
- Mel’čuk, I., A. Clas et A. Polguère (1995). *Introduction à la lexicologie explicative et combinatoire*, Coll. Champs linguistiques/Universités francophones, Louvain-la-Neuve/Paris : Éditions Duculot/AUPELF-UREF.
- Meyer, I. and K. Mackintosh (2000). “When Terms Move into our Everyday Lives: An Overview of De-terminologization.” *Terminology* 6(1), p. 111–38.
- Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross and K. J. Miller (1990). “Introduction to WordNet: An On-Line Lexical Database\*.” *International journal of lexicography* 3(4), p. 235-244.
- Miller, G. A. (2003). “The Cognitive Revolution: A Historical Perspective.” *Trends in cognitive sciences* 7(3), p. 141-144.
- Minsky, M. (1974). « A Framework for Representing Knowledge ». Repéré à <http://hdl.handle.net/1721.1/6089>
- Mousser, J. (2010). “A Large Coverage Verb Taxonomy for Arabic.” In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*, p. 2675-2681.
- Mousser, J. (2011). “Classifying Arabic Verbs Using Sibling Classes.” In *Proceedings of the Ninth International Conference on Computational Semantics*. Association for Computational Linguistics, p. 355-359.
- Omar A. M., H. Mahjoob, E. Dossoki, A. Fodah, S. Ebraheem and S. Salem. (2000). *AL-meknaz Al-kabeer*. Riyadh, Saudi Arabia: Sotoor Company.
- Osborne, T. (2013). “A Look at Tesnière’s *Éléments* through the Lens of Modern Syntactic Theory.” In *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*. Charles University: Prague, p. 262-271.
- Owens, J. (2013). “A House of Sound Structure, of Marvellous Form and Proportion.” In Owens, J. (dir.) *The Oxford Handbook of Arabic Linguistics*. USA: Oxford University Press, p. 1-22.

- Palmer M, P. Kingsbury and D. Gildea (2005). “The Proposition Bank: An Annotated Corpus of Semantic Roles.” *Computational Linguistics* 31 (1), p. 71–106.
- Pearson, J. (1998). *Terms in context*. Amsterdam: John Benjamins.
- Petrucci, M. R. L. (1996). “Frame Semantics.” In Verschueren, J., J.-O. Östman, J. Blommaert and C. Bulcaen (dir.), *Handbook of Pragmatics*. Amsterdam/Philadelphia: John Benjamins, p. 1-8.
- Pimentel, J. (2011). « Description de verbes juridiques au moyen de la sémantique des cadres », In *Actes de Conférence TOTH*, Annecy : Institut Porphyre, Savoir et Connaissance, p. 145-166.
- Pimentel, J. (2012). *Criteria for the Validation of Specialized Verb Equivalents: Applications in Bilingual Terminography*, Thèse de doctorat. Université de Montréal.
- Polguère, A. (2003). « Étiquetage sémantique des lexies dans la base de données DiCo », *TAL. Traitement automatique des langues* 44(2), p. 39-68.
- Pradet, Q., L. Danlos and G. De Chalendar. (2014). “Adapting VerbNet to French Using Existing Resources.” In *LREC'14-Ninth International Conference on Language Resources and Evaluation*. Reykjavik, Iceland, p. 1122-1126.
- Propbank (2015). Index of /propbank/framesets-english. Repéré à <https://verbs.colorado.edu/propbank/framesets-english/>
- Pustejovsky, J. (1991). “The Generative Lexicon.” *Computational linguistics* 17(4), p. 409-441.
- Rey, A. (1979). *La terminologie : noms et notions*. Coll. « Que sais-je ? », Paris : Presses universitaires de France.
- Roberts, A., L. Al-sulaiti and E. Atwell (2005). “aConCorde: Towards a Proper Concordance for Arabic.” In *Proceedings of the Corpus Linguistics Conference*. United Kingdom: University of Birmingham 1(1), p. 1-15.
- Rodríguez, H., D. Farwell, J. Farreres, M. Bertran, M. Alkhalifa, M. A. Martí, W. Black, S. Elkateb, J. Kirk, A. Pease, P. Vossen and C. Fellbaum. (2008). “Arabic WordNet: Current State and Future Extensions.” In *Proceedings of The Fourth Global WordNet Conference*. Szeged, Hungary, p. 387-405.

- Ruppenhofer, J., M. Ellsworth, M. R. L. Petruck, C. R. Johnson and J. Scheffczyk (2010). *FrameNet II: Extended theory and practice*. Repéré à <https://framenet.icsi.berkeley.edu/fndrupal/fnbibliography/author/169>
- Sadat, F. and E. Mohamed (2013). "Pre-processing and Language Analysis for Arabic to French Statistical Machine Translation." In *Proceedings of TALN*. Les Sables d'Olonne, France: ATALA, p. 620-627.
- Sager, J. (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.
- Sánchez Cárdenas, B. and M. Buendia Castro. (2012). "Inclusion of Verbal Syntagmatic Patterns in Specialized Dictionaries: The Case of EcoLexicon." In Vatvedt Fjeld, R. and J. Matilde Torjusen (dir.), *Proceedings of the 15th EURALEX International Congress*. Oslo: EURALEX, p. 554–562.
- Saraireh, M. A. (2001). "Inconsistency in Technical Terminology: A Problem for Standardization in Arabic." *Babel* 47(1), p. 10-21.
- Schmidt, T. (2009). "The Kicktionary – A Multilingual Lexical Resources of Football Language." In Boas, H.C. (dir.), *Multilingual FrameNets in Computational Lexicography. Methods and Applications*. Berlin/NewYork: Mouton de Gruyter: p. 101-134.
- Sieny, M. E. (1985). "Scientific Terminology in the Arab World: Production, Co-ordination, and Dissemination." *Meta* 30(2), p. 155-160.
- Solimando, C. (2015). "Word-formation in the Arabic Language of Tourism." In *Langues Cultures Meditation* 2(1), p. 157-168.
- Soudi, A., G. Neumann and A. V. D. Bosch (2007). "Arabic Computational Morphology: Knowledge-based and Empirical Methods." In Soudi, A., G. Neumann and A. V. D. Bosch (dir.), *Knowledge-based and Empirical Methods: Arabic Computational Morphology*. Nederland: Springer, p. 3-14.
- Sproat, R. (2007). "Preface." In Soudi, A., G. Neumann and A. V. D. Bosch (dir.), *Knowledge-based and Empirical Methods: Arabic Computational Morphology*. Nederland: Springer, p. vii-viii.
- Stetkevych, J. (1970). *The Modern Arabic Literary Language: Lexical and Stylistic Developments (No. 6)*. Washington D.C.: Georgetown University Press.

- Talmy, L. (1985). "Lexicalization Patterns: Semantic Structure in Lexical Forms." *Language Typology and Syntactic Description* 3, p. 57-149.
- Talmy, L. (2000). *Toward a Cognitive Semantics (Vol. 1)*. Cambridge: MIT press.
- Talmy, L. (2011). "Cognitive Semantics: An overview." In Maienborn, C., K. Von Heusinger et P. Portner (dir.), *Semantics: An International Handbook of Natural Language Meaning*. Berlin: De Gruyter Mouton, p. 622-642.
- taSaf~aHa (تصفح) (navigateur). Dans Almaany. Repérée à <http://www.almaany.com/ar/dict/ar-fr/%D8%AA%D8%B5%D9%81%D8%AD/>
- Temmerman, R. (2000). *Towards New Ways of Terminology Description*. Amsterdam/Philadelphia: John Benjamins.
- Temmerman, R. (2001). "Sociocognitive Terminology Theory." In Cabré, M. T. and J. Feliu (dir.), *Terminologia y Cognición*. Barcelona: Universitat Pompeu Fabra, p. 75-92.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Testen, D. (2016). "Root and Pattern System." In *Encyclopædia Britannica. Encyclopædia Britannica Online*. Repéré à <http://www.britannica.com/topic/root-and-pattern-system>.
- Teubert, W. (2009). « La linguistique de corpus : une alternative », *Semen. Revue de sémiolinguistique des textes et discours* (27), p. 185-211.
- Toutanova, K., D. Klein, C. D. Manning and Y. Singer (2003). "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network." In *Proceedings of HLT-NAACL*. Edmonton, Canada: Association for Computational Linguistics, p. 173-180.
- Valente, R. S. (2002). « *La lexicologie explicative et combinatoire* » dans *le traitement des unités lexicales spécialisées*, Thèse de Doctorat. Université de Montréal.
- Van Mol, M. (2000). "The Development of a New Learner's Dictionary for Modern Standard Arabic: The Linguistic Corpus Approach." In Heid, U., S. Evert, E. Lehmann and C. Rohrer (dir.), *Proceedings of the ninth Euralex international congress, Euralex 2000*. Stuttgart, Germany: Institut für Maschinelle Sprachverarbeitung, p. 831-836.
- Van Valin Jr, R. D. (2005). *Exploring the Syntax-Semantics Interface*. Cambridge University Press.
- Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca, New York: Cornell University Press.
- VerbNet (2015). Verb-Index. Repéré à <http://verbs.colorado.edu/verb-index/>

- Vossen, P. (dir.) (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer Academic Publishers.
- Wandji, Tchami, O., M.-C. L’Homme and N. Grabar (2013). “Discovering Semantic Frames for a Contrastive Study of Verbs in Medical Corpora.” In *Proceedings of 10th International Conference on Terminology and Artificial Intelligence*. Villetaneuse: Université Paris 13, p. 155.
- Wandji Tchami, O. (2014). « Les modèles de description du verbe dans les travaux de Linguistique, Terminologie et TAL », dans Bigi, B. (dir.) *Acte de 21ème conférence Traitement Automatique des Langues Naturelles (TALN)*, Marseille : Laboratoire Parole et Langage p. 37-38.
- WordNet (2015). WordNet Home Page. Repéré à <http://wordnet.princeton.edu>
- World International Property Organization. (2015). WIPO Pearl - User Guide. Repéré à <http://www.wipo.int/reference/en/wipopearl>.
- Wulff, H. R. (2004). “The Language of Medicine.” *Journal of the Royal Society of Medicine* 97(4), p. 187-188.
- yuEAlj* (يعالج) (traiter). (2015). Unified Medical Dictionary. Repéré à <http://www.emro.who.int/Unified-Medical-Dictionary.html>.



## Annexe 1 : Sources du corpus

SOURCE	DESCRIPTION	TYPE	ADRESSE URL	NOMBRE DE MOTS
<b>ABAHE.UK</b>	Manuel sur le dépannage informatique du système d'exploitation Windows	Manuel	<a href="http://www.abahe.co.uk">http://www.abahe.co.uk</a>	63353
<b>AITNEWS.COM</b>	Articles sur différents sujets portant sur la technologie informatique et cellulaire	Articles	<a href="http://www.aitnews.com">http://www.aitnews.com</a>	48047
<b>AL-JAZEERA.NET</b>	Articles sur différents sujets portant sur la technologie informatique	Articles	<a href="http://www.aljazeera.net/news">http://www.aljazeera.net/news</a>	17126
<b>A-PLUS</b>	Chapitre d'un manuel sur le dépannage informatique	Manuel	<a href="http://raypub.com/catalog/book_info.php?cPath=1_17&amp;products_id=1053&amp;PHPSESSID=0bd75cc75e3bbae22c161fd4c93562d">http://raypub.com/catalog/book_info.php?cPath=1_17&amp;products_id=1053&amp;PHPSESSID=0bd75cc75e3bbae22c161fd4c93562d</a>	15101
<b>APPLE-WD.COM</b>	Articles sur les différents produits de la compagnie Apple. C'est un site web arabe dédié aux produits de la compagnie.	Articles	<a href="http://www.apple-wd.com">http://www.apple-wd.com</a>	1353
<b>ARABHARDWARE</b>	Site Internet dédié aux composants informatiques matériels	Articles	<a href="http://www.arabhardware.net">http://www.arabhardware.net</a>	33312
<b>ASKMOST.COM</b>	Articles sur différents sujets portant sur le dépannage informatique et l'utilisation des applications des téléphones intelligents et tablettes	Articles	<a href="http://www.askmost.com/">http://www.askmost.com/</a>	7808
<b>COMPUTER-WD.COM</b>	Site web sur les différents sujets touchant aux solutions informatiques de la compagnie Microsoft	Articles	<a href="http://www.computer-wd.com">http://www.computer-wd.com</a>	4228
<b>ITP.COM</b>	Un réseau d'informations sur Internet sur le dernier cri technologique au Moyen-Orient et dans le monde.	Articles	<a href="http://www.itp.net/arabic">http://www.itp.net/arabic</a>	22284

<b>KEEF-TECH.COM</b>	Site Internet sur le dépannage informatique et cellulaire ainsi que sur leur mode de fonctionnement et utilisation	Articles	<a href="http://keeftech.com/">http://keeftech.com/</a>	10365
<b>KETTANEH.NET</b>	Des cours en ligne destinés aux personnes qui veulent s'initier en informatique	Articles	<a href="http://www.kettaneh.net">http://www.kettaneh.net</a>	19530
<b>LINUX</b>	Un manuel sur le système d'exploitation Linux	Manuel	<a href="http://librebooks.org">http://librebooks.org</a>	190460
<b>OS</b>	Un manuel sur les différents systèmes d'exploitation	Manuel	<a href="http://www.kutub.info/librariy/book/12863">http://www.kutub.info/librariy/book/12863</a>	4280
<b>PYTHON</b>	Un manuel de programmation sur le langage Python	Manuel	<a href="http://librebooks.org/learn-programming-with-python3/">http://librebooks.org/learn-programming-with-python3/</a>	194245
<b>TAIBAHU.EDU</b>	Des cours en ligne sur différents aspects de système d'exploitation Windows, DOS et Linux ainsi que les périphériques.	Articles	<a href="http://www.taibahu.edu.sa/pages.aspx?pid=8132">http://www.taibahu.edu.sa/pages.aspx?pid=8132</a>	51421
<b>TECH-WD.COM</b>	Un site web sur tout ce qui concerne la technologie informatique et cellulaire	Articles	<a href="http://www.tech-wd.com/wd/">http://www.tech-wd.com/wd/</a>	11645
<b>TIQANIAT.COM</b>	Un site qui regroupe une variété d'articles sur tout ce que l'utilisateur peut faire avec son PC, y compris des astuces et les fonctionnalités cachées du système Windows	Articles	<a href="http://www.tiqaniat.com">http://www.tiqaniat.com</a>	7276
<b>UBUNTU</b>	Un manuel sur le système d'exploitation Ubuntu	Manuel	<a href="http://itwadi.com/node/1478">http://itwadi.com/node/1478</a>	3416
<b>UNLIMITED-TECH</b>	Un blog sur les différents produits informatiques	Articles	<a href="http://www.unlimit-tech.com/blog/">http://www.unlimit-tech.com/blog/</a>	1639

## Annexe 2 : Liste partielle des UTN candidates

fréq-réf	fréq-spé	spécificité	UTN	Équivalent
2984	3557	233,561221	نظام	systeme
3411	2053	143,450261	تشغيل	exécution
818	1100	133,836875	شكل	forme
4013	2029	133,3585442	استخدام	utilisation
9521	2957	129,4173799	ملف	fichier
4172	1766	115,5246684	مستخدم	utilisateur
454	687	108,7008289	خدمة	service
8718	2368	108,1107183	جهاز	appareil
504	664	103,5044998	نسخة	copie
492	626	99,65451437	أداة	outil
970	792	98,94524443	حالة	état
653	683	98,89228288	عدد	nombre
1760	829	82,84201114	صورة	image
1083	577	72,64615481	وصول	accès
962	536	71,24628107	رسالة	message
455	381	69,22276734	نص	texte
222	289	68,15670903	وظيفة	fonction
540	387	66,31763051	حصول	obtention
4352	1021	65,53136936	برنامج	programme
3201	865	65,17835416	شركة	compagnie
1612	593	62,89358324	مفتاح	clef
1159	492	61,09443695	تنفيذ	exécution
8449	1363	60,00551457	شبكة	réseau
5284	1031	59,05334765	مثل	comme
8275	1313	58,22986625	تطبيق	application
3700	830	57,59666562	إصدار	version
1	106	56,18211445	mAwis_1	souris
1259	465	55,81055397	واجهة	interface
1325	451	52,89643442	سلسلة	série
2759	653	52,68316615	كمبيوتر	ordinateur
994	376	50,75364449	حاسوب	ordinateur
2687	618	50,42831553	قائمة	menu
6698	1004	48,93520287	اسم	nom

### Annexe 3 : Liste partielle des UTV candidates

freq_ref	freq_spec	specificite	lemme	Équivalent
5259	1480	87,07648698	عمل	tourner
1839	378	36,94526985	دعم	supporter
4180	576	34,87287876	عرض	afficher
10601	817	24,15703244	ظهر	s'afficher
5465	435	18,32166436	naqar-u_1	cliquer
1111	154	18,16751535	tadaE~am_1	supporter
7753	523	16,28702551	ضغط	appuyer
4001	303	14,42858367	كتب	écrire
1878	172	13,43509403	حدد	sélectionner
7439	455	13,12562574	اختار	choisir
684	72	9,979538667	أعاد	refaire
1303	107	9,469268338	حمل	télécharger
5707	299	8,02597789	ضبط	régler
5031	255	6,901815973	nasax-a_1	copier
7141	326	5,95848754	نقل	transférer
733	49	5,012392073	غلق	fermer
7467	290	2,957807836	حذف	supprimer
2678	95	0,874724146	انتقل	se transférer
9325	318	0,850994174	حفظ	sauvegarder
689	24	0,431878482	استجاب	répondre
3083	87	-1,234033779	أنشأ	créer
926	22	-1,345985366	اختفى	disparaître
525	8	-2,022279485	qal~aE_1	démarrer
914	17	-2,182783325	حمى	protéger
5022	126	-2,797886749	دخل	accéder
2066	29	-4,490885695	استقبل	recevoir
1912	14	-5,926486621	اشتغل	fonctionner
6080	104	-6,467912579	بحث	rechercher
6850	122	-6,551626254	taSaf~aH_1	naviguer
7860	141	-6,960692036	فتح	ouvrir
2164	10	-7,005661993	Har~ar_1	éditer
10412	200	-7,315287978	حدث	mettre à jour
4072	44	-7,477307029	اختبر	tester

## Annexe 4 : UTV retenues

UTV	Équivalent FR dans DiCoInfo	Équivalent EN dans DiCoInfo
1أدار	gérer <sub>1</sub>	X
1أدخل	entrer <sub>1</sub>	enter <sub>1</sub>
1أرسل	envoyer <sub>1</sub>	send <sub>1</sub>
1أغلق	fermer <sub>1</sub>	close <sub>1</sub>
1أقنع	démarrer <sub>1a</sub>	start <sub>1a</sub>
1أنشأ	créer <sub>1</sub>	create <sub>1</sub>
2أوقف	arrêter (pas dans DiCoInfo)	halt (pas dans DiCoInfo)
1إتصل	connecter <sub>1</sub>	connect <sub>1</sub>
1اختار	sélectionner <sub>1</sub>	X
1استجاب	répondre (pas dans DiCoInfo)	reply (pas dans DiCoInfo)
1استقبل	recevoir <sub>1</sub>	receive <sub>1</sub>
1بحث	rechercher <sub>1</sub>	X
1تحكم	contrôler (pas dans DiCoInfo)	control (pas dans DiCoInfo)
1تصفح	naviguer <sub>1</sub>	browse <sub>1</sub>
1تعرف	reconnaître (pas dans DiCoInfo)	recognize (pas dans DiCoInfo)
1تفحص	analyser <sub>1a</sub>	scan (pas dans DiCoInfo)
1توافق	être compatible (compatible <sub>1</sub> )	to be compatible (compatible <sub>1</sub> )
1توقف	s'arrêter (pas dans DiCoInfo)	stop (pas dans DiCoInfo)
1ثبت	installer <sub>2</sub>	install <sub>2</sub>
1حدث	mettre à jour <sub>1</sub>	update <sub>1</sub>
1حذف	supprimer <sub>1</sub>	delete <sub>1</sub>
1حسن	optimiser <sub>1a</sub>	X
1حفظ	sauvegarder <sub>1</sub>	backup <sub>1</sub>
1حمل	télécharger <sub>1</sub>	download <sub>1</sub>
1خزن	stocker <sub>1b</sub>	store <sub>1b</sub>
1دخل	accéder <sub>2</sub>	access <sub>2</sub>
1دعم	accepter <sub>2</sub>	support <sub>2</sub>

سجل 1	enregistrer <sub>1</sub>	save <sub>1</sub>
شغل 1	actionner (pas dans DiCoInfo)	run <sub>2b</sub>
ضبط 1	paramétrer <sub>1</sub>	configure <sub>1</sub>
ضغط 1	appuyer <sub>1</sub>	press <sub>1</sub>
طبع 1a	imprimer <sub>1a</sub>	print <sub>1a</sub>
طور 1	développer <sub>1</sub>	develop <sub>1</sub>
ظهر 1a	afficher <sub>1a</sub>	display <sub>1a</sub>
عالج 1	traiter <sub>1</sub>	process <sub>1</sub>
عدل 1	modifier <sub>1</sub>	X
عرض 1b	afficher <sub>1b</sub>	display <sub>1b</sub>
عمل 1	tourner <sub>2</sub>	run <sub>1a</sub>
عمل 2	tourner <sub>3</sub>	run <sub>2a</sub>
فتح 1	ouvrir <sub>1</sub>	open <sub>1</sub>
فحص 1	inspecter (pas dans DiCoInfo)	check (pas dans DiCoInfo)
قرأ 1	lire <sub>2</sub>	read <sub>2</sub>
كتب 1	écrire <sub>1</sub>	write <sub>1</sub>
نسخ 1	copier <sub>1</sub>	copy <sub>1</sub>
نفذ 1	exécuter <sub>2b</sub>	execute <sub>2b</sub>
نقر 1	cliquer <sub>1</sub>	click <sub>1</sub>
نقل 1	transférer <sub>1</sub>	transfer <sub>1</sub>
47		

## Annexe 5 : UTN retenues

UTN	Équivalent FR dans DiCoInfo	Équivalent EN dans DiCoInfo
.1.1 إتصال	connexion <sub>1.1</sub>	connection <sub>1.1</sub>
.1 إدارة	gestion <sub>1</sub>	X
.1 إدخال	entrée <sub>1</sub>	entry <sub>1</sub>
.1a إرسال	envoi <sub>1a</sub>	X
.1b إرسال	envoi <sub>1b</sub>	X
.1 إعادة-تثبيت	réinstallation <sub>1</sub>	reinstall <sub>1</sub>
.1 إعادة-تسمية	renommage <sub>1</sub>	X
.1a إعادة-تشغيل	redémarrage <sub>1a</sub>	restart <sub>1a</sub>
.1b إعادة-تشغيل	redémarrage <sub>1a</sub>	restart <sub>1b</sub>
.1 إغلاق	fermeture <sub>1</sub>	X
.1a إقلاع	démarrage <sub>1a</sub>	restart <sub>1a</sub>
.1b إقلاع	redémarrage <sub>1b</sub>	restart <sub>1b</sub>
.1 إنشاء	création <sub>1</sub>	creation <sub>1</sub>
.1 إنهاء	terminer <sub>1</sub>	X
.1 اختيار	sélection <sub>1</sub>	X
.1 استرجاع	restauration <sub>1</sub>	recovery <sub>1</sub>
.1 استرداد	récupération <sub>1</sub>	recovery <sub>1</sub>
.1 استعادة	restauration <sub>1</sub>	recovery <sub>1</sub>
.1 استقبال	réception <sub>1</sub>	reception <sub>1</sub>
.2 استقبال	réception <sub>2</sub>	reception <sub>2</sub>
.1.1 بحث	recherche <sub>1</sub>	X
.1 برمجة	programmation <sub>1</sub>	programming <sub>1</sub>
.2 برمجة	programmation <sub>2</sub>	programming <sub>2</sub>
.1 تثبيت	installation <sub>1</sub>	installation <sub>1</sub>
.2 تثبيت	installation <sub>2</sub>	installation <sub>2</sub>
.1 تحديث	mise à jour <sub>1,2</sub>	update <sub>2</sub>

تحرير 1.	édition <sub>1</sub>	editing <sub>1</sub>
تحسين 1.	optimisation <sub>1.1</sub>	optimization <sub>1.1</sub>
تحكم 1.1.	contrôler (pas dans DiCoInfo)	control (pas dans DiCoInfo)
تحميل 1.	téléchargement <sub>1</sub>	download <sub>1</sub>
تحويل 1.	transfert <sub>1</sub>	transfer <sub>1</sub>
تسجيل 1.	enregistrement <sub>1</sub>	X
تسريع 1.	accélération (pas dans DiCoInfo)	speed up (pas dans DiCoInfo)
تشغيل 1a.	exécution <sub>1a</sub>	execution <sub>1a</sub>
تشغيل 1b.	exécution <sub>1b</sub>	exécution <sub>1b</sub>
تشغيل 2a.	exécution <sub>2a</sub>	exécution <sub>2a</sub>
تشغيل 2b.	exécution <sub>2b</sub>	exécution <sub>2b</sub>
تصفح 1.1.	navigation <sub>1</sub>	browsing <sub>1</sub>
تطوير 1.	développement <sub>1</sub>	development <sub>1</sub>
تعديل 1.	modification <sub>1</sub>	X
تفعيل 1.	activation <sub>1</sub>	X
تنفيذ 1.	exécution <sub>2b</sub>	execution <sub>2b</sub>
توصيل 2.1.	connexion <sub>2.1</sub>	X
حماية 1.	protection <sub>1a</sub>	X
دعم 1.1.	support <sub>1</sub>	support <sub>1</sub>
ضغط 1.1.	appui <sub>1</sub>	X
ضغط 2.	compression <sub>1</sub>	compression <sub>1</sub>
طباعة 1a.	impression <sub>1a.1</sub>	printing <sub>1a</sub>
طباعة 1b.	impression <sub>1b</sub>	printing <sub>1b</sub>
عرض 1b.1.	affichage <sub>1b.1</sub>	display <sub>1b.1</sub>
قراءة 1.	lecture <sub>2</sub>	X
كتابة 1.	écriture <sub>1</sub>	X
كتابة 2.	écriture <sub>2</sub>	X
لصق 1.	collage <sub>1</sub>	X
مشاركة 1.	partage <sub>1.1</sub>	sharing <sub>1</sub>



1. معالجة	traitement <sub>1</sub>	processing <sub>1</sub>
1.1. نسخ	copie <sub>1.1</sub>	copy <sub>3.2</sub>
1.1. وصول	accès <sub>1</sub>	access <sub>1.1</sub>
1.2. وصول	accès <sub>2a</sub>	access <sub>2.1</sub>
<b>59</b>		

## Annexe 6 : Rôles sémantiques

Rôle Sémantique	Définition	Exemples
<b>Agent</b>	La notion d'Agent désigne le participant qui se trouve à l'origine de l'action.	<i>An\$A mustaxdim</i> [Agent] <i>HisAbN</i> [Patient] أنشأ مستخدم حسابا <i>L'utilisateur</i> [Agent] <i>a créé un compte</i> [Patient]
<b>Agent-a et Agent-b</b>	Cette structure désigne la réalisation linguistique où deux Agents (personne à l'origine de l'action) ont le même statut dans la phrase.	<i>hA*ihi hya Hal Samba, Al~aty tasmAh bilmu\$Arakap bayna wyndwz</i> [Agent-a] <i>wa lynks</i> [Agent-b] <i>Eabr Al\$abakap</i> هذه هي حال سامبا التي تسمح بالمشاركة بين ويندوز و لينكس على الشبكة C'est le cas de Samba, qui permet le <b>partage</b> entre Windows [Agent-a] et Linux [Agent-b] sur le réseau
<b>Destination</b>	L'actant Destination désigne l'objectif de l'action effectuée par l'Agent. De plus, il renvoie à la finalité typique d'une action.	yrafaE Almustxdim [Agent] AlSwrap [Patient] Eiy Flickr [Destination] على يرفع المستخدم الصورة فلكر L'utilisateur [Agent] téléverse la photo [Patient] sur Flickr [Destination]
<b>Environnement1 et Environnement2</b>	Ces deux actants désignent la plateforme informatique (logicielle ou matérielle) où une action s'effectue.	Almustaxdimwn [Agent] yastaTyEwn Almu\$Araka fy <b>ta\$gyl</b> AltaTbyqAt wa AlbarAmij Almuxtalifap [Patient] Eiy Ajhizap muxtalifap [Environnement1] المستخدمون يستطيعوا المشاركة في تشغيل التطبيقات و البرامج المختلفة على أجهزة مختلفة Les utilisateurs [Agent] peuvent participer à l'exécution des applications et des différents programmes [Patient] sur différents appareils [Environnement1].

<p><b>Instrument</b></p>	<p>L'actant Instrument renvoie à l'entité qui désigne le moyen par lequel l'action est effectuée.</p>	<p>&gt;unqur bizir AlmAws [Instrument]  Ely Almalaf [Patient]  أفقر بزر الماوس على الملف  Cliquez avec le bouton de la souris [Instrument] sur le dossier [Patient]</p>
<p><b>Lieu</b></p>	<p>Cet actant désigne l'endroit où l'activité se déroule</p>	<p>Aw~alaN, AiftaH Almaktabap  Almawjwdap Ely mawqiEak.  vum~, AibHav Ean  AlmustanadAt [Patient]  Almawjwdap Ely  Alkumbywtar [Lieu] wa Al~ty  turyd taHmylahA  أولاً، افتح المكتبة الموجودة على موقعك. ثم إبحث عن المستندات الموجودة على الكمبيوتر والتي تريد تحميلها  D'abord, ouvre la bibliothèque qui se trouve sur ton site, puis recherche les documents [Patient] qui se trouvent sur ton ordinateur [Lieu] et que tu veux télécharger</p>
<p><b>Matériau</b></p>	<p>Dans une plateforme informatique, cet actant désigne le langage utilisé lors de la conception d'un logiciel.</p>	<p>Ain kitAbap barnAmij [Patient]  bilugap rafiEap AlmstawY  [Matériau] Ashal kavyraN wa  yastagriq waqtAN Aqal  likitAbatih  ان كتابة برنامج بلغة رفيعة المستوى اسهل كثيرا و يستغرق وقتا اقل لكتابته  Écrire un programme [Patient] dans un langage sophistiqué [Matériau] est plus facile et prend moins du temps</p>
<p><b>Menace</b></p>	<p>Cet actant désigne le risque auquel un composant informatique peut être exposé.</p>	<p>lihimAyap AlniZam [Destination]  min ha*ihi AlAxTar [Menace]  yatim AEtimAd Aly~At  minhA: judrAn AlnAr  Firewalls [Instrument]  لحماية النظام من هذه الاخطار يتم اعتماد اليات منها : جدران النار  Pour protéger le système [Destination] de ces risques [Menace], certaines mesures sont prises en compte, comme Firewalls [Instrument]</p>

<p><b>Patient</b></p>	<p>Cet actant désigne l'entité subissant l'action. Patient peut renvoyer soit à l'élément créé ou utilisé par l'Agent.</p>	<p><i>wiHdap AlmuEAlajap</i>  <i>Almarkazy~ap</i> [Agent] <i>tunaf~i*</i>  <i>Al&gt;wAmir</i> [Patient]          وحدة المعالجة المركزية تنفذ الأوامر  <i>L'unité centrale</i> [Agent] <i>exécute</i>  <i>les commandes</i> [Patient]</p>
<p><b>Patient-a et Patient-b</b></p>	<p>Ces deux actants désignent une réalisation dans une phrase où les deux parties concernées subissent une action exprimée par des composants informatiques. Linguistiquement, ils sont toujours accompagné de maE (مع) (avec).</p>	<p><i>nusxap wyndwz jadydap</i> [Patient-a]  <b>tatawAfaq</b> maE Almasinjar  <i>Aljadyd</i> [Patient-b]          نسخة ويندوز جديدة تتوافق          مع الماسنجر الجديد          La nouvelle version de          Windows [Patient-a] est compatible          avec le nouveau          Messenger [Patient-b].</p>
<p><b>Résultat</b></p>	<p>Cet actant désigne le format final qu'un composant peut avoir après un processus de conversion.</p>	<p><i>liHuSwl EIY natA}ij jay~idap,</i>  <i>yustaHsan taHwyl malaf~At</i>  <i>AlmusyqY</i> [Patient] <i>iIY Suwar</i>          [Résultat]          للحصول على نتائج جيدة          يستحسن تحويل ملفات الموسيقى إلى          صور          Pour obtenir de bons résultats,          il est mieux de <b>convertir</b> les          fichiers musicaux [Patient] en          images [Résultat]</p>
<p><b>Source</b></p>	<p>Il s'agit de l'endroit à partir duquel l'action est réalisée.</p>	<p><i>lilrAgibyn</i> [Agent] <b>bitaHmyl</b> Ay  <i>min ha*ihi AltawziEAt</i> [Patient]  <i>maj~anaN min AlAintirnit</i> [Source]  <i>murajaEap AlrAbiT AltAly</i>          للراغبين بتحميل اي من هذه          التوزيعات مجاناً من الانترنت مراجعة          الرابط التالي          Pour ceux [Agent] qui désirent  <b>télécharger</b> une de ces          distributions [Patient] gratuitement          de l'Internet [Source], ils peuvent          consulter le lien suivant.</p>

## Annexe 7 : Correspondance entre les UTV et les UTN

UTN	Équivalent FR dans DiCoInfo	UTV	Équivalent FR dans DiCoInfo
.1.1 إتصال	connexion <sub>1,1</sub>	X	
.1 إدارة	gestion <sub>1</sub>	أدار <sub>1</sub>	gérer <sub>1</sub>
.1 إدخال	entrée <sub>1</sub>	أدخل <sub>1</sub>	entrer <sub>1</sub>
.1a إرسال	envoi <sub>1a</sub>	أرسل <sub>1</sub>	envoyer <sub>1</sub>
.1b إرسال	envoi <sub>1b</sub>	X	
.1 إعادة-تثبيت	réinstallation <sub>1</sub>	X	
.1 إعادة-تسمية	renommage <sub>1</sub>	X	
.1a إعادة-تشغيل	redémarrage <sub>1a</sub>	X	restart <sub>1a</sub>
.1b إعادة-تشغيل	redémarrage <sub>1a</sub>	X	restart <sub>1b</sub>
.1 إغلاق	fermeture <sub>1</sub>	أغلق <sub>1</sub>	fermer <sub>1</sub>
.1a إقلاع	démarrage <sub>1a</sub>	أقلع <sub>1a</sub>	démarrer <sub>1a</sub>
.1b إقلاع	redémarrage <sub>1b</sub>	X	
.1 إنشاء	création <sub>1</sub>	أنشأ <sub>1</sub>	créer <sub>1</sub>
.1 إنهاء	terminer <sub>1</sub>	X	
.1 اختيار	sélection <sub>1</sub>	اختار <sub>1</sub>	sélectionner <sub>1</sub>
.1 استرجاع	restauration <sub>1</sub>	X	
.1 استرداد	récupération <sub>1</sub>	X	
.1 استعادة	restauration <sub>1</sub>	X	
.1 استقبال	réception <sub>1</sub>	استقبل <sub>1</sub>	
.2 استقبال	réception <sub>2</sub>	X	
.1.1 بحث	recherche <sub>1</sub>	بحث <sub>1</sub>	rechercher <sub>1</sub>
.1 برمجة	programmation <sub>1</sub>	X	
.2 برمجة	programmation <sub>2</sub>	X	
.1 تثبيت	installation <sub>1</sub>	ثبت <sub>1</sub>	installer <sub>2</sub>

تثبيت <sub>2</sub>	installation <sub>2</sub>	X	
تحديث <sub>1</sub>	mise à jour <sub>1,2</sub>	حدث <sub>1</sub>	mettre à jour <sub>1</sub>
تحرير <sub>1</sub>	édition <sub>1</sub>	X	
تحسين <sub>1</sub>	optimisation <sub>1,1</sub>	حسن <sub>1</sub>	optimiser <sub>1a</sub>
تحكم <sub>1.1</sub>	contrôler (pas dans DiCoInfo)	تحكم <sub>1</sub>	contrôler (pas dans DiCoInfo)
تحميل <sub>1</sub>	téléchargement <sub>1</sub>	حمل <sub>1</sub>	télécharger <sub>1</sub>
تحويل <sub>1</sub>	transfert <sub>1</sub>	X	
تسجيل <sub>1</sub>	enregistrement <sub>1</sub>	سجل <sub>1</sub>	enregistrer <sub>1</sub>
تسريع <sub>1</sub>	accélération (pas dans DiCoInfo)	X	
تشغيل <sub>1a</sub>	exécution <sub>1a</sub>	X	
تشغيل <sub>1b</sub>	exécution <sub>1b</sub>	X	
تشغيل <sub>2a</sub>	exécution <sub>2a</sub>	X	
تشغيل <sub>2b</sub>	exécution <sub>2b</sub>	شغل <sub>1</sub>	actionner (pas dans DiCoInfo)
تصفح <sub>1.1</sub>	navigation <sub>1</sub>	تصفح <sub>1</sub>	naviguer <sub>1</sub>
تطوير <sub>1</sub>	développement <sub>1</sub>	طور <sub>1</sub>	développer <sub>1</sub>
تعديل <sub>1</sub>	modification <sub>1</sub>	عدل <sub>1</sub>	modifier <sub>1</sub>
تفعيل <sub>1</sub>	activation <sub>1</sub>	X	
تنفيذ <sub>1</sub>	exécution <sub>2b</sub>	نفذ <sub>1</sub>	exécuter <sub>2b</sub>
توصيل <sub>2.1</sub>	connexion <sub>2,1</sub>	X	
حماية <sub>1</sub>	protection <sub>1a</sub>	X	
دعم <sub>1.1</sub>	support <sub>1</sub>	دعم <sub>1</sub>	accepter <sub>2</sub>
ضغط <sub>1.1</sub>	appui <sub>1</sub>	ضغط <sub>1</sub>	appuyer <sub>1</sub>
ضغط <sub>2</sub>	compression <sub>1</sub>	X	
طباعة <sub>1a</sub>	impression <sub>1a,1</sub>	اطبع <sub>1a</sub>	imprimer <sub>1a</sub>
طباعة <sub>1b</sub>	impression <sub>1b</sub>	X	
عرض <sub>1b.1</sub>	affichage <sub>1b,1</sub>	عرض <sub>1b</sub>	afficher <sub>1b</sub>
قراءة <sub>1</sub>	lecture <sub>2</sub>	قرأ <sub>1</sub>	lire <sub>2</sub>

1.كتابة	écriture <sub>1</sub>	كتب1	écrire <sub>1</sub>
2.كتابة	écriture <sub>2</sub>	X	
1.لصق	collage <sub>1</sub>	X	
1.مشاركة	partage <sub>1.1</sub>	X	
1.معالجة	traitement <sub>1</sub>	عالج1	traiter <sub>1</sub>
1.1.نسخ	copie <sub>1.1</sub>	نسخ1	copier <sub>1</sub>
1.1.وصول	accès <sub>1</sub>	X	
1.2.وصول	accès <sub>2a</sub>	X	

