
L'analyse et l'annotation à base de FrameNet : contribution à l'étude contrastive des événements de mouvement en arabe et en anglais

Abdelaziz Lakhfif*, **Mohamed Tayeb Laskri****

* LRSD, département d'informatique, Université Ferhat Abbas Sétif I, Algérie.

** LRI, département d'informatique, Université Badji Mokhtar – Annaba, Algérie.

abdelaziz.lakhfif@univ-setif.dz, laskri@univ-annaba.org

RÉSUMÉ. Dans cet article, nous présentons une nouvelle approche computationnelle d'analyse et d'annotation de la langue arabe. L'approche proposée est fondée sur la théorie des frames sémantiques de Fillmore (Frame Semantics). Nous abordons la question de l'applicabilité de cette théorie à la langue arabe qui diffère typologiquement de l'anglais, langue sur laquelle la théorie a été fondée à l'origine. Nous allons aussi explorer l'utilisation de cette approche pour l'analyse sémantique de l'arabe, notamment, l'annotation en rôles sémantiques et son application à l'analyse contrastive des événements de mouvement et de déplacement en arabe et en anglais, en utilisant un outil développé spécialement afin de répondre aux spécificités de l'arabe et de satisfaire les principes du projet FrameNet. En dépit des différences entre l'arabe et l'anglais sur plusieurs aspects, allant du type d'écriture à la typologie de langue, les résultats de notre projet confirment, une fois de plus, la nature cross-langues de la théorie des frames sémantiques. Les résultats de ce travail sont fondés sur l'analyse d'un corpus constitué dans sa majorité des expressions du domaine des mouvements et déplacements.

ABSTRACT. In this paper, we describe a Frame Semantics based computational approach for Arabic language processing. We explore the representational of Frame Semantics approach to Arabic text semantics, the adaptability of Berkeley FrameNet database and the transferability of FrameNet tools for Arabic, a language that differ typologically from English. We describe our attempts to build an equivalent Arabic FrameNet and the use of such a semantic resource for Arabic text semantic analysis, representation and annotation. Here we present a frame based contrastive study of motion events expression in bilingual text (English-Arabic) using our FrameNet based tool for semantic annotation. Our study results confirm the cross-linguistic nature of Frame Semantics approach and the suitability of the theory for Arabic processing. The current work is based on an analysis of a representative corpus of motion events expressions.

MOTS-CLÉS : frames sémantiques, langue arabe, TAL, rôles sémantiques.

KEYWORDS: Frame Semantics, Arabic language, NLP, Semantic roles.

1. Introduction

La théorie des *frames* ou des *cadres* (Baker, 2009) sémantiques (*frame semantics*) (Fillmore, 1982 ; Fillmore et Baker, 2010), a connu, depuis son apparition, un essor considérable et elle est devenue rapidement une source d'inspiration pour des centaines de travaux de recherche dans le domaine du TAL. La théorie de Fillmore, dont l'idée était largement inspirée des travaux de recherche sur les formalismes de représentation des connaissances dans les domaines de l'intelligence artificielle et de la psychologie cognitive, durant les années soixante-dix, représente une évolution naturelle de son idée originale sur la grammaire des cas (Fillmore, 1968). La théorie affirme que « *les gens comprennent, en grande partie, la signification des mots en se référant aux frames (sémantiques) qu'ils évoquent* » (Ruppenhofer *et al.*, 2016a). Par exemple, pour comprendre le sens du mot anglais 'Arrive' (arriver), il faut en premier lieu avoir des connaissances sur le processus du mouvement, qui évoque principalement la notion de déplacement d'une entité (*Theme*) occupant une location initiale (*Source*), tout au long d'un chemin (*Path*), vers une location finale (*Goal*) et ainsi de suite (section 2). La théorie s'inscrit dans le domaine de la sémantique empirique qui accorde une grande importance à la relation entre la langue et l'expérience (Petrucci, 1996), où un mot prend sens dans un contexte conventionnel interprété par le biais d'une structure cognitive (frame) évoquée dans l'esprit des interlocuteurs, plutôt qu'explicite (Baker, 2009). Du point de vue lexicographique, la théorie, cherche à associer des formes linguistiques à leurs structures cognitives (frames) sous-jacentes décrivant leurs usages validés par un usage langagier.

L'application pratique de la théorie est la base sémantique lexicale bien connue FrameNet (Baker *et al.*, 1998 ; Fillmore *et al.*, 2003), une ressource lexicale disponible en ligne conçue et élaborée sur la base des principes des frames sémantiques. La base FrameNet est construite à partir d'un ensemble de frames organisées hiérarchiquement et considérées comme étant des classes sémantiques regroupant des unités lexicales, ainsi que sur un réseau de relations frame à frame, fournissant la possibilité de représenter des scénarios larges (Burchardt, 2008) permettant ainsi de capturer le sens d'un discours composé de plusieurs phrases. Malgré le fait que la base Berkeley FrameNet (BFN) ait été élaborée à l'origine pour l'anglais, la nature « cross-langue » des frames (Boas, 2009 ; Fillmore et Baker, 2010) et la libre disponibilité de la base ont joué pour motiver d'autres chercheurs à réutiliser la base BFN et ses outils d'annotation pour l'analyse d'autres langues (Pitel, 2009 ; Boas, 2009), rendant ainsi FrameNet la ressource lexicale et sémantique la plus intéressante en termes d'utilisation cette dernière décennie.

L'approche des frames sémantiques a été expérimentée avec succès dans plusieurs axes de recherche dans le domaine du TAL, principalement dans la construction des

bases lexicales multilingues (Boas, 2002 ; Torrent *et al.*, 2014), l'alignement de différentes ressources lexicales sémantiques (Shi et Mihalcea, 2005 ; Ferrandez *et al.* 2010 ; Baker *et al.*, 2017), l'extraction de connaissances à partir de textes (Nuzzolese *et al.*, 2011 ; Sjøgaard *et al.*, 2015), la construction d'ontologies (Reed et Pease, 2015), l'analyse sémantique de surface (étiquetage en rôles sémantiques) (Gildea et Jurafsky, 2002 ; Erk et Padó, 2006 ; Johansson et Nugues, 2008 ; Das *et al.*, 2014 ; Lakhfif et Laskri., 2015 ; Hartmann *et al.*, 2017), la catégorisation automatique de textes (Moschitti, 2008), les systèmes questions-réponses (Narayanan et Harabagiu, 2004 ; Shen et Lapata, 2007), la reconnaissance des paraphrases (Padó et Erk, 2005) et l'implication textuelle (Burchardt *et al.*, 2007 ; Burchardt et Pennacchiotti, 2017), la traduction automatique (Gimenez et Marquez, 2007 ; Wu et Fung, 2009, Lakhfif et Laskri, 2015 ; 2016), l'analyse des sentiments (Ruppenhofer et Michaelis, 2016b) et l'extraction d'événements (Agarwal *et al.*, 2014) dans les réseaux sociaux. L'approche a été expérimentée aussi dans des domaines spécialisés à l'image du projet « Kicktionary » (Schmidt, 2007) pour l'analyse des commentaires footballistiques, la génération de scènes visuelles à partir du texte (VigNet) (Coyne, *et al.*, 2010), le projet « BioFrameNet » (Dolbey *et al.*, 2006), une extension de BFN vers une ontologie biomédicale et le projet de représentation et d'annotation de textes dans le domaine juridique (Venturi *et al.*, 2009). Cependant, l'application de la théorie des frames sémantiques sur d'autres langues se heurte à des difficultés liées à l'applicabilité de l'outil d'annotation et d'affichage, conçu initialement pour l'anglais, sur d'autres langues (le chinois, l'arabe, etc.) et à la capacité des frames à représenter les divergences des langues dans la représentation linguistique des conceptualisations. Parmi les contraintes que l'approche de Fillmore doit résoudre dans un contexte multilingue, on trouve le phénomène de la divergence des langues (Dorr, 1993). La divergence dans l'expression d'un événement dans les langues se produit lorsque le sens sous-jacent est réparti sur différentes formes linguistiques, provoquant ainsi des transformations touchant un des aspects de l'expression tels que la structure, le lexique, la catégorisation des mots, etc. (Habash et Dorr, 2002). Dans cet article nous allons présenter les premiers résultats de l'adaptation de la théorie des frames sémantiques pour le traitement de la langue arabe, en nous intéressant, en particulier, à la conceptualisation des événements de mouvement et de déplacement en termes de frames (Ellsworth *et al.*, 2006 ; Ohara, 2007 ; Petruck, 2008) dans différentes langues (anglais, français et arabe). Nous allons montrer que la théorie des frames sémantiques dispose d'un mécanisme, fondé sur les relations *frame à frame*, pour la résolution de certaines divergences liées à la typologie des langues (Talmy, 1991).

Après une présentation concise de la notion des frames sémantiques (section 2), nous présentons dans la section 3, notre tentative de génération d'une base FrameNet pour l'arabe. Nous commençons par une brève introduction sur les caractéristiques de l'arabe. Ensuite, nous abordons la question de l'utilisation des ressources lexicales sémantiques

dans notre projet et l'application des principes de la théorie dans l'analyse et l'annotation de l'arabe moderne standard (MSA). La section 4 présente une étude contrastive de la conceptualisation des expressions des événements de mouvement selon le point de vue des frames sémantiques. Nous terminons par une conclusion dans la section 5.

2. Les frames sémantiques et le projet FrameNet

Une frame sémantique (figure 1) est une représentation schématique d'une situation, d'un objet, d'une relation ou d'un événement impliquant des participants sémantiques et regroupant tous les mots (FEE) qui peuvent évoquer cette situation dans une phrase. Chaque frame sémantique est associée avec un ensemble de participants sémantiques, appelés « éléments de la frame » (*frame elements*) (FE) (Pado et Pitel, 2007), qui représentent le sens véhiculé par les arguments syntaxiques du prédicat de la phrase. Fillmore et Baker (2010) attestent que la notion de frames est largement indépendante des langues, et qu'elle est plutôt proche de la notion de scripts (Schank et Abelson, 1977) et de la notion des frames proposée par Minsky (1975) qui représentent des situations stéréotypes et qui jouent un rôle important dans le processus de perception, de mémorisation et de raisonnement. Cependant, à la différence des ces anciens formalismes, qui ne tiennent pas compte du processus de la production langagière dans la mise en place de ses structures cognitives (Fillmore *et al.*, 2003 ; Fillmore et Baker, 2010), les frames de Fillmore sont considérées selon une perspective linguistique et, par conséquent, visent à fournir un cadre descriptif de l'interface entre le sens et les formes linguistiques (support du sens véhiculé).

Arriving	
Définition	<p>An object Theme moves in the direction of a Goal. The Goal may be expressed or it may be understood from context, but its is always implied by the verb itself.</p> <p>Our visitors ARRIVED yesterday. Amy ARRIVED home from school early one afternoon. The senator ARRIVED to a standing ovation.</p>
Core FE	<ol style="list-style-type: none"> 1. Goal [Goal] Semantic Type : Goal 2. Theme[Thm] Semantic Type : Sentient
UL anglaises	<p><i>approach.n, approach.v, arrival.n, arrive.v, come.v, crest.v, descend_(on).v, enter.v, entrance.n, entry.n, get.v, influx.n, make it.v, make.v, reach.v, return.n, return.v, visit.n, visit.v</i></p>

Arriving	
UL arabes	، أَقْبَلَ ، أَقْبَلَ ، رَجَعَ ، دَخَلَ ، اِنْتَهَى إِلَى ، وَصَلَ ، بَلَغَ ، دَنَا ، اِفْتَرَبَ ، اَبَ ، اِقْتَحَمَ ، تَوَجَّهَ إِلَى ، زَارَ ، عَادَ ، نَزَلَ بِ ، وَفَدَ ، قَدِمَ ، جَاءَ ، اِنْتَقَلَ إِلَى

Figure 1. Une illustration sommaire d'une entrée de la base FrameNet : 'Arriving'

La notion de la frame est à la base de notre compréhension des mots de la langue où un mot est associé à un cadre descriptif de son usage (frame sémantique). L'objectif principal de l'approche des frames sémantiques est la description de différentes significations véhiculées par un lexème, en s'appuyant sur des structures cognitives (frames), où chaque frame encapsule un de ses sens véhiculés, ainsi que les constructions grammaticales possibles en énumérant toutes les réalisations syntaxiques supportées par un usage langagier des rôles sémantiques joués par les différents arguments syntaxiques liés au lexème donné (le prédicat). Ces rôles sémantiques (FE) représentent les participants sémantiques spécifiques à la frame. Les FE nécessaires pour la compréhension de la situation sont des FE noyaux 'Cores'. Les frames sont reliées entre elles par des relations frame à frame (*inheritance, using, subframe, precedes, causative of, etc.*) (figure 2).

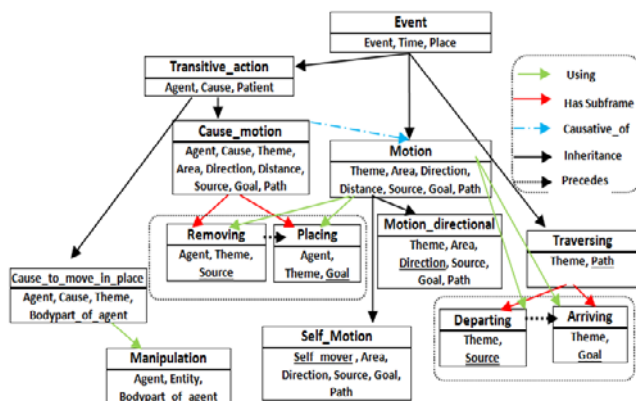


Figure 2. Une représentation partielle des relations frame à frame

Ces relations (BFN contient 1 687 relations) sont d'une importance capitale pour les applications à base de frames qui nécessitent des mécanismes d'inférence. Revenons sur notre exemple initial (section 1) concernant le processus mental de compréhension du

mot anglais ‘Arrive’ (arriver), l’image cognitive sous-jacente est organisée linguistiquement autour de la notion des frames. Dans notre cas, la frame caractérisant partiellement cette situation est appelée ‘Arriving’ et les participants sémantiques jouant les rôles thème (*Theme*) et but (*Goal*) sont indispensables à l’instanciation de la frame, parmi d’autres participants secondaires. Cette frame ‘Arriving’ avec la frame ‘Departing’ sont des sous-frames de la frame complexe ‘Traversing’ qui utilise (*using*) la frame ‘Motion’, qui de son côté caractérise un mouvement plus général que la frame ‘Arriving’. La frame ‘Arriving’ représente le mouvement d’un point de vue de la destination (*Goal-profiling*), alors que la frame ‘Departing’, qui est reliée à ‘Arriving’ avec la relation ‘Precedes’, représente le côté source (*Source-profiling*).

La frame ‘Arriving’ peut être évoquée par un groupe de mots partageant les mêmes caractéristiques sémantiques essentielles par rapport à l’image cognitive (frame), tels les verbes : *arrive* (arriver), *reach* (atteindre), *enter* (entrer), *visit* (visiter), etc. En général, un mot peut avoir plusieurs sens et, par conséquent, peut évoquer différentes frames. Le verbe ‘reach’ (atteindre) est un mot avec plusieurs sens et, par conséquent, il appartient à plusieurs frames. Parmi les sens du verbe ‘reach’ (atteindre) un qui correspond à la frame ‘Contacting’, dans laquelle un communicateur ‘Communicator’ adresse une ‘Communication’ à un destinataire ‘Addressee’ via une adresse particulière ‘Address’ (courriel, téléphone, etc.). Le reste des sens sont liés aux frames ‘Path_shape’ et ‘Body_mouvement’ respectivement. Dans ce cas de figure, la paire (lemme, frame évoquée) correspond à une unité lexicale (UL).

Le développement continu de la base BFN permet d’accroître le niveau de sa couverture linguistique, elle contient actuellement (juin, 2018) plus de 1 087 frames lexicales caractérisant plus de 13 640 UL. La base comporte aussi plus de 200 000 ensembles d’annotations de phrases extraites à partir du corpus « *British National Corpus* »(BNC), validant l’usage langagier des mots.

2.1. Procédure et outils d’annotation

La procédure d’annotation en frames sémantiques au sein du projet BFN consiste à associer les formes linguistiques avec leurs composants sémantiques (FE) de la structure cognitive (Fillmore et Baker, 2010).

Le projet BFN dispose d’un système client-serveur développé au centre ICSI Berkeley pour la création des frames, l’ajout des UL, et l’annotation en rôles sémantiques. Les données du système sont stockées dans une base de données relationnelle (MySQL). L’affichage des informations sur les frames et sur les annotations est assuré par un outil visuel à base de technologies Web. BFN a suivi une stratégie d’annotation qui consiste à procéder frame par frame dans la création des UL et l’annotation des phrases, tout en limitant le nombre des types de phrases (PT) et des

fonctions grammaticales (GF), afin de faciliter la tâche d'annotation. L'annotation de textes consiste à sélectionner pour chaque UL, des exemples de son utilisation sous forme de phrases à partir du corpus BNC qui attestent son usage et son adéquation avec le sens caractérisé par la frame. Une fois la phrase sélectionnée et importée dans l'outil d'annotation, l'opération d'annotation en rôles sémantiques va consister à regrouper en plusieurs couches et pour chaque phrase annotée l'étiquette représentant l'élément de la frame (FE) avec sa fonction grammaticale (GF), et son type de phrase (PT) sous la forme d'un triplet. Le processus d'annotation offre la possibilité d'ajouter d'autres couches pour les informations non couvertes par les couches précédentes (prédicats, relatifs, etc.).

وَصَلَ. V

Frame: Arriving

Definition:

وَصَلَ - [و ص ل]. (ف: ثلا. لازمتع. م. بحرف) إليه وُصِلًا وُوصِلًا وُصِلَ: بَلَّغَهُ وَاثْبَهَى إِلَيْهِ: QM

Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

Frame Element	Number Annotated	Realization(s)
Goal	(8)	PP[إلى].POBJ (5) .DNI (1)

- وَصَلَتْ الْحَالَةَ إِلَى بَابِ الدَّارِ وَضَعَطَتْ عَلَى الرِّزِّ فَأَضَاعَتْ كُلَّ أُنْوَارِ الْبَيْتِ وَ بَخِخَ الْجَمِيعَ بِمَسْتَقْبَلِهَا
- لَمَّا وَصَلَ الْكِتَابَ إِلَى رُبَيْدَةَ ، أَخَذَهَا أُزْجِيَةَ النَّدَى
- إِذَا وَصَلَ إِلَى مَقَرِّ الرِّئَاسَةِ ، وَبَعْدَ الْمَرَاجِعِينَ قَدْ أَخَذُوا أَفْكِنْتَهُمْ
- لَمَّا وَصَلَ الْحَمِيمِينَ فَتَوَجَّهَ الْأَصْدِقَاءُ إِلَى الضَّيِّعَةِ وَ بَعْدَ سَاعَةٍ مِنَ السَّيْرِ وَصَلُوا إِلَىهَا

Figure 3. Une partie de la table de réalisation pour “وَصَلَ (arriver)” « arrive »

La figure 3 montre une partie de la table récapitulative des réalisations syntaxiques des rôles sémantiques (FE) pour l'unité lexicale arabe وَصَلَ (arriver) « arrive » avec des exemples de son usage langagier, ainsi que sa définition extraite à partir d'un dictionnaire spécialisé (AL- Qamoos AL-Muheet ou « Le Dictionnaire complet », Al-Firoz-Abadi (1329-1415 AD)). Ce rapport offre ainsi aux utilisateurs un accès rapide aux informations sur la correspondance entre les formes syntaxiques et les composants sémantiques de la phrase.

Les données du projet BFN sont réparties entre deux bases différentes. La base lexicale qui regroupe toutes les informations sur les frames et les différents types de relations entre ces frames ainsi que les informations sur les UL et la base d'annotations

qui regroupe toutes les phrases annotées avec les sous-corpus à partir desquels ces mots sont extraits. Ces annotations riches en informations lexicographiques et en valences de mots sont stockées avec l'ensemble des valences de l'UL dans un fichier XML. Ce fichier XML contient aussi des données sur les modèles de valences des prédicats (UL), générées automatiquement durant l'opération d'annotation et qui peuvent être exploitées par des applications TAL.

2.2. La dimension cross-langue de la théorie des frames sémantiques

Suite aux succès de la théorie des frames sémantiques et de sa base lexicale FrameNet, l'équipe de Berkeley essaye actuellement de répondre aux questions liées à l'extension du projet BFN vers d'autres langues, notamment vers les langues dont la typologie est différente de celle de l'anglais, telles que l'espagnol, le français et les langues sémitiques (l'arabe, l'hébreu, le maltais, etc.). Ainsi, les chercheurs s'interrogent pour savoir dans quelle mesure les structures cognitives, ou frames sémantiques développées initialement pour l'anglais, sont universelles, et s'il existe des frames spécifiques à chaque langue ou à chaque classe de langues. Fillmore (1982) et d'autres chercheurs (Boas, 2005 ; Burchardt *et al.*, 2009) ont défendu la nature cross-langue de la théorie en avançant le fait que la description conceptuelle des frames, qui sont dans la majorité des cas des structures cognitives, est indépendante de la langue. Cependant, il est prévisible que certaines frames de la base BFN sont liées à l'anglais, principalement dans le cas de la description de certaines situations ou de certains événements liés à la culture et aux traditions américaines. Nous allons montrer (section 3.1) que l'utilisation et l'adaptation de la théorie des frames sémantiques pour l'arabe, langue sémitique qui diffère largement de l'anglais, doivent prendre en considération certaines caractéristiques linguistiques non rencontrées dans les langues indo-européennes.

3. Un FrameNet pour la langue arabe

Le succès du projet FrameNet pour l'anglais a encouragé le lancement des projets similaires pour d'autres langues, à l'image des projets pour l'espagnol (Subirats et Petruck, 2003), le japonais (Ohara *et al.*, 2004), le chinois (You et liu, 2005), l'allemand (Burchardt *et al.*, 2009), le brésilien (Salomão, 2009), l'italien (Johnson et Lenci, 2011), le danois (Bick, 2011), le coréen (Kim *et al.*, 2016), le finnois (Lindén *et al.*, 2017), et enfin le français. En effet, récemment ASFALDA, un FrameNet pilote pour le français (Candito *et al.*, 2014), a été mis en ligne. En utilisant l'outil SALTO (Burchardt *et al.*, 2006), le projet ASFALDA a opté, en première étape, pour une démarche ciblant

l'analyse et l'annotation d'un corpus orienté (Djemaa *et al.*, 2016) couvrant les sept domaines de discours (communications, transactions commerciales, relations spatiales, etc.) les plus discutés au sein de la communauté FrameNet. Cependant, ce genre de tentatives pour les langues sémitiques, spécialement pour l'arabe, n'a pas atteint le niveau attendu par rapport à cette importante ressource sémantique. En effet, la première étude visant l'application de la théorie de Fillmore sur une langue sémitique a été lancée par Petruck (2008), et, récemment, un FrameNet pour l'hébreu a vu le jour grâce à Hayoun et Elhadad (2016). Le projet suit les principes de développement adoptés par BFN, en utilisant la même base des frames avec leurs définitions, FE et toutes les relations frame à frame.

En revanche, très peu de travaux de recherche se sont intéressés à l'analyse de l'arabe sous la lumière des frames sémantiques. En effet, le travail de Sharaf et Atwell (2009) représente la première tentative sérieuse de l'utilisation de la théorie des frames sémantiques pour l'analyse de l'arabe en essayant de développer un prototype FrameNet explorant les valences des verbes extraits du Coran. De son côté, Alshehri (2014) a effectué une analyse contrastive (anglais et arabe) sur un petit corpus de textes contenant cinq verbes (*walk, run, fly, climb* et *crawl* et leurs équivalents en arabe) de la frame *Self_motion* dans le but de déterminer la capacité et la généralité de la structure de la frame à représenter les deux langues. En effet, le travail de Lakhfif et Laskri (2015) est la première tentative, grandeur nature, de la génération d'un *Arabic FrameNet*. Ce projet, qui a débuté en 2011, vise à construire une base lexicale sémantique pour l'arabe à partir de la base originale, en utilisant la ressource *Arabic WordNet* (Black *et al.*, 2006) (AWN) comme passerelle et en suivant la méthodologie de transfert recommandée par BFN. Le processus d'annotation consistait à extraire des phrases attestant l'usage des UL induites en se fondant sur le corpus *CCA Corpus of Contemporary Arabic* (Al-Sulaiti et Atwell, 2006), en plus d'autres sources scolaires et religieuses. La collection des exemples et l'annotation semi-automatique d'informations morphosyntaxiques ont été effectuées par trois personnes qualifiées en linguistique arabe, cependant que l'annotation en rôles sémantiques était faite par un expert. Cette base sémantique a été utilisée pour l'analyse, la représentation sémantique et l'annotation (Lakhfif, *et al.*, 2013) des phrases en arabe (ASM). Ce projet (thèse de doctorat) entre dans le cadre du développement d'un outil de traduction automatique de l'arabe vers la langue des signes algérienne (LSA) (Lakhfif, 2016).

Langue	# Frames lexicales	# Unités lexicales	# Ensembles annotés
Anglais (Berkley FrameNet)	1 087	13 640	174 022
Espagnol	–	1 268	11 000
Suisse	1 200	38 028	9 000
Japonais	565	8 500	60 480
Finnois	938	6 639	40 721
Brésil	179	196	12 000
Hébreu (Hayoun et Elhadad)	167	3 000	500
Arabe (Lakhfif et Laskri)	> 600	> 10 000	> 3 000
DiCoInfo : A Framed Version (Ghazzawi)	57	106	–

Tableau 1. Comparaison de l'état de certains projets FrameNet nationaux

Dans sa thèse de doctorat, Ghazzawi (2016), tout en s'appuyant sur la ressource BFN, a ajouté à la base DiCoInfo (L'Homme, 2008) la traduction arabe des mots appartenant au domaine de l'informatique. Cependant, malgré le fait que ce projet adopte l'approche de Fillmore, l'annotation en frames sémantiques est un peu différente de celle du projet BFN, ainsi que la codification des informations annotées. Le tableau 1 donne un résumé de la couverture de certains projets FrameNet nationaux, dont la majorité ont adopté une approche étendue (Vossen, 1999) dans leurs processus de construction, gardant ainsi l'intégralité de la base originale pour l'anglais (BFN), au vu des difficultés de développement de ce genre de ressource qui nécessite des efforts multidisciplinaires et qui consomme du temps.

3.1. L'arabe et les technologies du TAL

La langue arabe est classée à la quatrième position parmi les dix premières langues utilisées sur Internet (<http://www.internetworldstats.com/stats7.htm>). Avec l'émergence des réseaux sociaux comme facteur d'influence sociale, économique, et politique, le nombre d'utilisateurs de l'arabe (standard, moderne et dialectes) a été multiplié par trois, en moins de dix années (on est passé de 60 millions d'utilisateurs, en 2008, à 197 millions, en décembre 2017). L'importance de l'arabe est accentuée par le nombre de pays (22 pays, et plus de 420 millions d'habitants) qui l'utilisent comme langue officielle. En plus des pays arabes, l'arabe (classique) est la langue du Coran, livre sacré de l'islam, religion regroupant plus de 1,7 milliard de personnes dans le monde. L'arabe est une langue sémitique caractérisée par une riche morphologie inflectionnelle et un système de dérivation à base d'une douzaine de schèmes. L'arabe est une langue à sujet

pronominal vide « *pro-drop* », où la morphologie du verbe intègre un pronom sujet et peut, dans certains cas, intégrer aussi un pronom objet (encliticisé). En général, l'arabe est une langue cliticisante, où les prépositions et les conjonctions peuvent être aussi procliticisées au mot arabe. Par exemple (tableau 2), le mot arabe (*fa-jA-at-hu*) consiste en la conjonction “ف(fa)” “puis- et”, un verbe “جاء (jAa) venir” conjugué dans le passé avec une inflexion pour un pronom sujet à la première personne du féminin du singulier et avec une inflexion pour un pronom objet à la troisième personne du singulier masculin.

Arabe : فِجَاءَتْهُ
Translittération : <i>fà-jA-at-hu</i>
POS : <i>Conj--V.pass--Pro.1sf(subj)--Pro.3sm(obj)</i>
Glose-fr : <i>puis--venir--elle--lui (puis elle vint à lui)</i>

Tableau 2. La complexité de la morphologie arabe

La question du transfert des outils du projet BFN pour la représentation, l'analyse et l'annotation de l'arabe représente un défi majeur, vu les différences entre les deux langues, allant de la transcription, l'orientation, la flexion et l'agglutination (Ryding, 2005 ; Habash, 2010), en passant par les contraintes sur l'ordre des mots dans la phrase, jusqu'au système de dérivation. Parmi les traits linguistiques décisifs dans le traitement automatique de l'arabe, on note la terminaison et les voyelles de la fin des mots qui contribue à la définition des cas syntaxiques. Cependant, si l'on regarde la flexibilité dans la distribution des arguments syntaxiques au sein de la phrase arabe, on observe que la position des arguments syntaxiques (agent, objet, etc.) n'est pas un critère grammatical déterministe. L'arabe dispose aussi d'un mécanisme dérivationnel lui permettant de générer des catégories nominales ((le participe actif) اسم الفاعل, (forme infinitive) مصدر, et (le participe passif) اسم المفعول, etc.) qui peuvent occuper des fonctions grammaticales importantes (adjectif, adverbe, etc.), et jouer les mêmes rôles que le verbe, comme avoir un sujet, des objets, etc.

3.2. Génération d'une base FrameNet pour l'arabe

L'émergence, vers la fin de ce dernier siècle, des ressources lexicales et sémantiques, telles que Princeton WordNet (Fellbaum, 1998) (PWN), FrameNet et VerbNet (Kipper-Schuler, 2005), représente l'une des avancées les plus importantes dans le domaine du TAL. Ces bases lexicales sémantiques sont devenues indispensables dans les applications nécessitant une représentation profonde de la sémantique de la phrase (Fellbaum *et al.*, 2007), à l'image de la compréhension du texte, de la traduction

automatique, etc. Cependant, la seule ressource lexicale sémantique, disponible actuellement, pour l'arabe est la base « Arabic WordNet » (AWN) (Black et al., 2006). AWN est une base lexicale pour les mots arabes, inspirée du réseau sémantique PWN et qui suit la méthodologie de développement de la base EuroWordNet (Vossen, 1998). L'avantage de la base AWN est dû à sa liaison avec PWN (3.0) en termes d'équivalence entre synsets, permettant ainsi la projection des classes de l'ontologie SUMO des mots anglais vers les mots arabes. Parmi ces relations (tableau 3), la relation « *Hyponymy* » qui relie deux mots par la relation taxonomique (*has_hyponym*) et la relation d'implication « *sub_event* » liant un verbe à un autre verbe décrivant un sous-événement de l'événement du premier verbe.

Algorithme 1 – L'algorithme d'induction des LU arabes

Entrée : une base des alignements (lemme, PoS, des LU_Synset [ID, Frame + PWN_Mots])

/*

enSSID : le synset ID (SSID) anglais ; arSSID : le synset ID arabe

al_LU_SS : alignement d'un LU avec des synsets WN3.0

L_AW : la liste des LU arabes induites à partir de AWN : lemme, forme, pos, LU anglais, sumo, relation

L_ESS : liste des enSSIDs ; L_ASS : liste des : arSSID, sumo, AW relation

*/

L_AW = {} /*initialiser la liste des LU arabes induites à partir de AWN */

Pour chaque alignement faire

Pour chaque al_LU_SS faire

L_ESS = Lire_tous_les_enSSID_de_(lemme) /* tous les SSID anglais*/

Ajouter tous_les_enSSID_similaires à L_ESS /* relation WN 'similar'*/

en_SUMO = Lire_le_concept SUMO du enSSID

Pour chaque enSSID dans L_ESS faire /*synsetid anglais*/

arSSID = Lire le SSID arabe équivalent dans AWN

ajouter tous les mots de arSSID dans L_AW /*mots arabes */

ar_SUMO = le concept SUMO du synset arabe

Si(PoS = 'v' ou Pos = 'n') **Alors** /*la relation related_to*/

| Ajouter les SSID arabes reliés à arSSID avec la relation 'related_to' à la liste L_ASS

Fin Si

Ajouter les SSID arabes reliés à arSSID avec la relation 'has_derived' à la liste L_ASS

Ajouter les SSID arabes reliés à arSSID avec la relation 'has_hyponym' à la liste L_ASS

Si(PoS = 'v') **Alors** /*la relation 'verb_group', 'sub_event' */

| Ajouter les SSID arabes reliés à arSSID avec la relation 'verb_group' à la liste L_ASS

| Ajouter les SSID arabes reliés à arSSID avec la relation 'sub_event' à la liste L_ASS

Fin Si

Pour chaque arSSID dans L_ASS faire /*synsetid arabes, sumo, relation*/

| Lire le concept SUMO de arSSID

| **Si** (ar_SUMO subsume le concept SUMO de arSSID) **Alors**

| | **ajouter** tous les mots du sysnset **arSSID** dans **L_AW**

| | /*lemme, forme, pos, LU anglais, sumo, relation WN*/

| **Fin Si**

Fin Pour

Fin Pour

Fin Pour

Fin Pour

Sauvegarder la liste **L_AW** dans un Fichier XML

Ces relations parmi d'autres relations sont utilisées dans notre algorithme (algorithme 1) d'induction des unités lexicales pour construire un FrameNet pour l'arabe. La création manuelle d'une base similaire à BFN s'avère une tâche fastidieuse pour les langues moins dotées en ressources lexicales et sémantiques. Dans notre projet, en plus de la traduction manuelle d'un ensemble de Frames (100 Frames) et les unités lexicales qui les appartiennent à partir de la base originale BFN, nous avons opté pour une approche étendue (Vossen, 1999), en utilisant la ressource AWN comme une source dans notre algorithme (algorithme 1) d'induction de la base FrameNet pour l'arabe (Lakhfif et Laskri, 2015 ; Lakhfif, 2016). Comme les ensembles de mots AWN sont traduits et liés directement aux ensembles des mots de la base PWN, nous avons exploité un algorithme d'alignement (Ferrandez *et al.* 2010) qui induit les UL de la base BFN (version 1.3) à partir des synsets de la base PWN (version 3.0), en se fondant sur les relations ontologiques telles que les relations “*equivalent*”, “*related_to*”, “*verb_group*”.etc. (figure 4).

Relations	Synsets	Mots arabes	Équivalents anglais
sub_event	balaEa_v1AR	بَلَعَ	<i>swallow</i>
sub_event	maDaga_v1AR	مَضَعُ	<i>chew</i>
related_to	>akol_n1AR	أَكَلَ	<i>food</i>
verb_group	>akol_v2AR	أَكَلَ، تَنَاوَلَ الطَّعَامَ	<i>eat</i>
has_hyponym	qaDama_v1AR	قَضَمَ	<i>nibble</i>

Tableau 3. Les relations associées au synset : >akala_v1AR أَكَلَ-*eat*'

Wn30synset ID = 2005948 PoS = verb corrélation = 1.09		
Synsetid = arrive_v1EN sumo = Motion + Members : <i>arrive, get, come</i>		
Gloss : reach a destination ; arrive by movement or progress ;		
frame = <i>Arriving</i> ID = 940 lemma = <i>arrive</i> PoS = <i>v</i> Sumo = Motion +		
Projection		
Relation	Lemmes arabes	SynsetID
equivalent	وَصَلَ	waSala_v4AR
related_to	مَجِيءٌ - قُدُومٌ , وَصُولٌ	quduwm_n1AR, wuSuwol_n1AR

Figure 4. Une entrée de l'induction des lemmes arabes pour "Arrive.v"

Cet algorithme, fondé sur une mesure de ressemblance sémantique entre les mots, rapporte une précision de 77 %, ce qui lui permet d'obtenir le meilleur score parmi les algorithmes d'induction proposés jusqu'à présent. Les UL induites automatiquement sont passées par une phase de vérification manuelle, afin d'éliminer les unités qui n'ont pas d'exemple d'utilisation en arabe, attestant sa cohérence avec la frame d'appartenance. Une partie de l'évaluation de notre projection ainsi que le nombre d'UL dans la base originale BFN sont présentés dans le tableau 4 qui montre aussi le nombre d'UL dans la base BFN utilisées dans la projection et le nombre d'UL générées pour l'arabe à partir des AWN respectivement. Il est à noter que malgré le fait que la taille de la base BFN soit plus large de treize fois par rapport à la base AWN (tableau 5), l'induction valable des UL arabes a atteint un rapport avoisinant la moitié de la taille de la base BFN. Cela ne veut pas dire que notre alignement couvre 50 % de la base BFN (seul le tiers (32 %) des UL dans la base BFN avait des équivalents dans AWN), mais plutôt qu'il donne un indice de la richesse du lexique arabe.

Catégorie de frames	# Frames	# UL (verbes) dans BFN	# (En) UL utilisées	# (Ar) UL	Précision
Mouvement	35	765	202	689	0,687
Causation	10	189	60	188	0,534
Activités	5	30	12	43	0,699
Frames vérifiées	80	1 523	484	1 490	0,711

Tableau 4. La précision de l'induction par catégorie de Frames (Lakhfif, 2016)

	# lemmes	# verbes	# noms	# adjectifs	# autres
PWN 3.0	155 287	13 600	81 000	19 000	3 600
FrameNet	13 640	5 200	5 558	2 396	486
AWN	11 256	2 525	7 960	500	271
Ar. FrameNet	10 273	4 500	5 530	232	11

Tableau 5. Comparaison de la couverture de différentes ressources sémantiques

3.3. L'analyse et l'annotation de l'arabe en frames sémantiques

La base FrameNet offre des informations sur les mots (UL) anglais et sur les frames sémantiques qui les caractérisent sous forme de données (XML) exploitables par les applications TAL. Parmi les informations utiles offertes par cette base, un nombre important de phrases extraites du corpus BNC sont annotées avec la triple information (FE, GF et PT). Ces informations illustrent, avec un niveau de détail sans précédent (Fillmore *et al.*, 2003), comment ces rôles sémantiques sont exprimés à travers des compléments et des modifiants du prédicat déclencheur de la frame. À cet effet, le projet offre une base lexicale avec un outil logiciel d'annotation et de visualisation à base des technologies XML. L'outil d'annotation BFN a été réutilisé et porté avec succès par plusieurs projets d'annotation en frames sémantiques pour des langues indo-européennes. Cependant, vu les divergences entre l'arabe et l'anglais, la réutilisation de cet outil pose plusieurs problèmes techniques, liés principalement à l'écriture des caractères et des mots arabes (langue cursive), rendant l'adaptation des outils de BFN pour la visualisation et l'annotation de l'arabe inappropriée. À cet effet, et afin de répondre aux défis soulevés par les spécificités de la langue arabe (transcription, orientation, flexion, agglutination, etc.), nous avons opté pour le développement d'un outil d'analyse et d'annotation de textes arabes (Lakhfif *et al.*, 2013), un système multiniveau permettant l'analyse lexico-morphologique, syntaxique et sémantique de l'arabe (ASM). Le système intègre en cascade l'analyseur morphologique AraMorph (Buckwalter, 2002) et la ressource AWN dans un système d'analyse fournissant plusieurs couches d'informations (PoS, grammaticale, sémantique). L'intégration de ces outils et de ces ressources offre une riche description des phrases arabes à l'image d'AWN qui ajoute des informations ontologiques sur les mots et FrameNet qui fournit des informations sémantiques sur les arguments du prédicat. Ces informations et d'autres informations utiles sont stockées dans un fichier XML, compatible avec le système de codification du projet BFN, et peuvent être exploitées par d'autres applications TAL tierces.

3.3.1. *Traitement de certaines caractéristiques de l'arabe*

Dans notre application de l'approche des frames sémantiques sur l'arabe, et afin de prendre en charge certaines spécificités linguistiques, nous avons opté pour des solutions qui préservent l'intégrité de l'approche adoptée par BFN et qui prennent en charge les caractéristiques propres de l'arabe. L'un des paris concerne l'annotation en rôles sémantiques des pronoms incorporés dans la morphologie du verbe (inflexion). Ces pronoms occupent généralement la position de sujet et dans certaines constructions peuvent aussi occuper la position de complément d'objet. Ce phénomène linguistique, très répandu dans certaines langues à sujet nul (pro-drop) telles que l'espagnol, l'italien et les langues sémitiques, se sert de la richesse morphologique afin de réaliser certaines constructions complexes. En arabe, le pronom sujet caché joue un rôle important dans la résolution de l'anaphore et de la référence dans certaines phrases complexes où le sujet est omis (CNI). Par exemple, dans le verbe 'شَرَعْنَا-' (tableau 6, figure 5), le pronom sujet 'نا-نا' incorporé dans le verbe correspond au pronom pronominal non réalisé morphologiquement 'nous – نحن'. Ce pronom contient des informations sur la personne, le nombre et le genre (1PM). Dans notre projet, dans le cas d'absence de sujet lexical dans une phrase, nous avons opté pour l'ajout d'une étiquette grammaticale 'SBJp' pour 'sujet pro-drop' pour marquer ce type de pronom explicitement avec l'étiquette sémantique adéquate (rôle sémantique). Cette solution nous a permis de garder des traces sur les arguments dans la phrase. Le deuxième cas concerne l'annotation en FE des constituants syntaxiques non locaux dans les constructions grammaticales à montée (*raising*) et à contrôle. Dans ces constructions à deux prédicats (un prédicat gouverne un autre), chaque prédicat évoque une frame différente, et l'argument du premier prédicat est aussi considéré comme un argument pour le deuxième prédicat (cible). Dans l'exemple ci-dessous (Ruppenhofer *et al.*, 2016a), 'John' est aussi un argument du prédicat 'retaliate' et, par conséquent, doit avoir le rôle sémantique 'Offender'.

– We expect [John Avenger] to retaliate [against us Offender] [INI Punishment] [DNI Injury].

Target السَّيْرُ Target شَرَعْنَا Target [Activity_start] وَ [Self_motion] في [Agent ;pr-drop-1MP] [Activity] [Goal] نَحْنُ مِيرِيبِتْ هَاوَسْ
Glose : <i>et nous commençons à marcher vers 'Merripit House'</i>
we started to [walk Self_motion] [to Merripit House Goal]

Tableau 6. *Une clause avec une structure à contrôle (The Hound of the Baskerville)*

Dans BFN, la fonction grammaticale 'externe' (GF : Ext) est attribuée aux deux arguments ('John' et 'We') des deux prédicats. Cette démarche pose des problèmes d'ambiguïté pour les applications TAL qui utilisent les données BFN dans des

mécanismes d'inférence. Dans notre projet, et afin d'éviter cette ambiguïté, un argument est considéré uniquement vis-à-vis du prédicat qu'il gouverne, tout en étendant les fonctions grammaticales vers les constituants pronoms, attachés morphologiquement au verbe, et qui occupent la position syntaxique 'sujet' ou 'objet'. Cette solution donne une distinction claire entre les rôles sémantiques dans ce type de construction.

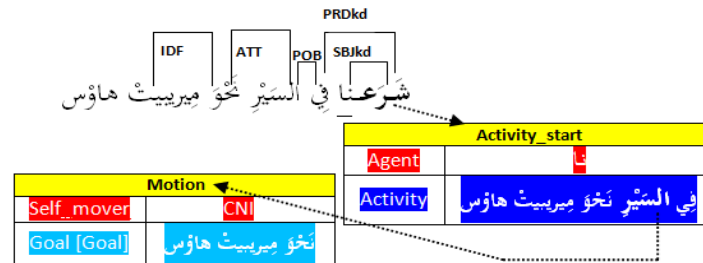


Figure 5. Conceptualisation en frames d'une clause arabe (sujet nul, éléments non locaux)

Dans l'exemple de la figure 5, le prédicat (la marche) السَّيْرِ déclenche la frame 'Self_motion', et son rôle sémantique 'Self_mover' doit être réalisé par le sujet du prédicat de la construction à contrôle, du verbe appartenant au groupe des verbes de rapprochement et de commencement (أفعال الشروع و المقاربة). La liaison entre le sujet (étiqueté par SBJkd c'est-à-dire sujet des verbes de rapprochement et de commencement) du premier prédicat 'à commencer- شَرَع' et le sujet du prédicat cible (la marche) السَّيْرِ est facilement établie en se fondant sur les informations d'agrément du pronom sujet caché (SBJp) qui fait référence sémantiquement à un argument non local. Le troisième cas de différence concerne l'annotation des verbes à préposition. Dans BFN, un verbe à préposition (*take off*, *take up*, etc.) est considéré avec sa particule comme une expression polylexicale 'multiword' et, par conséquent, cette particule sera marquée comme une partie du verbe dans la couche 'Target'. Cependant, dans notre projet, et en suivant les principes de la grammaire arabe, la préposition est marquée avec le syntagme prépositionnel qu'elle gouverne. Dans notre cas, l'association de la préposition avec le verbe dans l'annotation, génère des incohérences dans la description syntaxique du syntagme.

3.3.2. L'annotation de corpus

Suite aux succès des projets d'annotation de textes pour l'anglais, plusieurs tentatives ont été faites pour l'annotation de l'arabe. Les projets à l'image de Penn Arabic Treebank (PATB) (Maamouri *et al.* 2004), Prague ADTB (Hajic *et al.*, 2004), CATiB (Habash et Roth, 2009) et The Quranic Arabic Dependency Treebank (QADT) (Dukes *et al.* 2010) représentent les projets ayant le plus d'impact sur les travaux de recherche sur l'arabe. Cependant, à l'exception du projet Arabic PropBank (Diab *et al.*, 2008) qui

s'intéresse à l'annotation de la sémantique des arguments du prédicat, on note que l'annotation sémantique des corpus en langue arabe n'a pas encore attiré l'attention méritée. Dans notre projet d'annotation en frames sémantiques, le processus d'annotation de corpus suit les mêmes procédures de collection et d'organisation des données adoptées par le projet BFN qui est décrit en détail dans (Ruppenhofer *et al.*, 2016a). Pour chaque UL, les phrases contenant l'entrée avec le sens adéquat sont sélectionnées à partir de plusieurs sources. Comme notre objectif de l'annotation est de fournir une riche description de textes selon le niveau morpholexical, syntaxique ou sémantique intéressant pour les applications TAL (figure 6), notre outil semi-automatique d'analyse fournit plusieurs types de descriptions selon ces trois niveaux. Au début d'analyse, notre outil établit automatiquement une analyse complète de la phrase en entrée et propose tous les résultats possibles pour chaque lexème (ou groupe de lexèmes), dans chaque niveau d'analyse, selon le degré d'ambiguïté.

```
<text>مَنْزَعْنَا فِي الْمَيْتْرِ نَحْوُ مِيرِييْتْ مَاؤُس</text>
<annotationSet cDate="04/10/2017" status="S_MANUAL" ID="480">
  <layer rank="1" name="Target">
    <label cBy="Aziz" end="4" start="0" name="Target"/>
  </layer>
  <layer rank="1" name="FE">
    <label cBy="Aziz" feID="2542" end="6" start="5" name="Agent"/>
    <label cBy="Aziz" feID="2085" end="42" start="8" name="Activity"/>
  </layer>
  <layer rank="1" name="GF">
    <label end="6" start="5" name="SBJp"/>
    <label end="42" start="8" name="POBJ"/>
  </layer>
  <layer rank="1" name="PT">
    <label end="6" start="5" name="NP"/>
    <label end="42" start="8" name="PP"/>
  </layer>
  <layer rank="1" name="AWP">
    <label end="6" start="0" name="V;CAT:VPKd;TEN:PV;VOI:A;TRA:VPKd;GEN:_;1">
    <label end="6" start="5" name="SBJp;Drop;pron:ل;nA;pgn:1_P"/>
    <label end="12" start="10" name="P;CAT:PREP;TEN:_"/>
    <label end="21" start="14" name="N;DEF:D;CAT:N;NUM:_;GEN:M;PER:_;CAS:GI">
    <label end="29" start="24" name="P;CAT:ADVL;TEN:_"/>
    <label end="39" start="31" name="N;DEF:_;CAT:NOP;NUM:S;GEN:M;PER:1;CAS">
    <label end="45" start="41" name="N;DEF:_;CAT:NOP;NUM:S;GEN:M;PER:1;CAS">
  </layer>
```

Figure 6. Une représentation en XML pour une phrase annotée selon différentes couches

Par la suite, l'utilisateur choisit la solution adéquate parmi les suggestions, elle sera enregistrée comme un résultat final de l'analyse (figure 6) (Lakhfif, 2016).

Dans le niveau morphologique, en plus de la description des lemmes après segmentation, l'outil offre la possibilité d'annotation et de génération des fichiers annotés en symboles préterminaux (tags) les plus utilisés pour l'arabe tels que Penn TreeBank (Marcus *et al.*, 1994), BAMA (BuckWalter tags) et CATiB, QADT. Le système d'analyse et d'annotation génère automatiquement des couches d'annotations

concernant les classes SUMO pour les mots reconnus durant l'analyse à base d'AWN en plus des mots membres des ensembles « synsets » équivalents et leurs traductions anglaises 'Gloses'. Les données d'analyse et d'annotation sont organisées par unité lexicale, ce qui signifie que toutes les couches d'annotations pour une unité lexicale donnée sont enregistrées dans un seul fichier. Cependant, en plus de l'annotation style BFN qui s'appuie sur une représentation syntagmatique de constituants, l'annotation en *fonctions grammaticales* (GF) dans notre projet, et à l'instar des projets SALSA (Burchardt *et al.*, 2009), FrameNet danois, et ASFALDA, est fondée sur la grammaire de dépendance (Tesnière, 1959 ; Mel'čuk, 1988), une grammaire bien adaptée pour l'arabe, une langue ayant un réseau de relations de dépendance dans chaque phrase ou proposition (Ryding, 2005). L'annotation utilise un ensemble large d'étiquettes syntaxiques (supérieur à cinquante étiquettes) telles que le complément accusatif absolu (المفعول المطلق), les compléments circonstanciels (المفعول فيه), le complément de manière (الحال), etc., permettant ainsi d'améliorer la tâche de classification des arguments syntaxiques (Johansson et Nugues, 2008). Ce genre de relation est nécessaire dans certaines applications telles que la traduction automatique et la compréhension du texte.

4. Étude de cas : la théorie des frames sémantiques et le domaine des événements de mouvement et de déplacement.

Les langues naturelles présentent des divergences dans leurs descriptions des événements de mouvement et de déplacement, notamment dans les moyens linguistiques accordés à l'expression de la trajectoire et de la manière. Talmy (1991, 2000) propose de classer les langues, selon la façon d'exprimer le composant décrivant la trajectoire (*path*) du mouvement, en deux grandes familles : la famille des langues à cadre satellitaire (*satellite-framed languages* ou *s-languages*) (comme l'anglais), qui utilisent des particules (satellites) pour exprimer la trajectoire, comme stratégie de base, tandis que la manière est généralement confondue avec la sémantique du verbe (*run, limp, crawl*, etc.) ; la deuxième famille, concerne les langues à cadre verbal (*verb-framed languages* ou *v-languages*), (comme l'espagnol) qui fusionnent le concept de la trajectoire avec celui du mouvement au sein du verbe (*entrer, arriver, sortir*). Dans cette dernière famille, la manière du mouvement est généralement exprimée à travers des compléments de phrase (*il entre à la maison en courant*). Le français, l'italien, l'hébreu (Talmy, 1991) et l'arabe (Talmy, 1991 ; Bernini, 2010) sont classés parmi les *v-languages*. Par exemple, les traductions en arabe de *go in, go out* et *go down* sont نَحَلَ (*daxala*) « entrer », خَرَجَ (*xaraja*) « sortir » et نَزَلَ (*nazala*) « descendre » respectivement. Cependant, l'anglais et l'allemand, parmi d'autres, sont des *s-languages* connues pour leur utilisation intensive des verbes de manière dans leurs descriptions des événements de mouvement. Deux conséquences capitales résultent de cette divergence en typologie. La première

conséquence concerne la divergence dans l'expression des informations de la manière du mouvement, et la deuxième conséquence concerne la divergence dans l'expression des informations de la trajectoire du mouvement.

4.1. Une analyse contrastive à base de FrameNet

Depuis des décennies, le domaine des expressions de mouvement et de déplacement est devenu le champ préféré des études contrastives cross-langues. Parmi ces études, les travaux de Slobin (2006) et son équipe essaient de montrer l'importance accordée par la langue à l'expression de la manière et le degré d'élaboration des éléments de la trajectoire dans l'expression des événements de mouvement. Ces études ont porté sur des corpus narratifs multilingues. C'est dans cette perspective que le livre d'images *Frog story* et le chapitre 6 du roman *The Hobbit* (Tolkien, 1937) représentent les supports les plus utilisés dans les études contrastives sur la typologie des langues. De son côté, et afin de montrer l'apport des frames sémantiques dans la compréhension du texte, l'équipe du projet BFN a lancé plusieurs projets d'annotation des corpus constitués à partir de romans célèbres tels que *The Tiger of San Pedro* (Doyle, 1908) et le chapitre 14 de *The Hound of the Baskerville* (Doyle, 1901). Ce dernier a été choisi comme objet d'étude par Ellsworth *et al.*, (2006), Ohara (2007) et Petruck (2008), dans leurs analyses contrastives sous la lumière de la théorie des frames sémantique. Ellsworth *et al.*, (2006) ont effectué une étude comparative en termes de prédicats déclencheurs de frames dans la version anglaise et de ses traductions (espagnole et japonaise). En plus de la validation des hypothèses établies par Talmy et Slobin sur la typologie des langues, l'étude a révélé l'existence de différentes stratégies de conceptualisation d'un même événement. Cette divergence apparaît également au sein d'une même famille de langues. Par exemple, en anglais, une scène décrivant une action avec des participants peut se voir traduite, en japonais, en une expression décrivant la scène entière ou l'état résultant de l'action. De son côté, Ohara (2007) a mis l'accent sur l'importance de l'interaction entre la sémantique du prédicat déclencheur de la frame et ses constructions grammaticales dans la conceptualisation des événements. Ces études contrastives à base de FrameNet ont révélé que la structure conceptuelle des frames et les relations frame à frame permettent de comparer les langues avec plus de détails que dans les études contrastives fondées uniquement sur la typologie des langues.



Figure 7. Annotation multilingue de texte en rôles sémantiques

Dans cette direction de recherche, et afin de savoir si l'anglais et l'arabe conceptualisent les événements de mouvement avec les mêmes structures conceptuelles (frames), nous avons procédé à l'annotation multilingue (EN-FR-AR) en rôles sémantiques du chapitre 6 du célèbre roman *The Hobbit*, en utilisant notre outil d'annotation (Lakhfif *et al.*, 2013) (figure 7). Le tableau 7, qui résume les divergences en termes de conceptualisation (frames), montre les types et le nombre des frames évoqués dans les expressions de mouvement des créatures ressemblant aux humains (Bilbo, nains, etc.) dans la version originale et sa traduction en arabe. Notre étude révèle que parmi les soixante-douze frames utilisées dans la version originale, la version traduite en arabe a employé les mêmes frames dans soixante et une expressions. En effet, parmi les soixante expressions évoquant une '*Self_motion*' dans la version anglaise, uniquement quatre expressions dans la version arabe ont montré différentes frames, avec trois expressions qui étaient caractérisées par les frames '*Motion_directional*' et '*Arriving*' évoquées par les prédicats نَزَلَ (descendre), عَادَ (retourner), اقْتَرَبَ (s'approcher de), respectivement. Cette divergence est due à la différence en typologie des deux langues, car les verbes de manière '*climb*' et '*creep*' n'ont pas d'équivalent dans les *v-langues* et ils se traduisent généralement par des verbes directionnels. Cette remarque confirme l'hypothèse de Slobin (2004) qui affirme que les *v-langues* n'ont pas d'équivalent pour l'expression '*climb down*'.

Frames évoquées (EN)	Frames évoquées (AR)	# d'expressions	
<i>Self_motion</i>	<i>Self_motion</i>	56	-
<i>Self_motion</i>	<i>Motion_directional</i>	01	(نَزَلَ) - descendre
<i>Self_motion</i>	<i>Arriving</i>	02	(عادَ - اقْتَرَبَ) – retourner, s’approcher de
<i>Self_motion</i>	<i>Manipulation</i>	01	(تَعَلَّقَ) - s’accrocher
<i>Motion</i>	<i>Motion_directional</i>	02	(تَدَخَّرَجَ) - dégringoler
<i>Motion</i>	<i>Self_motion</i>	02	(انزَلَقَ) – se glisser
<i>Motion_directional</i>	<i>Motion_directional</i>	04	(سَقَطَ, وَقَعَ) - tomber
<i>Motion_directional</i>	<i>Cause_motion</i>	01	(أَوْقَعَ) – faire tomber
<i>Cause_to_move_in_place</i>	<i>Manipulation</i>	01	(تَعَلَّقَ) – s’accrocher
<i>Fleeing</i>	<i>Fleeing</i>	01	(هَرَبَ) – s’échapper
<i>Dispersal</i>	<i>Self_motion</i>	01	(تَفَرَّقَ) – se disperser

Tableau 7 Type de frames évoqués dans les expressions de mouvement des créatures ressemblant aux humains dans *The Hobbit* chap. 6 (EN-AR)

Un autre point important, que nous avons constaté, est que, lorsque la version originale utilise deux verbes de manière (*run*, *creep*) qui évoquent la frame ‘*Self_motion*’ pour caractériser un mouvement à manière en direction d’un but, exprimée via ‘*back – en arrière*’ et ‘*still nearer – plus près*’ respectivement, la traduction arabe se focalise sur le but à atteindre, en utilisant la frame ‘*Arriving*’ (qui caractérise le mouvement du côté but (*Goal-profiling*)). Cette frame ‘*Arriving*’ est évoquée par les deux prédicats عاد (*retourner*) et اقْتَرَبَ (*s’approcher*), respectivement. Cette façon de conceptualisation est prévisible dans la version arabe où la manière du mouvement est généralement omise, comme l’atteste la première expression (*run*). Cependant dans la deuxième expression (*creep*), la traduction était fidèle même dans la description de la manière, et ce, grâce à l’adverbe (participe actif) arabe زاحفا (*creeping*) « en rampant ». Il est à noter que cette stratégie représente une exception dans les *v-Langues* (la version française a omis la manière : ‘*Il se glissa encore plus près*’). En dépit des différences en termes de caractérisation en frames (*Self_motion* vs *Arriving*, qui partagent certaines caractéristiques de la frame générale ‘*Motion*’), les deux représentations ont capturé l’essentiel du sens (manière et location) véhiculé par chaque expression avec des points de vue différents. En résumé, 85 % des expressions ont été

caractérisées par les mêmes frames, présentant ainsi un parallélisme conceptuel remarquable vu la divergence dans la typologie.

	Goal	Source	Path	Area	Place	Direction
Anglais	21	14	37	10	5	1
Arabe	35	09	15	6	5	3

Tableau 8. *Distribution des éléments des frames (FE) (EN/AR)*

Cependant, le tableau 8 montre les conséquences de la divergence de typologie (v-langue vs s-langue) sur la réalisation sémantique des composants du mouvement dans les deux langues. Cette divergence est due aux différentes stratégies de lexicalisation des éléments de la trajectoire. L'anglais utilise généralement des verbes de manière associés à des particules (satellites : *in, out, up, down, etc.*) pour l'expression des éléments du trajet, alors que l'arabe utilise des verbes qui, en plus de la notion de mouvement, expriment aussi certains éléments de la trajectoire.

FR	Vers le milieu du pont, Cosette, ayant les pieds engourdis, désira marcher Self_motion . [Il <i>Agent</i>] [la <i>Theme</i>] [posa Placing] [à terre <i>Goal</i>] et la reprit par la main.
EN	Towards the middle of the Bridge, Cosette, whose feet were benumbed, wanted to walk Self_motion . [He <i>Agent</i>] [set Placing] [her <i>Theme</i>] [on the ground <i>Goal</i>] and took her hand again.
AR	و حوالي منتصف الجسر رَغبت كوزيت، و قد خدرت رجلاها، في أن [تسير Self_motion] [فأنزلها] [Cause_motion] إلى الأرض [Goal] و أمسك بيدها. Trad. : <i>Et environ la moitié du pont, Cosette, ces deux pieds s'engourdirent, désira de marcher et donc il la fit descendre à la terre.</i>
Glose	w HawAly munotaSafi AljisorI ragibato kuwziyto wa qado xadirato rijolAhA fiy >no tasiyra fa>anozalahaA <IY Al>aroDi w >amosaka biyadihaA

Tableau 9. *Conceptualisation en Frames (Les Misérables) (EN/AR/FR)*

Une autre divergence dans les points de vue en termes de conceptualisation est présentée dans le tableau 9 représentant un exemple à partir du célèbre roman français *Les Misérables* (Hugo, 1845-1862). Le texte original (français) et sa traduction en anglais utilisent la frame 'Placing' (un 'Agent' place un 'Theme' dans une location 'Goal') pour caractériser la scène où *Jean Valjean* essaye de poser par terre *Cosette* qui est dans ses bras. Cependant, dans la version arabe, le traducteur a préféré utiliser le verbe أنزل (faire descendre) évoquant, ainsi la frame 'Cause_motion' (un 'Agent' cause à

un ‘Theme’ de subir un mouvement direct). Selon la figure 2 (section 2), la conceptualisation arabe de l’événement n’est qu’un point de vue général de la scène caractérisée par ‘Placing’ qui est une sous-frame de ‘Cause_motion’. Un point fort de FrameNet et que, contrairement aux autres bases lexicales, elle offre des mécanismes (relations frame à frame) permettant de relier toutes les UL qui peuvent concevoir la même scène. En dépit de ces divergences de conceptualisation en termes de frames, la théorie de Fillmore dispose de mécanismes d’interprétation et de moyens de prise en charge des effets qui résultent des différences dans les stratégies de lexicalisation des événements de mouvement et de déplacement.

5. Conclusion

Dans cet article nous avons présenté les premiers résultats de notre tentative d’application de la théorie des frames sémantiques pour l’analyse de l’arabe, langue qui diffère de l’anglais typologiquement. Nous avons montré qu’une intégration de plusieurs ressources linguistique peut améliorer la qualité des systèmes TAL (Baker et Felbaum, 2009) (section 3.3). Nous avons aussi montré qu’il était possible d’appliquer les procédures d’analyse et d’annotation du projet BFN sur l’arabe, avec quelques adaptations nécessaires afin de respecter les caractéristiques linguistiques propres à l’arabe. Nous avons effectué une étude contrastive à base des frames sémantiques sur la conceptualisation des événements de mouvement et de déplacement. Nous avons aussi montré qu’une approche par l’analyse sémantique peut apporter des réponses aux questions relatives à la divergence des langues dans l’expression des événements de mouvement.

Nous croyons que notre analyse contrastive offre un apport aux études sur la typologie des langues initiées par Talmy et Slobin et représente les premiers résultats pour l’arabe fondés sur un objet d’étude de référence (*The Hobbit*) dans les analyses cross-langues. Sur cette base, nous avons confirmé les conclusions des chercheurs (Fillmore, Ohara, Ellsworth *et al.*, Petruck) sur le rôle important de la structure des frames et des relations frame à frame dans l’interprétation des divergences en conceptualisation entre les différentes langues. Cependant, afin d’arriver à une généralisation de ces conclusions, une étude approfondie doit être établie sur un large corpus représentatif. Notre travail peut être étendu sur d’autres phénomènes langagiers afin de répondre aux questions liées à l’applicabilité de la théorie des frames sémantiques sur l’arabe.

6. Bibliographie

- Al-Sulaiti L., Atwell E., « The design of a corpus of contemporary Arabic », *International Journal of Corpus Linguistics*, vol. 11, p. 135-171, 2006.
- Alshehri A. M., « The frame semantics of 'selfmotion' frame in Arabic and English », Thèse de doctorat, San Francisco : San Francisco State University, 2014.
- Baker, C.F., Fillmore, C.J., Lowe, J.B., « The Berkeley FrameNet Project », *In COLING-ACL '98 : Proceedings of the Conférence*, p. 86-90, 1998.
- Baker C., « La sémantique des cadres et le projet FRAMENET : une approche différente de la notion de « valence » », *Langages*, vol. 4, n° 176, p. 32-49, 2009.
- Baker C., Fellbaum C., « WordNet and FrameNet as complementary resources for annotation », *In : Association for Computational Linguistics* (ed.), *In Proceedings of the Third Linguistic Annotation Workshop*, Singapore, p.125-129, 2009.
- Baker C., Fellbaum C., Passonneau R.J., « Semantic Annotation of MASC », *Handbook of Linguistic Annotation*, Springer, Dordrecht, p. 699-717, 2017.
- Bernini, G., « Word classes and the coding of spatial relations in motion events : A contrastive typological approach », *In Space in Language*, p. 29-52, Edizioni ETS, 2010.
- Bick E., « A FrameNet for Danish », *In : Proceedings of Nodalida*, Riga, Latvia, NEALT Proceedings Series, *Riga, Latvia*, vol. 11, p.34-41, 2011.
- Black W., Elkateb S., Rodriguez H., Alkhalifa M., Vossen P., Pease A., Fellbaum C.. « Introducing the Arabic wordnet project », *In Proceedings of the third international WordNet conference*, p. 295-300, Seogwipo, Korea, 2006.
- Boas H. C., « Bilingual FrameNet Dictionaries for Machine Translation », *In LREC*, 2002, p. 1364-1371, Las Palmas, Iles Canaries, 2002.
- Boas H. C., « Semantic frames as interlingual representations for multilingual databases », *International Journal of Lexicography*, vol. 18, n° 4, p. 445-478, 2005.
- Boas H. C., « Recent trends in multilingual computational lexicography », *Multilingual FrameNets in Computational Lexicography : Methods and Applications. Berlin and New York : Mouton de Gruyter*, p. 1-36, 2009.
- Buckwalter T., « Buckwalter Arabic morphological analyzer version 1.0 », linguistic data consortium, *University of Pennsylvania, LDC Catalog n° : LDC2002L49*, 2002.
- Burchardt A., Erk K., Frank A., Kowalski A. Pado S., « SALTO : A versatile multi-level annotation tool », *In Proceedings of LREC-*, Genoa, Italy, 2006.
- Burchardt A., Reiter N., Thater S., Frank A., « A semantic approach to textual entailment : system evaluation and task analysis ». *In : Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, Prague, République Tchèque, 2007.
- Burchardt A., Erk K., Frank A., Kowalski A., Padó S., Pinkal M., « Using FrameNet for the semantic analysis of German : annotation, representation, and automation », *Multilingual FrameNets in Computational Lexicography : methods and applications*, p. 209-244, 2009.
- Burchardt A., Pennacchiotti M., « FATE : Annotating a Textual Entailment Corpus with FrameNet », *In Handbook of Linguistic Annotation*, p. 1101-1118, Springer, 2017.
- Candito M., Amsili P., Barque L., Benamara F., De Chalendar G., Djemaa M., Haas P., Huy-ghe R., Mathieu Y. Y., Muller P., Sagot B., Vieu L., « Developing a French FrameNet : Methodology and First results », *Proceedings of the International Conference on Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland, 2014.

- Coyne O. Rambow Hirschberg J., Sproat R., « Frame Semantics in Text-to-Scene Generation », *In Proceedings of the KES'10 workshop on 3D Visualisation of Natural Language*, Cardiff, Wales, 2010.
- Diab M., Mansouri A., Palmer M., Babko-Malaya O., Zaghouni W., Bies A., Maamouri M., « A Pilot Arabic Propbank », *In Proceedings of the 7th International Conference on Language Resources and Evaluation*, 2008.
- Djemaa M., Candito M., Muller P. Vieu L., May., « Corpus annotation within the french framenet : a domain-by-domain methodology », *In Tenth International Conference on Language Resources and Evaluation (LREC)*, Portorož, Slovenia, 2016.
- Dolbey, A., Ellsworth, M. and Scheffczyk, J., « BioFrameNet : A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies », *In KR-MED*, Vol. 222, 2006.
- Doyle, A. C. *The hound of the Baskervilles*. London : Strand Magazine, 1901. [Trad.arabe. M.H. Mahmoud. (1997). “نسيح باسكرفيل”.] [Trad. Franç. A. de Jassard. *Le Chien des Baskerville*, 1905. La Bibliothèque électronique du Québec]
- Dukes K., Buckwalter T., « A Dependency Treebank of the Quran using Traditional Arabic Grammar », *In Proceedings of the 7th international conference on Informatics and Systems*, Cairo, Égypte, 2010.
- Ellsworth M., Ohara K., Subirats K., Schmidt T., « Frame-semantic analysis of motion scenarios in English, German, Spanish, and Japanese », Paper presented at *The Fourth International Conference on Construction Grammar*, Tokyo, 2006.
- Erk K., Pado S., « Shalmaneser a toolchain for shallow semantic parsing », *In Proceedings of LREC*, vol. 6, 2006.
- Ferrández Ó., Ellsworth M., Muñoz R., Baker C.F., « Aligning FrameNet and WordNet based on semantic neighborhoods », *In Proceedings of the 7th international language resources and evaluation conference (LREC)*, p. 310-314, 2010.
- Fellbaum C., « WordNet : An Electronic Database ». MIT Press, Cambridge, MA. 1998.
- Fellbaum C., Osherson A., Clark P.E., « Putting semantics into WordNet's “morphosemantic” links », *In Proceedings of the 3th Language and Technology Conference, Poland*, 2007.
- Fillmore C.J., « The case for case », *In : Bach, E., Harms, R. (eds.), Universals in Linguistic Theory*, Holt, Rinehart & Winston, New York, 1968.
- Fillmore C.J., « Frame Semantics », *Linguistics in the Morning Calm*, p. 111-38, Seoul, 1982.
- Fillmore C.J., Johnson C.R., Petruck M.R.L., « Background to Framenet », *International Journal of Lexicography*, vol. 16, n° 3, p. 235-250, 2003.
- Fillmore C. J., Baker C., « A frames approach to semantic analysis », *In Heine, Bernd and Heiko Narrog (Eds.), The Oxford Handbook of Linguistic Analysis*, p. 313-341, Oxford University Press, 2010.
- Ghazzawi N., « Du terme prédicatif au cadre sémantique : méthodologie de compilation d'une ressource terminologique pour les termes arabes de l'informatique », *Thèse de doctorat*, mai 2016.
- Gildea D., Jurafsky D., « Automatic labeling of semantic roles », *Computational linguistics*, vol. 28, n° 3, p. 245-288, 2002.
- Habash N., Dorr B., « Handling translation divergences : Combining statistical and symbolic techniques in generation-heavy machine translation », *Machine Translation : From Research to Real Users*, p. 84-93, 2002.

- Habash N., Roth, R. M., « CATiB : The columbia arabic treebank », In *Proceedings of the ACL-IJCNLP, Conference Short Papers*, p. 221-224, 2009.
- Habash N. Y., « Introduction to Arabic natural language processing », *Synthesis Lectures on Human Language Technologies*, vol. 3 n° 1, p. 1-187, 2010.
- Hajic J., Smrz O., Zemánek P., Šnaidauf J., Beška E., « Prague Arabic dependency treebank : Development in data and tools ». In *Proceeding of the NEMLAR Intern. Conf. on Arabic Language Resources and Tools*, Cairo, Egypt, p. 110-117, 2004.
- Hartmann S., Kuznetsov I., Martin T., Gurevych I., « Out-of-domain FrameNet Semantic Role Labeling », In *Proceedings of the 15th Conference of the European Chapter of the ACL : vol. 1, Long Papers*, p. 471-482, 2017.
- Hayoun A., Elhadad M., « The Hebrew FrameNet Project », In *Proceedings of the International Conference on Language Resources and Evaluation Conference (LREC)*, Portorož Slovenia, 2016.
- Hugo, V., *Les Misérables*.1845-1862 [1- Trad. arabe. M. Albalabki. (1979), « المؤسسة ». Dar El Ilm Lilmalayin, Liban] [2- Trad. anglaise. I. F. Hapgood. *Les Misérables*,1887, Thomas Y. Cowell :New York, USA]
- Johansson R., Nugues P., « The effect of syntactic representation on semantic role labeling », In *Proceedings of International Conference on Computational Linguistics (COLING)*, Manchester, UK, 2008.
- Jeonguk k., Hahm y., Choi K.S., « Korean FrameNet Expansion Based on Projection of Japanese FrameNet », In *Proceeding COLING (Demos)*, Osaka, Japan, 2016.
- Kipper-Schuler K., « VerbNet : A broad-coverage, comprehensive verb lexicon », PhD thesis, University of Pennsylvania, 2005.
- Lakhfif A., Laskri M. T., Atwell E., « Multi-Level Analysis and Annotation of Arabic Corpora for Text-to-Sign Language MT », In *Second Workshop on Arabic Corpus Linguistics (WACL-2)*, Lancaster University, UK, 2013.
- Lakhfif A., Laskri M. T., « A frame-based approach for capturing semantics from Arabic text for text-to-sign language », *International Journal of Speech Technology*, doi : 10.1007/s10772-015-9290-8, 2015.
- Lakhfif A., « Un signeur virtuel 3D pour la traduction automatique de textes arabes vers la langue des signes algérienne », thèse de doctorat en sciences, avril 2016.
- L'Homme M.C., *Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés*, Traduire, vol. 217, p. 78-103, 2008.
- Johnson M., Lenci A., « Verbs of visual perception in Italian FrameNet », *Constructions and Frames*, vol. 3, n° 1, p. 9-45, 2011.
- Lindén K., Haltia H., Luukkonen J., Laine A. O. Roivainen H., Väisänen N., « FinnFN 1.0 :The Finnish frame semantic database », *Nordic Journal of Linguistics*, vol. 40, n° 3, p. 287-311, 2017.
- Maamouri M., Bies A., Buckwalter T., Mekki W., « The penn arabic treebank : Building a large-scale annotated arabic corpus », In *NEMLAR conference on Arabic language resources and tools*, p. 102-109, 2004.
- Marcus M. P., Santorini B., Marcinkiewicz M. A., « Building a large annotated corpus of English : The Penn TreeBank », *Computational Linguistics*, vol. 19, n° 2, p. 313, 1994.
- Minsky M., « A framework for representing knowledge », In *Proceeding Winston (Ed.), The Psychology of Computer Vision*, McGraw-Hill, 1975.

- Mel'čuk I., « *Dependency Syntax : Theory and Practice* », State University of NY Press, Albany, 1988.
- Moschitti, A., « Kernel methods, syntax and semantics for relational text categorization », In *Proceedings of the 17th ACM conference on Information and knowledge management*, p. 253-262, ACM, 2008.
- Narayanan S., Harabagiu S., « Question answering based on semantic structures », In *Proceedings of the 20th international conference on Computational Linguistics (ACL)*, p. 693-, 2004.
- Nuzzolese A. G., Gangemi A., Presutti V., « Gathering lexical linked data and knowledge patterns from FrameNet », In *Proceedings of the sixth international conference on Knowledge capture*, p. 41-48, ACM, 2011.
- Ohara K. H., Fujii S., Ohori T., Suzuki R., Saito H., Ishizaki S., « The japanese framenet project : An introduction », In *Proceedings of LREC-04, Satellite Workshop « Building Lexical Resources from Semantically Annotated Corpora* », p. 9-11, 2004.
- Ohara, K.H., « Frame Semantics in Action : A Frame-based Contrastive Text Analysis Using FrameNet », In *the 10th International Cognitive Linguistics Conference*, Poland, 2007.
- Pado S., Pitel G., « Annotation précise du français en sémantique de rôles par projection cross-linguistique », *Proceedings of TALN-07*, Toulouse, France, 2007.
- Petruck M.R.L., « Frame semantics », In *Verschuereen, J., Östman, J-O. 2007, Blommaert, J. & Bulcaen, C., eds, Handbook of pragmatics*, Amsterdam : John Benjamins, p. 1-13, 1996.
- Reed S. K., Pease A., « A framework for constructing cognition ontologies using WordNet, FrameNet, and SUMO », *Cognitive Systems Research*, vol. 33, p.122-144, 2015.
- Ruppenhofer J., Ellsworth M., Petruck M. R., Johnson C. R., Scheffczyk J., « *FrameNet II : Extended theory and practice* », Institut für Deutsche Sprache, Bibliothek, 2016a.
- Ruppenhofer J., Michaelis L. A., « Frames, polarity and causation », *Corpora*, vol. 11, n° 2, p. 259-290, 2016b.
- Ryding KC., *A Reference Grammar of Modern Standard Arabic*, Cambridge University Press, 2005.
- Salomão, M.M.M., « FrameNet Brasil : um trabalho em progresso », *Calidoscópico*, vol. 7, n° 2, p. 171-182, 2009.
- Schank R. C., Abelson R. P., « *Scripts, Plans, Goals, and Understanding* », Hillsdale, N. J. : Erlbaum Associates, 1977.
- Sharaf A., Atwell E., « Knowledge Representation of the Quran Through Frame Semantics : A Corpus-Based Approach », In *Proceedings of the Fifth Corpus Linguistics Conference*. University of Liverpool, UK, 2009.
- Shen D., Lapata M., « Using Semantic Roles to Improve Question Answering », In *EMNLP-CoNLL*, p. 12-21, 2007.
- Shi L., Mihalcea R., « Putting pieces together : Combining FrameNet, VerbNet and WordNet for robust semantic parsing », In *Computational linguistics and intelligent text processing*, p. 100-111, Springer Berlin Heidelberg, 2005.
- Schmidt T., « The Kicktionary – a Multilingual Lexical Recourse of Football Language », [w :] Boas, HC (red.).*Multilingual FrameNets*, 2007.
- Slobin Dan I., « The many ways to search for a frog : Linguistic typology and the expression of motion events », In *Stromqvist & Verhoeven (eds.)*, p.219-257, 2004a.
- Slobin D. I., « Relating events in translation », In *Perspectives on language and language development : Essays in honor of Ruth A. Berman*. Dordrecht : Kluwer, p. 115-130, 2004b.

- Søgaard A., Plank B., Alonso H. M., « Using Frame Semantics for Knowledge Extraction from Twitter », *In AAAI*, p. 2447-2452, 2015.
- Subirats C., Petruck M.R.L., « Surprise : Spanish FrameNet », *In E. Hajicova, A. Kotesovcova & J. Mirovsky (eds.), Proceedings of CIL 17*. CD-ROM. Prague : Matfyzpress, 2003.
- Talmy L., « Toward a cognitive semantics », Cambridge, MA : MIT Press, 2000.
- Talmy L., « Path to realization : A typology of event conflation », *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, p. 480-519, 1991.
- Tolkien J. R. R., *The Hobbit or there and back again*, London : George Allen & Unwin, 1937. [1-Trad.arabe. H. Fahmy, M. Ghanim. (2008), « الهوية (أو ذهابا و عودة) ». Dar Lila-Boeken.] [Trad.française. F. Ledoux. *Bilbo le Hobbit*, 1980, Paris : Hachette.]
- Tesnière L., « Éléments de syntaxe structurale », Librairie C. Klincksieck, 1959.
- Torrent T., Salomão M. M., Campos F. A., Braga R. M., Matos E. E., Gamonal M., Gonçalves J., Souza B. C., Gomes D., Peron S., « Copa FrameNet Brasil : a frame -based trilingual electronic dictionary for the Football World Cup », *In COLING (Demos)*, 10-14, 2014.
- Venturi G., Lenci A., Montemagni S., Vecchi E. M., Sagri M. T., Tiscornia D., Agnoloni T., « Towards a FrameNet resource for the legal domain », *LOAIT*, p. 67-76, 2009.
- Vossen P. « Introduction to eurowordnet », *In EuroWordNet : A multilingual database with lexical semantic networks*, Springer Netherlands, p. 1-17, 1998.
- Vossen P. (ed) « EuroWordNet », (LE2-4003, LE4-8328), Part A, Final Document Deliverable D032D033/2D014, 1999.