

Les modèles de description du verbe dans les travaux de Linguistique, Terminologie et TAL

Ornella Wandji Tchami

CNRS UMR 8163 STL, Université Lille 3, 59653 Villeneuve d'Ascq, France

ornella.wandjitchami@univ-lille3.fr

Résumé. Dans le cadre de notre projet de recherche, qui a pour but l'implémentation d'un outil de simplification des emplois spécialisés de verbes dans des corpus médicaux à partir de l'analyse syntaxico-sémantique de ces verbes en contexte, nous proposons une analyse de quelques approches et travaux qui ont pour objet principal la description du verbe dans les trois domaines de recherche à l'interface desquels se situe notre projet : linguistique, TAL et terminologie. Nous décrivons plus particulièrement les travaux qui peuvent avoir une incidence sur notre étude. Cet état de l'art nous permet de mieux connaître le cadre théorique dans lequel s'intègre notre projet de recherche et d'avoir les repères et références susceptibles de contribuer à sa réalisation.

Abstract. As part of our research project, which aims to implement a text simplification tool for the specialized usages of verbs in medical corpora using the syntactic and semantic analysis of these verbs in context, we propose an overview of some approaches and work whose main research object is the description of verbs, within the three research areas which interface our study is : linguistics, terminology and NLP. We pay a particular attention to studies that can have an impact on our work. This state of the art allows us to better understand the theoretical framework related to our research project. Moreover, it allows us to have benchmarks and references that might be useful for the realization of our project.

Mots-clés : Verbe terminologique ou spécialisé, sémantique des cadres, sémantique lexicale, structure argumentale, étiquetage en rôles sémantiques.

Keywords: Specialized verb, Frame Semantics, lexical semantics, argumental structure, Semantic Role Labeling.

1 Introduction

Contexte général

L'intérêt porté au verbe change selon que l'on se situe dans le domaine de la linguistique, de la terminologie ou celui du Traitement Automatique des Langues (TAL). En effet, selon leurs objectifs respectifs, chacune de ces disciplines octroie au verbe une place différente reconnaissable à travers l'importance qui lui est donnée dans les diverses études et les différents cadres théoriques propres au domaine concerné. Quelles sont les éclairages proposés par les différentes approches (linguistique, terminologie et TAL) qui prennent le verbe comme objet d'étude ? Comment est-ce que le verbe est abordé dans ces travaux ? Est-il traité au même titre que les autres catégories grammaticales en l'occurrence le nom ? En quoi est-ce que le verbe et sa structure argumentale peuvent-ils être utiles en vue de la simplification des textes spécialisés ? Telles sont les questions auxquelles nous allons essayer de répondre dans ce travail qui a pour objectif de dresser un état de l'art de différents modèles de descriptions du verbe dans les trois disciplines concernées.

Travail envisagé

Le projet que nous entreprenons a pour objectif de proposer une méthode de simplification de textes médicaux, à partir d'une analyse syntaxico-sémantique des verbes en contexte. Au terme de ce travail, nous souhaitons implémenter un outil de simplification des textes écrits en français (et éventuellement en anglais), spécialisés en cardiologie (ou en d'autres domaines médicaux). L'outil devra repérer les emplois verbaux peu communs au discours des patients et devra ensuite proposer des emplois sémantiquement similaires, mais plus adaptés au niveau de spécialisation de ces utilisateurs. La méthode proposée est basée sur différentes hypothèses. En effet, nous pensons que le prédicat verbal peut être un excellent

point de départ pour cerner la sémantique des textes spécialisés puisqu'il sert à exprimer l'expertise portée par les mots qui l'entourent dans la phrase (L'Homme & Bodson, 1997). Par conséquent, nous considérons la structure argumentale du verbe comme une importante source d'informations sur les propriétés sémantique et syntaxique du verbe.

Ce projet de recherche s'inscrit dans le cadre de la simplification des textes spécialisés. Il s'agit d'une tâche du TAL qui consiste à cibler et à simplifier automatiquement les éléments, qui empêchent la compréhension aisée d'un texte, afin de faciliter l'accès au contenu de ce texte. Les travaux existants se focalisent sur la simplification syntaxique (Brouwers *et al.*, 2014), la simplification lexicale (Elhadad, 2006; Leroy *et al.*, 2012), la combinaison des fonctions lexicales, grammaticales, syntaxiques et discursives (Heilman *et al.*, 2007, 2008; Pitler & Nenkova, 2008), les caractéristiques de surface des textes (nombre de caractères et syllabes par mot), la capitalisation, la ponctuation et les ellipses (Tapas & Orr, 2009), ou la modélisation statistique de la langue (Thompson & Callan, 2004). L'approche que nous proposons se situe à mi-chemin entre la simplification lexicale et la simplification syntaxique et vise à réduire les difficultés de compréhension des textes médicaux fortement spécialisés à travers la simplification des constructions verbales. À notre connaissance, il n'existe pas de travaux en simplification de textes autant orientés sur l'analyse des verbes et de leurs argumentales. Une étude comparative du fonctionnement des verbes dans des textes de corpus médicaux rédigés par des experts et des non-experts en médecine a permis d'observer que les verbes ont tendance à s'entourer d'arguments fortement spécialisés dans les écrits des experts, rendant parfois leur compréhension difficile pour les non-experts (Wandji Tchami *et al.*, 2013). Notre travail de recherche vient donner une suite à cette observation. L'objectif principal étant d'améliorer certains aspects de la méthode (l'annotation automatique des arguments, l'analyse des verbes) et de la développer davantage, en y intégrant un travail de simplification.

Le travail présenté ici a pour objectif de nous aider à mieux cerner le cadre théorique dans lequel s'inscrit le projet de recherche que nous envisageons de réaliser, en nous donnant une idée précise des travaux et outils existants, centrés sur le verbe et susceptibles de nous aider pour réalisation du travail envisagé. Il est organisé autour de 4 grandes parties dont les trois premières sont consacrées à l'exploration des travaux portant sur le prédicat verbal, respectivement en linguistique (section 2), terminologie (section 3) et TAL (section 4). Dans la dernière partie (section 5), nous faisons une discussion de l'impact que les travaux de l'état de l'art peuvent avoir sur la réalisation de notre projet de recherche et nous abordons les perspectives de travail que nous envisageons d'explorer.

2 Les approches linguistiques dédiées au verbe

En linguistique, de nombreux cadres théoriques placent le verbe au coeur de leurs travaux. Nous nous attardons plus particulièrement sur les cadres théoriques qui s'intéressent au verbe en tant qu'élément régisseur (c'est-à-dire un élément dont la réalisation syntaxique et sémantique dépend grandement de la présence d'autres constituants qui lui sont subordonnés) et décrivent son rapport avec les autres constituants de la phrase. Il existe diverses approches de description du verbe, mais nous ne sommes pas en mesure de fournir une présentation exhaustive de toutes les approches théoriques existantes. Nous nous limitons à celles qui servent de bases à la réalisation de différentes tâches du TAL (sections 4.2 et 4.1), et à la conception des ressources (section 4.3).

2.1 Le nœud verbal au coeur de la syntaxe structurale

La syntaxe structurale (Tesnière, 1959) est la première théorie à avoir mis le verbe au centre de la phrase. En syntaxe structurale, l'ensemble des mots d'une phrase constitue une véritable hiérarchie au sein de laquelle les constituants sont liés les uns aux autres par des liens de dépendance. La phrase, encore appelée *stemma*, est décrite comme étant un schéma arborescent, ou un ensemble de nœuds. Le nœud quant à lui désigne un ensemble constitué d'un régissant et de tous ses subordonnés. Dans cette configuration, le nœud central correspond en général au nœud verbal. Le verbe, étant au centre du nœud verbal, est par conséquent au coeur de la phrase. Il est pour ainsi dire le régissant de toute la phrase. La notion de nœud verbal est définie à travers une métaphore du drame : « *le nœuds verbale ... exprime un tout petit drame. Comme un drame, ... il comporte obligatoirement un procès et plus souvent des acteurs et des circonstants* ». C'est dans cette optique que cette approche postule l'existence des actants ou participants au procès verbal (Tesnière, 1959). L'ensemble des actants d'un verbe constitue sa structure actancielle. Un verbe peut avoir zéro, un ou plusieurs actants, comme le montre l'exemple suivant : *Alfred donne le livre à Charles*. Dans cette phrase, le verbe *donner* a trois actants : *Alfred*, *le livre* et *Charles*. Chaque actant joue un rôle bien déterminé dans le procès verbal.

2.2 La théorie des cadres sémantiques

Encore appelée *Frame semantics*, la sémantique des cadres est une approche qui remonte aux années 1980. Elle est une extension de la grammaire des Cas (Fillmore, 1968), qui évoquait déjà l'existence des rôles sémantiques (agent, lieu, etc.) dans la structure syntaxique profonde du verbe. La sémantique des cadres (Fillmore, 1982) vise à l'origine à faciliter la compréhension des textes. Son principal objectif est de décrire la syntaxe et la sémantique des unités lexicales (noms, adjectifs, verbes). L'idée principale de Fillmore est que le sens d'un mot ne peut être interprété que si l'on a accès aux informations (linguistiques, extralinguistiques ou encyclopédiques) essentielles faisant référence à ce mot. Ces informations peuvent être accessibles grâce à un frame ou cadre au sein duquel les unités lexicales sont organisées. Le cadre est défini comme un scénario, un schéma ou une structure conceptuelle qui sous-tend l'utilisation d'un item lexical ainsi que son interprétation (Fontenelle, 2009). Il décrit une situation particulière ainsi que les participants *Frame elements* (FE) qui peuvent être obligatoires (*core elements*) ou facultatifs (*non core elements*). Un cadre est évoqué par une unité lexicale (LU). Par exemple, le frame de la transaction commerciale (Fillmore, 1976) a plusieurs unités évocatrices : *acheter, vendre, payer, récupérer* et plusieurs participants : obligatoires (VENDEUR, ARGENT, BIEN, ACHETEUR) et facultatifs (MOYEN), etc. Lorsque l'unité évocatrice du cadre est un verbe, l'analyse est focalisée sur les arguments de ce dernier qui représentent les éléments du cadre.

2.3 Les classifications de verbes

2.3.1 La classification des verbes (anglais) selon Levin

Beth Levin (Levin, 1993) propose une classification lexico-sémantique de verbes anglais à partir d'une analyse de leur fonctionnement (syntaxe, classe sémantique des arguments sélectionnés, etc.). Les verbes qui affichent un ensemble d'alternances (de diathèses ou frames) identiques ou similaires dans la réalisation de leurs structures argumentales sont supposés partager certains éléments de sens et, de ce fait, sont regroupés dans une classe sémantiquement homogène. L'alternance de diathèses (la relation entre deux réalisations de surface d'un même prédicat), qui est le principal critère d'identification des classes verbales dans cette approche, est appuyée par des propriétés supplémentaires liées à la sous-catégorisation, à la morphologie et aux verbes ayant un sémantisme complexe. À partir de ces critères, la classification couvre 3 024 verbes, 4 186 sens, 240 classes de verbes construites autour de 79 alternances. Par exemple, la classe des prédicats dénotant une configuration spatiale contient les verbes suivants : *balance, bend, bow, crouch, dangle, flop, fly, hang, hover, jut, kneel, lean, lie, loll, loom, lounge, nestle, open, perch, plop, project, protude, recline, rest, rise, roost, sag, sit, slope, slouch, slump, sprawl, squat, stand, stoop, straddle, swing, tilt, tower* (Levin, 1993). Une extension substantielle de cette classification intègre 57 nouvelles classes pour les verbes qui n'ont pas été couverts initialement (Korhonen & Briscoe, 2004). Parmi les nouvelles classes, *FORCE class* regroupe les verbes tels que *manipulate, pressure, force*.

2.3.2 Les classes d'objets de Gaston Gross

Une classe d'objets est un « *ensemble de substantifs, sémantiquement homogènes, qui détermine une rupture d'interprétation d'un prédicat donné, en délimitant un emploi spécifique* » (Gross, 2008). En d'autres termes, les classes d'objets déterminent l'interprétation donnée d'un prédicat parmi d'autres possibles. Elles sont induites par les prédicats (verbes et adjectifs) et permettent d'identifier en contexte les mots avec lesquels ils entretiennent une relation conceptuelle telle que la synonymie, l'antonymie, etc. Ces entités sont construites sur des bases syntaxiques et concernent particulièrement les compléments qui apportent beaucoup plus d'informations que le sujet dans l'interprétation d'un prédicat (Gross, 2012). Par exemple, la phrase *vous suivez* n'est pas assez précise, ce qui rend son interprétation difficile. Par contre, si l'on y ajoute un complément, l'interprétation sera plus aisée et la signification du verbe sera plus transparente. Ainsi, dans la phrase *vous suivez ce chemin*, l'objet *chemin* peut être remplacé par un autre substantif comme *route, rue, voie, sentier* et le verbe garde le même sens. Ces substantifs peuvent donc être considérés comme appartenant à une même classe d'objets, celle de <voies>. Par contre, si on remplace *chemin* par le mot *cours*, on est face à un autre emploi du verbe car *cours* appartient à une autre classe d'objets, appelée <enseignements>. Elle contient les mots comme *séminaire, stage, formation, cycle études, etc.* Le principal intérêt des classes d'objets est de rendre compte des différents emplois des prédicats, en déterminant leurs schémas d'arguments et en rattachant à ceux-ci un ensemble de propriétés qui les caractérisent (Gross, 2008).

2.3.3 Lexique-Grammaire des verbes français

Le lexique-grammaire des verbes du français (Gross, 1975) est un dictionnaire syntaxique électronique téléchargeable¹. Il est organisé en plusieurs tables, chacune regroupant les verbes du lexique qui ont un fonctionnement comparable : constructions types, distribution des actants, sémantique, etc. Chaque table comprend un ensemble de propriétés, et un codage qui précise si l'élément a ou non cette propriété. Chaque entrée d'une table contient les informations suivantes : l'élément vedette, une construction type dans laquelle il peut apparaître, et des constructions associées à cette construction type. Les différents emplois des verbes, énumérés dans les tables, sont décrits grâce à des propriétés structurelles, distributionnelles et sémantiques. Par exemple, les tables de constructions sans compléments prépositionnels contiennent des constructions types, parmi lesquelles $N_0 V$ et $N_0 V N_1$. La construction $N_0 V$ accueille les verbes tels que *pleuvoir*, *bêtifier*, *bouillir*, *pisser* et selon le verbe, elle peut accepter un N_0 humain (*Luc bêtifie*), non humain (*l'eau bout*), impersonnel (*il pleut*), et le verbe peut être modifié par un adverbe (*ça ne pisse pas loin*) (Leclere, 1990).

3 Le verbe dans les travaux en terminologie

Les entités nominales ont longtemps occupé la place centrale dans les travaux sur les langues de spécialité au détriment des autres parties du discours, plus particulièrement des verbes, mis à l'écart pour diverses raisons. En effet, les travaux en terminologie se focalisent la plupart du temps sur la description des concepts, ou des entités nominales (particulièrement les noms) et la mise au jour des relations qu'elles partagent (genre-espèce, partie-tout, etc.). L'un des motifs principaux énoncés justifiant l'exclusion du verbe est la place accordée aux objets et à leurs dénominations dans l'approche de Wüster (Wüster, 1985). Cette situation trouve également une explication dans le fait que les entités nominales sont généralement utilisées pour le développement des terminologies, ontologies, thésaurus, glossaires, ou des vocabulaires. Ce constat s'explique également par les besoins croissants des applications : l'indexation et l'extraction d'informations sont des tâches typiquement basées sur les entités nominales. Pour ces raisons, la plupart des approches théoriques et méthodologiques sont adaptées aux entités nominales. Néanmoins, quelques travaux s'inspirant de la sémantique lexicale s'intéressent aux verbes et à leur mode de fonctionnement dans les domaines spécialisés. Ces travaux montrent que l'étude du verbe est quasi indispensable dans le cadre des activités comme l'extraction d'informations (Tateisi *et al.*, 2004), la conception des dictionnaires terminographiques (Tellier, 2008) ou encore la traduction spécialisée (Pimentel, 2011). Les structures argumentales des verbes peuvent également servir pour la détection automatique des relations sémantiques (Massimiliano *et al.*, 2008). Nous parlerons de deux approches d'analyse du verbe terminologique : l'approche conceptuelle (section 3.1) et l'approche lexico-sémantique (section 3.2).

3.1 L'approche conceptuelle

Le principe de l'approche conceptuelle stipule que l'on ne s'intéresse au verbe que s'il a l'aptitude de désigner un « concept d'activité », c'est-à-dire une activité (L'Homme, 2012). Telle est la condition qui détermine l'intégration des verbes dans des ressources terminologiques. Autrement dit, le verbe ne peut être considéré comme terme que s'il est fortement assimilable à un nom sur le plan conceptuel. Rey définit clairement le statut du verbe selon la perspective conceptuelle en ces termes : « *la terminologie ne s'intéresse aux signes (mots et unités plus grandes que le mot) qu'en tant qu'ils fonctionnent comme des noms dénotant des objets et comme des « indicateurs de notions » (de concepts) et dans cette optique, les verbes sont des noms de processus, d'actions* » (Rey, 1979). Cette conception justifie en partie la discrimination observée entre les parties du discours traitées dans un dictionnaire spécialisé. En général, on y compte très peu de verbes et d'adjectifs, mais beaucoup d'entités nominales. Les résultats d'une étude portant sur la présence des verbes dans les dictionnaires de spécialité évaluent à 2,44% (entre 0 et 4 verbes par dictionnaire) la moyenne d'apparition des verbes dans quatre dictionnaires terminologiques (L'Homme, 2003). L'approche conceptuelle a débouché de nos jours sur une démarche conceptuelle, incarnée par les ontologies, qui permet de distinguer les concepts d'activité, exprimés par les noms ou par les verbes, dans les domaines de spécialité. Ainsi, dans le domaine médical par exemple, les verbes tels que *traiter*, *observer* et *activer* peuvent devenir terminologiques puisqu'ils permettent de rendre compte des notions comme *traitement de la maladie*, *observation du patient* et *activation des cellules* (L'Homme, 2012).

1. <http://infolingu.univ-mlv.fr/DonneesLinguistiques/Lexiques-Grammaires/Telechargement.html>

3.2 L'approche lexico-sémantique

La sémantique lexicale est le cadre théorique qui a montré l'importance de la structure argumentale du verbe et du réseau lexical auquel le verbe appartient. Dans ce cadre, la caractérisation de la nature spécialisée du verbe est basée sur la description de sa structure argumentale ou son appartenance à un ou plusieurs réseaux lexicaux, (morpho-)sémantiques ou paradigmatiques. Ces tâches reposent sur l'observation et l'analyse des différentes occurrences du verbe en corpus.

3.2.1 La structure argumentale

L'analyse de la structure argumentale du verbe peut avoir pour but de démontrer sa nature terminologique. En effet, la nature prédicative du verbe fait qu'il a besoin des éléments qu'il régit pour la réalisation de son sens. Une étude propose de prendre en considération la nature des arguments du verbe qui détermine son degré de spécialisation (L'Homme, 1998). Ce raisonnement illustre l'hypothèse selon laquelle le verbe n'est pas spécialisé par lui-même, mais grâce à la prise en compte de sa structure argumentale (L'Homme, 2012). C'est ce critère qui permet d'admettre *installer* comme verbe spécialisé dans l'exemple suivant :

L'utilisateur installe la nouvelle version du traitement de texte sur son PC.

Dans cette phrase, les termes (*utilisateur, version, PC*) qui représentent les têtes des arguments du verbe appartiennent au domaine de l'informatique. Par conséquent, *installer* peut être considéré comme verbe terminologique dans ce domaine.

L'analyse des arguments des verbes constitue également un critère de poids chez (Tellier, 2008) qui y trouve un moyen de sélection des verbes, à partir d'un corpus spécialisé relevant du domaine de l'infectiologie, représentant de bons candidats termes à ajouter dans un dictionnaire spécialisé. Ce critère est également utilisé dans d'autres travaux (Lerat, 2002; Pimentel, 2011). Cependant, la caractérisation des arguments du prédicat verbal n'a pas pour unique but l'identification des verbes terminologiques. D'autres objectifs peuvent être poursuivis : l'extraction d'informations dans les corpus spécialisés du domaine de la biologie moléculaire (Tateisi *et al.*, 2004), l'élaboration d'un dictionnaire juridique portugais-anglais (Pimentel, 2011), l'analyse contrastive des corpus médicaux de niveaux de spécialisation différents (expert vs profane) (Wandji Tchami *et al.*, 2013).

3.2.2 Le réseau lexical

Outre la nature des actants, d'autres paramètres peuvent être pris en compte par les chercheurs lors du repérage des verbes terminologiques. L'un de ces paramètres, qui revient très souvent, est le lien qu'un verbe peut avoir avec un nom. Ainsi, si le nom est terminologique, et si le verbe est sémantiquement et le plus souvent morphologiquement apparenté à celui-ci, alors, il est fort possible que le verbe soit spécialisé lui aussi (L'Homme, 2012). Ce critère s'observe avec les couples tels que *développement - développer, téléchargement - télécharger, rechauffement - rechauffer*, le verbe et le nom correspondant désignent tous les deux une activité. Cependant, il existe des cas où le sens du verbe et celui du nom sont distincts malgré le lien morphologique qui existe entre eux. C'est le cas du couple *programme - programmer*, où le nom *programme* désigne le résultat de l'activité que dénote le verbe *programmer*. Comme nous pouvons le constater, dans l'approche lexico-sémantique, les noms peuvent servir de point de départ à partir duquel les verbes spécialisés sont identifiés en fonction des liens qu'ils partagent avec eux. C'est d'ailleurs cette méthode qui permet de retenir les verbes *évoluer, excréter, infecter* et *sécréter* comme termes du domaine de l'infectiologie, de part leur parenté aux noms *évolution, excrétion, infection* et *sécrétion* (Tellier, 2008). Comme ces noms sont fortement spécialisés dans ce domaine, les verbes correspondant héritent de cette caractéristique. Toutefois, il est possible de déplacer le point de départ de l'analyse vers le verbe. Cette technique peut permettre de découvrir d'autres unités reliées au verbe et d'élargir ainsi le réseau lexical construit autour de ce dernier (L'Homme, 2012).

Cette démarche a été appliquée lors de la conception du DicoInfo (Dictionnaire fondamental de l'informatique et de l'Internet), une base de données lexicales contenant des termes (les verbes y compris) fondamentaux du domaine de l'informatique et de l'internet (L'Homme, 2009). L'approche utilisée s'inspire grandement des principes théoriques et méthodologiques de la Lexicologie explicative et combinatoire (Mel'cuk *et al.*, 1995) et permet de fournir pour chaque entrée différents types d'informations : la réalisation linguistique des actants, les liens lexicaux, les synonymes, les contextes d'apparition du terme, etc. Pour le verbe *programmer* par exemple, DicoInfo propose divers types d'unités lexicales appartenant au réseau lexical notamment, *programmation (action de programmer), programme (résultat de l'action de programmer), informaticien (agent de l'action de programmer), langage (instrument utilisé pour programmer), logiciel*

(*résultat l'action de programmer*), *écrire* (synonyme de *programmer*), *développer* (synonyme de *programmer*), etc. Cet exemple permet d'observer que les mots repérés sont liés au verbe par différentes relations exprimées de façon implicite à travers de courtes gloses explicatives.

4 Le verbe en TAL

Le traitement des verbes dans le domaine du TAL s'appuie le plus souvent sur la caractérisation de leur structure argumentale : la valence verbale (Eynde & Mertens, 2003), les possibilités combinatoires et les relations de dépendances (Marneffe *et al.*, 2006), les fonctions grammaticales et rôles sémantiques des arguments (Gildea & Jurafsky, 2002), la désambiguïstation du sens des verbes (Ide & Véronis, 1998; Ye & Baldwin, 2006; Wagner *et al.*, 2009; Brown *et al.*, 2011), l'acquisition de schémas de sous-catégorisation à partir de l'analyse automatique de gros corpus (Messiant *et al.*, 2010), etc. Dans cette section, nous nous focalisons sur deux types de travaux : l'étiquetage des rôles sémantiques (section 4.1) et la désambiguïstation du sens des verbes (section 4.2). Par la suite, nous faisons la description de quelques ressources dédiées au verbe (section 4.3).

4.1 Étiquetage des rôles sémantiques

L'étiquetage des rôles sémantiques (Gildea & Jurafsky, 2002; Palmer *et al.*, 2005; Swier & Stevenson, 2004; Ye & Baldwin, 2006), ou *Semantic Role Labeling (SRL)*, est une tâche du TAL qui consiste à identifier de façon automatique les relations ou les rôles sémantiques (agent, patient, recipient, etc.) que jouent les constituants d'une phrase dans un cadre sémantique donné. Cette tâche est nécessaire pour la conception de différents types d'applications, et plus particulièrement celles qui touchent la compréhension et l'interprétation de la langue. Il s'agit par exemple de systèmes de questions-réponses (Miller *et al.*, 1996), d'extraction d'informations (Surdeanu *et al.*, 2003), de traduction automatique (Boas, 2002), ou de résumé automatique (Melli *et al.*, 2005). Les unités prédicatives (verbes, noms, adjectifs) occupent généralement le coeur des études qui concernent la SRL. En ce qui concerne le verbe, l'annotation consiste généralement à identifier dans la phrase les limites de ses arguments et éventuellement des circonstants, et ensuite de leur associer des rôles sémantiques selon le contexte. La démarche la plus utilisée pour la réalisation d'une SRL comprend trois étapes principales : (1) l'identification des arguments du verbe, basée le plus souvent sur des heuristiques (Xue & Palmer, 2004) qui permettent de réduire le nombre de candidats ; (2) le calcul des probabilités pour chacune des étiquettes à représenter les rôles sémantiques possibles ; (3) l'attribution de scores à chaque étiquette, éventuellement combinée à d'autres facteurs de prédiction, pour assigner des étiquettes appropriées aux arguments des verbes.

De nos jours, les modèles d'apprentissage statistiques sont très sollicités pour l'annotation des textes en rôles sémantiques. L'un des travaux de référence propose un système de SRL statistique, qui peut être utilisé aussi pour l'analyse syntaxique, l'étiquetage des parties du discours (Church, 1988), et la désambiguïstation du sens des mots (Lapata & Brew, 2004). Ce système, conçu pour les verbes, les noms et les adjectifs, atteint 82% de précision sur des phrases pré-annotées manuellement, tandis qu'il montre 65% de précision et 61% de rappel sur des phrases non annotés (Gildea & Jurafsky, 2002). Il a d'ailleurs été utilisée dans le cadre du projet FrameNet.

4.2 Désambiguïstation du sens des verbes

La désambiguïstation du sens des verbes ou *Verb Sense Disambiguation (VSD)* est une sous-tâche de la WSD (word sense disambiguation). Elle consiste à sélectionner automatiquement, parmi ses différents sens, le sens le plus approprié d'un verbe polysémique, selon son contexte d'apparition. Par exemple, le verbe *read* (lire) a plusieurs sens. Pour faire la distinction entre les phrases telles que *I read a book* (*je lis un livre*) et *I read you loud and clear* (*je te comprends parfaitement*), il est nécessaire de désambiguïser le contexte d'apparition du verbe, en suivant une des méthodes existantes. La désambiguïstation du sens est une tâche nécessaire pour la traduction automatique (Carpuat & Wu, 2007) ou l'extraction d'informations (Schütze & Pedersen, 1995; Sanderson, 1994). Deux types d'approches sont utilisées habituellement pour la désambiguïstation du sens des mots : approche à base de règles et approche à base d'apprentissage. L'approche à base de règles requiert des ressources comme les bases de données lexicales, les dictionnaires électroniques, qui fournissent des descriptions lexicales, syntaxiques et sémantiques des mots. À partir de ces ressources, des règles sont définies pour déterminer le sens exact du mot parmi l'ensemble des sens possibles. En traduction automatique, un ensemble constitué de 63 règles est proposé comme source de connaissances (Specia *et al.*, 2005). L'approche la plus utilisée actuellement est

basée sur l'apprentissage automatique (Ye & Baldwin, 2006; Brown *et al.*, 2011; Yarowsky, 1995). L'apprentissage peut être supervisé (exigeant un ensemble d'exemples manuellement annotés) ou non supervisé (appliqué sur des textes non annotés). En effet, certains chercheurs proposent une technique non supervisée de désambiguïsation du sens des verbes, qui regroupent les verbes ayant les préférences sélectionnelles et de sous-catégorisation similaires (Wagner *et al.*, 2009). Cette méthode montre 57.06% de précision. D'autres travaux identifient les préférences sémantiques (Lapata & Brew, 2004) et les marques de sous-catégorisation (Lapata & Brew, 1999) des verbes apparaissant dans plusieurs classes de Levin.

En ce qui concerne les méthodes supervisées, les premières expériences se focalisaient sur les bi-grammes et les fonctions linguistiques et contextuelles (Pedersen, 2000, 2001; Hoa Trang & Palmer, 2002). Par la suite, les chercheurs se sont intéressés à l'apport des bases de connaissances fournissant les informations telles que la catégorie grammaticale des mots voisins, la forme morphologique, les collocations, la relation syntaxique verbe-objet, utiles pour lever certaines ambiguïtés (Yoong Keok & Hwee Tou, 2002). De plus en plus, les chercheurs abordent les rôles sémantiques des arguments des verbes comme des fonctions contribuant à l'amélioration des performances des systèmes lors de la désambiguïsation du sens des verbes (Hoa Trang & Palmer, 2005; Ye & Baldwin, 2006). Cette technique est d'ailleurs recommandée car les rôles sémantiques associés à un mot peuvent donner des indices pour la déduction de son sens, surtout lorsque ces rôles sont associés à des frames de sous-catégorisation syntaxique (Gildea & Jurafsky, 2002). Certains travaux suivent une approche supervisée basée sur connaissances extraites des ressources lexicales externes (VerbNet, WordNet, etc.) (*knowledge based WSD*) (Brown *et al.*, 2011). Une autre approche, inspirée par les travaux en psycholinguistique, propose de nouveaux critères de regroupement des sens d'un mot, en fonction de la différence faible ou importante qui existe entre ces mots (Brown, 2008). Pour le français, il existe une approche d'analyse sémantique des textes, basée sur des réseaux lexicaux et les relations de dépendance entre les mots ambigus et les autres mots de la phrase (Mouton, 2010).

La réalisation de la WSD sur des textes spécialisés est actuellement une tâche relativement difficile selon les domaines, à cause de l'absence des ressources terminologiques nécessaires ou de l'insuffisance des données disponibles. Néanmoins, dans certains domaines comme l'informatique biomédicale, différentes études proposent des systèmes de WSD basés sur des méthodes non supervisées (Liu *et al.*, 2001) ou supervisées (Stevenson & Guo, 2010), utilisant des terminologies existantes.

4.3 Quelques ressources lexicales dédiées au verbe

Dans la suite de cette section, nous décrivons brièvement quelques ressources lexicales : FrameNet (section 4.3.1), VerbNet (section 4.3.2), VerbOcean (section 4.3.3) et WordNet (section 4.3.4). Nous nous intéressons particulièrement à la manière dont l'information sur le verbe est présentée.

4.3.1 FrameNet

FrameNet (Ruppenhofer *et al.*, 2006) est une base de données lexicales² initialement conçue pour l'anglais. Elle contient plus de 10 000 sens des unités lexicales décrits à travers plus de 1 000 cadres sémantiques liés hiérarchiquement les uns aux autres et illustrés par plus de 170 000 phrases. Le projet FrameNet propose une description des unités lexicales prédicatives (verbes, noms et adjectifs), basée sur l'annotation en cadres sémantiques (Fillmore, 1982) des phrases dans lesquelles ces unités apparaissent.

His **\$20** TRANSACTION **with Amazon.com** **for a new TV** had been very smooth.

Dans cette phrase, chaque couleur représente un élément du cadre : bleu foncé=ACHETEUR, bleu ciel=ARGENT, rouge=VENDEUR, vert=BIEN.

Ces frames mettent en évidence des informations sémantiques nécessaires pour capturer les sens de l'unité lexicale clé. Ainsi, pour chacune de ses entrées, FrameNet est capable de fournir un cadre sémantique complet, une description du frame, ses éventuelles relations avec d'autres frames, une description des éléments du frame et une illustration des schémas valenciels de l'entrée à l'aide d'exemples (Ruppenhofer *et al.*, 2006).

2. <https://framenet.icsi.berkeley.edu/fndrupal/about>

4.3.2 VerbNet

Contrairement à FrameNet, VerbNet³ (Kipper *et al.*, 2000; Kipper-Schuler, 2005) est totalement focalisé sur les verbes. Cette ressource lexicale propose une description des verbes basée sur la classification de Levin (section 2.3.1). Elle consiste à regrouper les verbes en différentes classes, qui mettent en évidence leurs propriétés syntaxiques et sémantiques communes. Cette méthode de description permet de faire des généralisations sur le comportement des verbes. Par exemple, les verbes appartenant à la classe *Hit 18.1* : *bang, bash, hit, kick...* sont des transitifs directs. Ils exigent un agent et un patient, et peuvent être modifiés par des prédicats sémantiques exprimant la manière, la cause, la direction, etc.

VerbNet est donc un lexique hiérarchique de verbes anglais regroupés en classes, indépendamment des domaines de spécialités auxquels ils peuvent appartenir. Chaque classe est décrite à travers : l'ensemble d'arguments possibles, présentés sous forme de rôles thématiques ; les éventuelles restrictions de sélection d'arguments (comme *animé, humain, organisation*) ; les cadres, décrivant les possibles réalisations de surface de la structure argumentale (constructions transitives, intransitives, syntagmes prépositionnels, résultatives) ; les alternances de diathèse, c'est-à-dire les variations des différents cadres. Selon le site officiel, après son extension (Korhonen & Briscoe, 2004), VerbNet compte 274 classes de premier niveau, 23 rôles thématiques, 94 prédicats sémantiques, 55 restrictions syntaxiques, 5 257 sens des verbes et 3 769 lemmes.

4.3.3 VerbOcean

VerbOcean (Chklovski & Pantel, 2004) est une ressource lexicale qui propose un réseau sémantique de relations entre les verbes, et recense uniquement des paires de verbes sémantiquement proches. Elle contient 22 306 relations entre 3 477 verbes et identifie 5 types de relations : *similitude* (la similitude), *strength* (la force), *antonymy* (l'antonymie), *enablement* (l'habilitation), et la relation temporelle *happens-before* (a lieu avant). L'approche appliquée pour la conception de cet outil est basée sur deux étapes : (1) la détection des paires de verbes qui apparaissent en co-occurrence fréquente, grâce à des requêtes effectuées sur le portail Google ; (2) pour chaque paire, le calcul du score de chaque relation possible, grâce à 35 schémas lexico-syntaxiques. Par exemple, les verbes *discover* (*découvrir*) et *refine* (*affiner, améliorer*) sont considérés comme une paire illustrant la relation *happens-before* si la chaîne *discovered and refined* (instantiant le schéma *Xed and then Yed*) est identifiée de façon très fréquente sur Google.

4.3.4 WordNet

WordNet (Fellbaum, 1998) est une base de données lexicale qui propose une description des verbes, mais également des noms et des adjectifs, sur la base de différentes relations sémantiques : la synonymie, l'antonymie, l'hyponymie, l'hyperonymie, la méronymie, la troponymie et l'implication (Miller, 1995). Contrairement à VerbOcean qui s'intéresse uniquement aux paires de verbes sémantiquement proches, WordNet traite plusieurs catégories d'unités prédicatives et ces unités sont regroupées dans des *synsets*, 117 000 au total. Un *synset* est un groupe de mots (synonymes) sémantiquement homogènes. Il contient des pointeurs qui marquent ses relations conceptuelles avec d'autres *synsets*. En outre, un *synset* contient une brève définition et, dans la plupart des cas, une ou plusieurs courtes phrases illustrant l'utilisation des membres de ce *synset*. Les formes des mots ayant plusieurs significations sont représentées par autant de *synsets* distincts.

5 Discussion et travaux futurs

Comme indiqué plus haut, le projet que nous entreprenons a pour objectif de proposer une méthode de simplification de textes médicaux écrits en français, à partir d'une analyse syntaxico-sémantique des verbes en contexte. Nous avons vu dans la section 3 que les travaux sur les langues de spécialité sont le plus souvent focalisés sur les entités nominales et, par conséquent, les travaux sur les verbes terminologiques sont peu nombreux. De même, dans la section 4, nous démontrons que peu de travaux en TAL appliquent la sémantique des cadres à des textes spécialisés et qu'il existe encore des cadres théoriques dans lesquels le verbe et sa structure argumentale sont peu considérés. En rupture avec ces constats, nous proposons d'exploiter l'étude de la structure argumentale des verbes pour la simplification des textes spécialisés. Nous partons de l'hypothèse selon laquelle le verbe, en tant que prédicat central dans la phrase, peut être le point de départ pour cerner la syntaxe et la sémantique des textes spécialisés puisqu'il sert à articuler l'expertise et les connaissances portées

3. <http://verbs.colorado.edu/mpalmer/projects/verbnet.html>

par les mots qui l'entourent dans la phrase (L'Homme & Bodson, 1997). Les travaux de l'état de l'art peuvent nous venir en aide à différentes étapes de la méthode, dont les principales sont :

1. *Annotation automatique des arguments des verbes selon leurs catégories sémantiques.* Comme la SRL, cette annotation vise à détecter automatiquement les verbes et leurs arguments, et à associer des étiquettes sémantiques à ces arguments. Toutefois, dans notre travail, les arguments seront associés non pas à des rôles sémantiques, mais à des types sémantiques proposés par une ressource terminologique existante (Côté, 1996). Pour ce faire, nous allons implémenter un programme qui prend en entrée le fichier résultant de l'analyse morpho-syntaxique, la terminologie et éventuellement d'autres ressources. Afin d'optimiser l'identification des arguments et de repérer plus de termes, nous prévoyons de définir des heuristiques basées sur la coordination, la relation hypéronyme/hyponyme, les têtes lexicales, etc. De même, pour augmenter la couverture de cette annotation, nous allons enrichir la terminologie avec des entrées supplémentaires.
2. *Analyse des verbes.* L'analyse des verbes est effectuée grâce aux informations obtenues lors de l'annotation sémantique des arguments. Il s'agit typiquement d'analyser leur nombre et leurs types sémantiques. Des méthodes d'apprentissage non supervisé peuvent être utilisées pour distinguer entre les emplois des verbes. Plus particulièrement, à partir des annotations, nous souhaitons savoir si ces emplois véhiculent les sens spécialisés ou non. Pour cette étape aussi, nous pouvons nous appuyer sur les ressources existantes, qui viendront compléter ou renforcer les annotations obtenues à l'étape précédente. Bien que l'apport de ces ressources est limité lorsqu'il s'agit de traiter des textes spécialisés, elles restent utiles, plus particulièrement pour l'analyse des emplois non spécialisés des verbes. Par exemple, les ressources comme VerbNet et FrameNet peuvent fournir des informations standard sur les schémas valenciels des verbes. Les ressources de type classes d'objets ou WordNet peuvent fournir les séries d'arguments qui sont sémantiquement proches entre eux. L'analyse des verbes peut permettre d'effectuer plusieurs types d'appréciation : la complétude des annotations, l'importance de certains arguments des verbes, la déviation des schémas argumentaux qui peuvent être révélateurs des emplois non standard et spécialisés des verbes. Des heuristiques dédiées seront nécessaires pour combiner les annotations avec les ressources, et pour départager ces différents cas de figure ;
3. *Simplification.* Dans le cas des textes spécialisés, rédigés pour un public de non experts, les emplois spécialisés des verbes peuvent être considérés comme des sources de difficulté. Grâce à l'étape précédente, de tels emplois spécialisés peuvent être détectés automatiquement. La simplification a pour objectif de rendre ces emplois de verbes plus abordables pour les utilisateurs non spécialistes. A ce niveau, l'absence de ressources du type WordNet pour les langues de spécialité représente une difficulté cruciale que nous allons devoir affronter. Dans un premier temps, pour pallier à ce problème, les phrases simplifiées seront conçues sur un modèle que proposent les définitions des termes du DicoInfo (L'Homme, 2009). Il s'agira de fournir une définition typique de la construction verbale ambiguë dans laquelle entre le verbe. Cette définition sera enrichie par un ou plusieurs synonymes du verbe qui seront recherchés dans WordNet ou d'autres ressources qui proposent les synonymes des mots de la langue générale. Une étude comparative des corpus de textes spécialisés, écrits par des experts et ceux écrits par des non-experts, effectuée au préalable, sera utile lors de la simplification, pour l'identification, si possible, des constructions verbales synonymes. La simplification concernera également les constituants syntaxiques de la phrase et éventuellement des temps verbaux (Brouwers *et al.*, 2014). En exploitant la méthode appliquée dans FrameNet et grâce aux observations en corpus, nous pouvons détecter les arguments nécessaires (*core*) et non nécessaires (*non core*) et alléger les phrases en supprimant les éléments non nécessaires. De la même manière, si les éléments nécessaires à la compréhension sont absents, nous pouvons les déduire et compléter ainsi la structure argumentale de verbes, en espérant que cela facilite la compréhension des phrases.

Au terme de ces différentes étapes, nous pensons pouvoir améliorer la lisibilité du texte et de rendre le sens des verbes plus accessibles aux utilisateurs non spécialistes en médecine. Les résultats de notre approche lexico-syntaxique seront évalués et comparés à ceux des méthodes de simplification focalisées uniquement sur les entités nominales, c'est-à-dire sur les arguments des verbes.

6 Conclusion

Tout au long de ce travail, nous avons exploré les principaux cadres théoriques et travaux qui s'intéressent particulièrement au prédicat verbal dans trois domaines de recherche : terminologie, où le verbe a tardé à s'imposer comme unité pouvant exprimer des connaissances spécialisées, face à la place dominante des entités nominales ; linguistique, où le verbe a toujours fait partie des catégories grammaticales les plus étudiées ; TAL, où de nos jours, de nombreuses ressources

et méthodes se consacrent partiellement ou entièrement aux verbes à travers l'étude de sa structure argumentale ou de ses relations sémantiques avec d'autres verbes. De façon générale, notre travail met en évidence le fait que la frontière entre ces trois disciplines n'est pas étanche, car les techniques et approches utilisées en linguistique sont réutilisées en terminologie et en TAL, et de la même façon, la linguistique et la terminologie contemporaines font très souvent recours aux ressources, outils et applications développées en TAL. Nous allons nous servir de cette interdisciplinarité pour mener à bien notre projet.

Références

- BOAS H. (2002). Bilingual fraMENET dictionaries for machine translation. In *LREC*, p. 1364–137, Las Palmas de Gran Canaria, Spain.
- BROUWERS L., DELPHINE B., ANNE-LAURE L. & THOMAS F. (2014). Syntactic sentence simplification for french. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)@EACL*, p. 47–56.
- BROWN S. (2008). Choosing sense distinctions for wsd : Psycholinguistic evidence. In *Proceedings of ACL/HLT*, p. 249–252, Columbus, OH.
- BROWN S., DLIGACH D. & PALMER M. (2011). Verbnets class assignment as a wsd task. In *9th International Conference on Computational Semantics*, Oxford, UK.
- CARPUAT M. & WU D. (2007). Improving statistical machine translation using word sense disambiguation. In *EMNLP-CoNLL*, p. 61–72.
- CHKLOVSKI T. & PANTEL P. (2004). Verbocean mining the web for fine-grained semantic verb relations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Barcelona, Spain.
- CHURCH K. (1988). A stochastic parts program and noun phrase parser for unrestricted text. In *Second Conference on Applied Natural Language Processing*, Austin, Texas.
- CÔTÉ R. (1996). *Répertoire d'anatomopathologie de la SNOMED internationale, v3.4*. Université de Sherbrooke, Sherbrooke, Québec.
- ELHADAD N. (2006). Comprehending technical texts : Predicting and defining unfamiliar terms. In *AMIA*, p. 239–243.
- EYNDE K. & MERTENS P. (2003). La valence : l'approche pronominale et son application au lexique verbal. *French Language Studies*, **13**(1), 63–104.
- FELLBAUM C. (1998). *WordNet : An Electronic Lexical Database*, In M. PRESS, Ed., *Language Speech and Communications*.
- FILLMORE C. (1968). *The case for case*, In UNIVERSALS, Ed., *Linguistic Theory*, p. 1–88.
- FILLMORE C. (1976). *Topics in lexical semantics*, In I. U. PRESS, Ed., *Current Issues in Linguistic Theory*, p. 76–138.
- FILLMORE C. (1982). *Frame Semantics*, In H. P. CO, Ed., *Linguistics in the morning calm*, p. 111–137.
- FONTENELLE T. (2009). sémantique des cadres et lexicographie. *Lexique*, (19), 162–177.
- GILDEA D. & JURAFSKY D. (2002). Automatic labeling of semantic roles. *Computational Linguistics*, **28**(3), 245–288.
- GROSS G. (2008). Les classes d'objets. *Lalies*, (28), 111–165.
- GROSS G. (2012). *Manuel d'analyse linguistique : approche sémantico-syntaxique du lexique*. Villeneuve-d'Ascq : Presses universitaires du Septentrion.
- GROSS M. (1975). *Méthodes en syntaxe*. Hermann.
- HEILMAN M., THOMPSON C., CALLAN J. & ESKENAZI M. (2007). Combining lexical and grammatical features to improve readability measures for first and second language texts. In *HLT-NAACL*, p. 460–467, Rochester, New York.
- HEILMAN M., THOMPSON C. & ESKENAZI M. (2008). An analysis of statistical models and features for reading difficulty prediction. In COLUMBUS, Ed., *Third Workshop on Innovative Use of NLP for Building Educational Applications*, p. 71–79, Ohio.
- HOA TRANG D. & PALMER M. (2002). Combining contextual features for word sense disambiguation. In A. FOR COMPUTATIONAL LINGUISTICS, Ed., *Proceedings of the ACL-02 Workshop on Word Sense Disambiguation : Recent Successes and Future Directions*, p. 88–94, Stroudsburg, PA, USA.

- HOA TRANG D. & PALMER M. (2005). The role of semantic roles in disambiguating verb senses. In A. FOR COMPUTATIONAL LINGUISTICS, Ed., *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, p. 42–49, Stroudsburg, PA, USA.
- IDE N. & VÉRONIS J. (1998). Introduction to the special issue on word sense disambiguation : The state of the art. *Computational Linguistic*, **24**(1), 2–40.
- KIPPER K., DANG H. & PALMER M. (2000). Class-based construction of a verb lexicon. In *The Seventh National Conference on Artificial Intelligence AAAI/IAAI*, p. 691–696.
- KIPPER-SCHULER K. (2005). *VerbNet : A broad-coverage comprehensive verb lexicon*. Thèse de doctorat, niversity of Pennsylvania, Philadelphia, PA.
- KORHONEN A. & BRISCOE T. (2004). Extended lexical-semantic classification of english verbs. In *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics*, Boston, MA.
- LAPATA M. & BREW C. (1999). Using subcategorization to resolve verb class ambiguity. In *JOINT SIGDAT CONFERENCE ON EMPIRICAL METHODS IN NLP AND VERY LARGE CORPORA*, p. 266–274.
- LAPATA M. & BREW C. (2004). Verb class disambiguation using informative priors. *COMPUTATIONAL LINGUISTICS*, p. 45–73.
- LECLERE C. (1990). Organisation du lexique-grammaire des verbes français. *Langue française*, (87), 112–122.
- LERAT P. (2002). Qu'est-ce que le verbe spécialisé ? le cas du droit. *Cahiers de Lexicologie*, **80**, 201–211.
- LEROY G., ENDICOTT J., MOURADI O., KAUCHAK D. & JUST M. (2012). Improving perceived and actual text difficulty for health information consumers using semi-automated methods. In *American Medical Infomatics Association*.
- LEVIN B. (1993). *English Verb Classes and Alternation, A Preliminary Investigation*. The University of Chicago : Press.
- L'HOMME M. (1998). Le statut du verbe en langue de spécialité et sa description lexicographique. *Cahiers de lexicologie*, **73**(2), 61–84.
- L'HOMME M. (2003). Capturing the lexical structure in special subject fields with verbs and verbal derivatives a model for specialized lexicography. *IJL*, **16**(4), 403–422.
- L'HOMME M. (2009). *Le DiCoInfo. Dictionnaire fondamental de l'informatique et de l'Internet*. Rapport interne, Observatoire de linguistique Sens-Texte (OLST).
- L'HOMME M. (2012). Le verbe terminologique un portrait de travaux récent. In *Congrès Mondial de Linguistique Française-CMLF*, p. 93–107.
- L'HOMME M. & BODSON C. (1997). Modèle de description des verbes specialises combinant base de connaissances et hypertexte. In *Congres international de terminologie*, p. 381–398, San Sebastian, Espagne.
- LIU H., LUSSIERB Y. & FRIEDMAN C. (2001). Disambiguating ambiguous biomedical terms in biomedical narrative text : An unsupervised method. *Journal of Biomedical Informatics*, **34**, 249–261.
- MARNEFFE M., MACCARTNEY B. & MANNING C. (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, p. 449–454.
- MASSIMILIANO C., ALDO G., ESTHER R., J S. & ISABEL R. (2008). Unsupervised learning of semantic relations for molecular biology ontologies. In *Proceeding of the 2008 conference on Ontology Learning and Population : Bridging the Gap between Text and Knowledge*, p. 91–104.
- MELLI G., WANG Y., LIU Y., KASHANI M., SHI Z., GU B., SARKAR A. & POPOWICH F. (2005). Description of squash the sfu question answering summary handler for the duc-2005 summarization task. In *HLT/EMNLP*.
- MEL'CUK I., CLAS A. & POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot / Aupelf-UREF.
- MESSIANT C., GÁBOR K. & POIBEAU T. (2010). Acquisition de connaissances lexicales à partir de corpus : la sous-catégorisation verbale en français. *Traitement Automatique des Langues*, **51**(1), 65–96.
- MILLER G. A. (1995). Wordnet : A lexical database for english. *Communication ACM*, **38**(11), 39–41.
- MILLER S., STALLARD D., BOBROW R. & SCHWARTZ R. (1996). A fully statistical approach to natural language interfaces. In *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*, p. 55–61, Stroudsburg, PA, USA.
- MOUTON C. (2010). *Ressources et méthodes semi-supervisées pour l'analyse sémantique de texte en français*. PhD thesis, Université Paris Sud-Paris XI.

- PALMER M., GILDEA D. & KINGSBURY P. (2005). The proposition bank an annotated corpus of semantic roles. *Computational Linguistics*, **31**(1), 71–105.
- PEDERSEN T. (2000). A simple approach to building ensembles of naive bayesian classifiers for word sense disambiguation. *CoRR*, **cs.CL/0005006**.
- PEDERSEN T. (2001). A decision tree of bigrams is an accurate predictor of word sense. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, p. 1–8, Stroudsburg, PA, USA.
- PIMENTEL J. (2011). Description de verbes juridiques au moyen de la sémantique des cadres. In *TOTH*.
- PITLER E. & NENKOVA A. (2008). Revisiting readability : A unified framework for predicting text quality. In *EMNLP*, p. 186–195, Waikiki, Honolulu, Hawaii.
- REY A. (1979). *La terminologie : noms et notions*, In P. UNIVERSITAIRES DE FRANCE, Ed., "Que sais-je ?".
- RUPPENHOFER J., ELLSWORTH M., PETRUCK M., JOHNSON C. & SCHEFFCZYK J. (2006). *FrameNet II Extended Theory and Practice*. Berkeley, California : International Computer Science Institute. Distributed with the FrameNet data.
- SANDERSON M. (1994). Word sense disambiguation and information retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 142–151, New York, NY, USA.
- SCHÜTZE H. & PEDERSEN J. (1995). Information retrieval based on word senses. In *Proceedings of the 4th Annual symposium on document analysis and information retrieval*, p. 161–175, Las Vegas.
- SPECIA L., DAS GRAÇAS V NUNES M. & STEVENSON M. (2005). Exploiting rules for word sense disambiguation in machine translation.
- STEVENSON M. & GUO Y. (2010). Disambiguation of ambiguous biomedical terms using examples generated from the umls metathesaurus. *Journal of Biomedical Informatics*, **43**, 762–773.
- SURDEANU M., HARABAGIU S., WILLIAMS J. & AARSETH P. (2003). Using predicate-argument structures for information extraction. In A. FOR COMPUTATIONAL LINGUISTICS, Ed., *Proceedings of the ACL*, p. 8–15, Sapporo, Japan.
- SWIER R. & STEVENSON S. (2004). Unsupervised semantic role labelling. In *EMNLP*.
- TAPAS K. & ORR D. (2009). Predicting the readability of short web summaries. In *WSDM*, p. 202–211, Barcelona, Spain.
- TATEISI Y., OHTA T. & TSUJII J. (2004). Annotation of predicate-argument structure on molecular biology text. In SPRINGER, Ed., *In Proceedings of the Workshop on the 1st International Joint Conference on Natural Language Process (IJCNLP)*, Hainan Island, China.
- TELLIER C. (2008). *Verbes spécialisés en corpus médicale : une méthode de description pour la rédaction d'articles terminologiques*. Thèse de doctorat, Université de Montréal.
- TESNIÈRE L. (1959). *Éléments de syntaxe structurale*. Paris : Klincksieck.
- THOMPSON C. & CALLAN P. (2004). A language modeling approach to predicting reading difficulty. In *HTL-NAACL*, p. 193–200.
- WAGNER W., SCHMID H. & SCHULTE IM WALDE S. (2009). Verb sense disambiguation using a predicate-argument-clustering model. In *In Proceedings of the CogSci Workshop on Distributional Semantics beyond Concrete Concepts*.
- WANDJI TCHAMI O., L'HOMME M. & GRABAR N. (2013). Discovering semantic frames for a contrastive study of verbs in medical corpora. In *Terminologie et intelligence artificielle (TIA)*, Villetaneuse.
- WÜSTER E. (1985). *Introduction to the General Theory of Terminology and Terminological Lexicography*.
- XUE N. & PALMER M. (2004). Calibrating features for semantic role labeling. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.
- YAROWSKY D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*, p. 189–196, Stroudsburg, PA, USA.
- YE P. & BALDWIN T. (2006). Verb sense disambiguation using selectional preferences extracted with a state-of-the-art semantic role labeler. In *Australasian Language Technology Workshop*, p. 141–148, Sydney, Australia.
- YOONG KEOK L. & HWEI TOU N. (2002). An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, p. 41–48, Stroudsburg, PA, USA.