



HAL
open science

Vers l'annotation automatique des noms prédicatifs d'évènements médiatiques

Dina Garnier-Oeliarisoa, Delphine Battistelli, Jean-Luc Minel

► **To cite this version:**

Dina Garnier-Oeliarisoa, Delphine Battistelli, Jean-Luc Minel. Vers l'annotation automatique des noms prédicatifs d'évènements médiatiques : Perspectives actuelles. Fryni Kakoyianni-Doa. Penser le Lexique-Grammaire, Honoré Champion, pp.413-426, 2014, Sciences du Langage, histoire de la langue et des dictionnaires, 978-2-7453-2512-9. halshs-00951182

HAL Id: halshs-00951182

<https://shs.hal.science/halshs-00951182>

Submitted on 24 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dina Garnier-Oeliarisoa¹, Delphine Battistelli², Jean-Luc Minel¹

(1) MODYCO, Université Paris-Ouest Nanterre La Défense, (2) STIH, Université Paris Sorbonne

dina.garnier.oeliarisoa@gmail.com, delphine.battistelli@paris-sorbonne.fr, jminel@u-paris10.fr

VERS L'ANNOTATION AUTOMATIQUE DES NOMS PREDICATIFS D'EVENEMENTS MEDIATIQUES

Résumé : Nous proposons une approche linguistique pour la reconnaissance automatique dans un corpus de dépêches d'agence d'événements dits « médiatiques » (et qualifiés comme tels par une partie de la communauté des journalistes). Dans cet article, nous nous intéressons aux noms prédictifs d'événements. Nous présentons notre démarche visant à construire une base de connaissance structurée à partir des travaux de description linguistique de S-H.Lee et G.Gross.

Mots-clés: événement, nom prédictif d'événement, prédicat

Abstract : We present a linguistic approach for the automatic processing of “media events” . In this paper, we focus on the predicative nouns of events. We present our approach for a structured knowledge base that we developed from the linguistic framework of S-H.Lee and G.Gross.

Keywords : event, predicative noun of event, predicate

1 Introduction

La notion d'événement est complexe à définir dans le champ des sciences humaines et sociales comme dans le champ du traitement automatique des langues (TAL) et de ses applications en recherche d'information (RI). Elle peut par exemple être abordée selon un principe de description de la catégorisation aspectuelle de situations (voir les nombreux travaux de linguistique générale comme de TAL portant sur ce thème) comme selon un principe de détection de situations perçues comme « nouvelles » dans un flux d'informations disponibles à un moment donné (voir notamment la tâche de RI appelée *Event Detection Tracking*). Il existe du reste (presque) autant de définitions que de champs disciplinaires dans lesquels la notion d'évènement est amenée à être considérée. Selon (Prestini-Christophe, 2006), il existe cependant des points communs à toutes ces définitions : (i) un évènement correspond à une « rupture » ; (ii) un « fait » devient « évènement » en fonction du monde dans lequel il advient (autrement dit, l'évènement est « subjectif »). Dans une démarche relevant essentiellement du paradigme de l'analyse du discours, quelques travaux se sont ainsi interrogés sur le rôle de la langue dans la construction, la représentation et l'interprétation des évènements en lien avec la mémoire, la culture, l'histoire, la société (voir notamment (Van de Velde, 2006), (Lecolle, 2009)). Ils ont pour objet de travailler sur les relations entre évènements du monde et évènements discursifs, entre un évènement mondialisé et ses différents traitements, entre évènements et normes/sens commun.

A la croisée de travaux de linguistique générale et d'analyse du discours, nos travaux ont pour objectif de participer à la « clarification » de la notion d'évènement *via* la définition de procédures TAL ancrées dans la spécification de besoins de communautés d'utilisateurs données. Nous nous intéressons au repérage, dans des textes de type dépêches d'agence, d'évènements dits « médiatiques » et qualifiés comme tels par les journalistes de l'AFP avec lesquels nous sommes amenés à collaborer dans le cadre du projet RMM2¹. Leurs besoins en matière de reconnaissance des évènements sont décrits dans les termes suivants :

¹ Relax Multimédias 2 : <http://www.rmm2.org>

« informations qui annoncent un événement à venir ou qui vient de commencer (et qui va se poursuivre pendant quelques temps), participation ou non d'une personnalité à un événement culturel ou encore informations qui peuvent avoir un intérêt pour le grand public ».

Cherchant à caractériser l'expression de ces besoins dans les termes d'une analyse linguistique automatisable (au moins en partie), nous proposons de procéder à une annotation de plusieurs types de marqueurs linguistiques liés à la notion d'événement (parmi ces marqueurs, ceux de temps, de lieu et de participant). Nous présentons ici les premiers éléments de notre démarche quant au repérage de noms prédictifs d'événements, forçant à une analyse approfondie des notions de prédicat et d'argument, à l'instar des travaux de (Harris, 1976) et de ceux menés dans le cadre du lexique-grammaire (Gross, 1975, Gross, 1981). La section 2 de cet article présente la manière dont le TAL aborde généralement la notion d'évènement. Nous présentons en section 3 la typologie sémantique de noms prédictifs d'événements établie par (Lee, 2001). La section 4 évoque une proposition de structuration en termes de traits (c'est à dire dans un format compatible avec un traitement informatique) de la typologie présentée précédemment. Nous présentons enfin deux expérimentations ainsi que l'évaluation de ces dernières en section 5. La première expérimentation vise à identifier les différentes structures linguistiques des événements médiatiques dans des textes journalistiques comportant des expressions simples mais également longues et pouvant être complexes ; la seconde a pour objectif de mesurer la couverture et la pertinence de la notion de nom prédictif d'évènement « décontextualisé ».

2 La notion d'évènement dans le champ du TAL

Dans le champ du TAL et de ses applications en RI, de plus en plus de travaux s'intéressent à la détection et à l'annotation des événements dans les textes (voir par exemple pour l'anglais (LDC, 2005)). Certains, notamment ceux décrits dans (Arnulphy et al, 2010), les considèrent comme des entités nommées à part entière ; ils se basent sur une liste pré-établie de noms déclencheurs d'évènement, puis considèrent l'analyse de certains indices pour détecter des expressions dites « candidates » représentant des événements dans certains contextes. D'autres travaux, notamment ceux adoptant le schéma d'annotation TimeML (Pustejovsky et al, 2005) comme par exemple (Sauri et al, 2005), (Bittar, 2008, 2009a) ou (Parent et al, 2008), inscrivent l'analyse des événements dans une problématique plus large qui est celle de la temporalité linguistique. (Sauri et al, 2005) décrit ainsi un module informatique nommé *Evita* (« *Events In Text Analyzer* »), où des méthodes symboliques sont utilisées pour détecter les verbes, adjectifs et noms événementiels considérés comme des événements candidats². La particularité de ce système est qu'il ne possède pas une liste pré-établie de patrons d'événements et qu'il n'est pas associé à un domaine en particulier. Tous ces travaux autour de la norme TimeML s'intéressent comme nous à l'annotation de différents marqueurs de surface. Encore plus proches de nous, (Bittar, 2009b, 2010) s'intéresse plus particulièrement à l'intégration des constructions à verbe support dans TimeML et propose des consignes linguistiquement motivées pour l'annotation de la relation entre un verbe support et un nom prédictif d'évènement.. Les événements verbaux y sont classés selon les classifications des verbes dans TimeML ; les noms déverbaux sont classés selon la classification du verbe associé au nom déverbal tandis que les noms événements qui ne sont pas des déverbaux appartiennent à la classe *OCCURRENCE*.

L'objet du travail présenté dans cet article est plus restreint puisqu'il ne concerne pas à proprement parler l'analyse de cette relation entre un verbe support et un nom prédictif ; et différent dans la mesure où nous nous intéressons à la classification sémantique des noms prédictifs d'événements en nous fondant sur des travaux linguistiques issus de l'approche de G. Gross, pour en mesurer à la fois la pertinence et la couverture. Dans une perspective de

² Pour les événements de type nom, des méthodes statistiques sont également utilisées pour la désambiguïsation.

traitement automatique, il est souvent indispensable de structurer les connaissances décrites par les travaux en linguistique générale afin qu'elles puissent être exploitables informatiquement. Les tables du lexique-grammaire (Gross, 1975, 1981) établissent par exemple une correspondance très riche entre d'une part les verbes, adverbess, noms prédicatifs, expressions figées et d'autre part leurs constructions syntaxiques respectives. Si les informations linguistiques présentes dans les tables sont structurées, elles restent cependant difficiles à exploiter informatiquement. (Constant & Tolone, 2010) ont mis en place un système permettant de transformer les données des tables du lexique grammaire dans un format exploitable informatiquement. Un second travail (Sagot & Tolone, 2009) a été réalisé afin de transformer les données des tables dans un format compatible avec un analyseur syntaxique. La dimension sémantique est présente dans cette base de connaissances mais cette dernière a été réalisée principalement dans un objectif d'analyse syntaxique. Les travaux de (Lee, 2001), qui s'inscrivent eux dans les travaux sur les classes d'objets de G. Gross³, offrent une description détaillée du comportement des noms prédicatifs d'événements en particulier ; une typologie sémantique de ces derniers, fondée sur des critères syntaxiques, y est proposée, mais cette analyse n'est pas exploitable directement en TAL. Dans notre perspective de traitement des événements médiatiques, nous proposons une structuration de ces connaissances et une implémentation.

3 Typologie sémantique des noms prédicatifs d'événements

Nous avons choisi d'annoter les marqueurs d'événements médiatiques en nous fondant sur la typologie des noms (prédicatifs) d'événements décrite par (Lee, 2001) qui s'inscrit dans la théorie des classes d'objets de G. Gross. (Lee, 2001) s'intéresse en particulier aux noms d'événements définis par la construction en *avoir lieu*, c'est à dire aux noms d'événements qui acceptent la construction rappelée en (1).

(1) Dét Nom d'événement *avoir lieu* W

où **Dét** désigne '**un**', '**le**' ou '**ce**' et **W** d'éventuels **compléments circonstanciels**.

(Lee, 2001) propose par ailleurs les définitions suivantes :

- Les *verbes supports* sont des « verbes vides ou vidés de leur sens lexical d'origine (...) et qui n'ont pour rôle que d'actualiser⁴, dans une phrase simple, un terme prédicatif n'appartenant pas à la catégorie du verbe » (Lee, 2001). Par exemple, le verbe 'donner' dans la phrase « *L'Orchestre de Paris donnera 41 de ses 85 concerts ou représentations à la Salle Pleyel* » et le verbe 'être' dans la phrase « *Usher est en concert à Paris Bercy ce lundi 24 Janvier 2011* » sont des verbes supports.
- Contrairement aux verbes supports, les *verbes opérateurs* sont prédicatifs et en relation avec des noms prédicatifs. Par exemple, les verbes 'jouer' et 'diriger' dans la phrase « *Le jour de ses 70 ans (samedi 20 février), Christoph Eschenbach jouera et dirigera deux concertos pour piano de Mozart* » et le verbe 'célébrer' dans la phrase « *Le festival de danse contemporaine Les Hivernales d'Avignon célébrera dès jeudi soir et jusqu'au 28 février ses 30 ans* » constituent des verbes opérateurs.

Le travail de Lee (2001) a débouché sur la création de classes d'objets-événements. Ces derniers permettent la classification de tous les noms d'événements selon des critères à la fois syntaxiques et sémantiques. (Lee, 2001) part du principe qu'il existe une corrélation entre les types aspectuels des situations décrites (état, action, événement, ...) et les différents *supports* et *opérateurs* qu'il qualifie de « *généraux* ». En d'autres termes, les verbes supports

³ Voir par exemple (Pesant & Mathieu-Colas, 1998) pour une introduction à la notion de classe d'objets.

⁴ Ou de « conjuguer » selon (Giry-Scheider, 1987), c'est-à-dire situer dans le temps un prédicat.

et verbes opérateurs généraux permettent d'aider à la caractérisation de traits aspectuels comme état, action, événement,... Par exemple, les constructions verbales « il y a » et « avoir lieu » désignent des verbes supports généraux du trait « événement » ; les verbes « faire » et « donner » sont des supports généraux du trait « action » et le verbe « avoir » ou « être Prép⁶ » sont des supports généraux du trait « état ». Lee (2001) a établi, pour chaque classe d'objet d'événements, une description du nom prédicatif d'événement ainsi que des différents supports et opérateurs généraux et ceux qualifiés d'« appropriés » qui lui sont associés. Ces derniers, contrairement aux supports et opérateurs généraux, permettent de caractériser les classes d'objets qui sont des sous-classes sémantiques des traits (qui, rappelons-le, peuvent eux être définis à l'aide des supports et opérateurs « généraux »). Par exemple, la classe <crimes et délits> est définie grâce aux verbes supports appropriés « commettre » ou « perpétuer » ainsi que grâce au verbe opérateur approprié « commanditer » tandis que la classe <cérémonie d'union> est caractérisée par les verbes opérateurs appropriés « célébrer » ou « bénir ». Ces différents supports et opérateurs (généraux et appropriés) peuvent être exploités d'une part comme des déclencheurs d'événements et d'autre part pour permettre la désambiguïsation de l'usage événementiel ou non du mot ; par exemple, pour distinguer l'emploi du mot 'mariage' dans les deux phrases suivantes : « *Autre incorrection culinaire annoncée : le mariage du champagne et des fromages les plus coulants, voire les plus forts, tels le munster...* » et « *Hélène Mandroux se dit prête à célébrer un mariage homosexuel....* ».

La typologie proposée consiste donc en un regroupement de noms (ou de groupes nominaux) selon des critères à la fois syntaxiques et sémantiques. Elle comporte les quatre classes principales suivantes, qui se divisent à leur tour en plusieurs sous-classes :

- événements <**fortuits**>. Ils concernent les événements de type *séisme, inondation, incendie, explosion, fuite de gaz,....*
- événements <**à double caractère**>. Ils concernent les événements de type *mutinerie, révolution, manifestation, grève, coup d'Etat, échanges de tirs, altercation,....*
- événements <**cycliques**>. Ils correspondent aux événements itératifs comme *Noël, 14 juillet, Mardi gras,*
- événements <**créés**>. Ils concernent les événements de type *colloque, mariage, congrès, conférence, concert, exposition, foire, marché, championnat, Grand Prix,....*

4 Construction d'une base de connaissances structurées et annotation

Dans une première étape de notre travail, nous proposons de structurer et d'implémenter les connaissances liées aux noms prédicatifs d'événements de façon « décontextualisée ». Cela signifie que nous considérons les noms prédicatifs ainsi que les classes et catégories décrites par (Lee, 2001) sans prendre en compte les différents supports et opérateurs qui leur sont associés. Nous avons fait ce choix pour deux raisons. D'une part, nous souhaitons mesurer la couverture des données dans la base de connaissances et tester la pertinence ou non des catégories et classes de la typologie. D'autre part, la prise en compte des verbes supports et opérateurs associés aux noms prédicatifs nécessite des développements plus longs à mettre en œuvre. Nous traiterons donc ce problème plus complexe dans une seconde étape. La structure de traits associée à un nom ou à un groupe nominal que nous avons retenue à ce jour est décrite dans la figure 1.

TYPE nom prédicatif d'événement TRAITS graphie : conférence, concert, cyclone,... categoGram : N,.. classe : <spectacles>, <réunions intermittentes>, ... categorie : cree, cyclique,...

Figure 1. Structure de traits associée à un nom (ou groupe nominal) événementiel

⁶ Prép : préposition

Cette structure est décrite dans des fichiers XML et implémentée dans les dictionnaires au format Unitex (Paumier, 2008). Comme notre démarche est d'interroger des bases de connaissances (structurées) pour la recherche et l'annotation des informations linguistiques, les grammaires locales nous semblent en effet plus appropriées que les techniques avec expressions régulières. L'implémentation se déroule en trois étapes : analyse linguistique, annotation puis validation humaine permettant l'enrichissement incrémental de la base de connaissances. L'étape d'analyse linguistique est une étape classique d'analyse morphologique et lexicale. Nous faisons appel aux ressources et programmes d'Unitex (Paumier, 2008). Pour l'analyse du temps morphologique des verbes, nous faisons appel au Treetagger (Schmid, 1994). Notre système reçoit en entrée soit un fichier au format texte (.txt), soit un fichier au format xml. Pour l'étape d'annotation, nous faisons appel aux connaissances (structurées) stockées dans les dictionnaires (au sens d'Unitex) et à des grammaires locales d'Unitex. Notre système produit en sortie un fichier annoté et structuré au format xml. Dans la troisième étape, notre système offre la possibilité d'une validation humaine des annotations produites avec dès lors un enrichissement incrémental de la base de connaissances.

5 Expérimentation et évaluation

Rappelons que notre objectif est de proposer une approche linguistique pour le repérage automatique d'événements dits médiatiques. Il est donc important de déterminer les différentes formes ou structures linguistiques associés à ces événements. Nous avons mis en place deux expérimentations. La première vise à mesurer les distributions des différentes « formes » (ou structures linguistiques) d'événements médiatiques identifiées. La seconde expérimentation consiste à mesurer la couverture des données dans la base de connaissances et à mesurer la pertinence de la classification de (Lee, 2001) à partir de noms prédicatifs d'événements « décontextualisés ».

Nos corpus sont constitués de dépêches AFP fournis par les journalistes qui collaborent avec nous dans le cadre du projet RMM2. Nous disposons de deux types de corpus : un corpus d'étude composé de 33 dépêches contenant 332 phrases et un corpus test composé de 100 dépêches contenant 1431 phrases. Nous n'avons pour l'instant pas mis en place une étude de la représentativité de notre corpus.

Une première étape d'analyse nous a permis d'identifier cinq types de structures. Elles sont présentées en relation avec les exemples (a) à (e) dans la figure 2. Les éléments en gras désignent les supports ou opérateurs tandis que les éléments soulignés désignent les noms prédicatifs d'événements ; nous ne mettons pas en gras les supports ou opérateurs des noms prédicatifs qui ne sont pas à usage événementiel (cf. exemples (x) et (x')).

x. Ce rendez-vous avait été lancé en août 2006 dans le cadre intime de l'abbaye cistercienne du XVIII^e siècle de Valloires par le Britannique Adam Gatehouse, producteur de concerts à la BBC et chef d'orchestre.

x'. Trois concerts auront lieu le 29 mars, dans les salons de la mairie du XIII^eème, le 31 mars dans l'amphithéâtre Richelieu de la Sorbonne et enfin le 2 avril à l'Unesco

a. Seul le rez-de-chaussée du musée d'Orsay est ouvert (moins l'exposition Masques, mais avec les expositions Pastels et Picasso) et de ce fait gratuit

b. Le jour de ses 70 ans (samedi 20 février), Christoph Eschenbach **jouera** et **dirigera** deux concertos pour piano de Mozart, puis s'entourera de musiciens amis (le violoniste Gidon Kremer, le violoncelliste Yo-Yo Ma et le pianiste Tzimon Barto) pour un concert de gala

b'. Une exposition inédite sur "l'ombre" dans les oeuvres d'art, de la Renaissance au XX^e siècle, **est organisée** à partir de mardi et jusqu'au 17 mai par le musée madrilène Thyssen-Bornemisza

c. La dixième édition de la Biennale de Lyon, intitulée "Le spectacle du quotidien", **aura lieu** du 16 septembre 2009 au 3 janvier 2010 et sera conduite par Hou Hanru, commissaire, et Thierry Raspail, directeur artistique, indiquent les organisateurs

d. L'Orchestre de Paris **donnera** 41 de ses 85 concerts ou représentations à la Salle Pleyel, sa résidence jusqu'à l'ouverture de la Philharmonie de Paris à l'horizon 2012-2013

d'. Après Chaillot, le spectacle sera donné en mars à Combs-la-Ville (Seine-et-Marne, les 17 et 18), Cusset (Allier, le 26), le 10 avril à Rethel (Ardennes) et le 28 mai à Cahors

e. Usher **est en concert** à Paris Bercy ce lundi 24 Janvier 2011 !!

Figure 2. Structures linguistiques possibles

Les noms prédicatifs d'événements « décontextualisés » que nous étudions plus particulièrement dans cet article sont des noms intervenant dans une construction en « avoir lieu » telle que celle donnée en (1). Les critères de différenciation des structures sont à la fois syntaxiques (fonction syntaxique du nom par rapport au verbe, mode de la phrase,...) et sémantiques (rôle sémantique du nom, type de procès, sens lexical du nom,...). Les structures s'analysent de la façon suivante. Dans (x) et (x') les noms ne sont pas employés dans un sens événementiel ; le sens lexical du nom est monosémique dans (x) tandis qu'il peut être polysémique dans (x'). De (a) à (e), les noms sont employés dans un sens événementiel. Dans (a), les noms semblent n'être en relation (ou du moins difficilement reliables) avec aucun verbe (support ou opérateur). Dans (b) et (b'), les noms sont en relation avec des verbes opérateurs indiquant un procès de type « action » ; le nom est complément du verbe dans (b) tandis qu'il est sujet dans (b'). Dans (c), le nom est en relation avec un verbe support indiquant un procès de type « événement » et est sujet du verbe. Dans (d) et (d'), les noms sont en relation avec des verbes supports indiquant un procès de type « action » ; le nom est complément du verbe dans (d) tandis qu'il est sujet dans (d'). Dans (e), le nom est en relation avec un verbe support indiquant un procès de type « état » et est complément de ce verbe. La distribution des événements médiatiques selon leurs structures linguistiques est présentée dans le tableau 1.

Type de structure	Distribution	combinaison
(a)	43,91%	Nom prédicatif « isolé »
(b) et (b')	21,26%	Nom prédicatif-Opérateur
(c),(d),(d') et (e)	34,45%	Nom prédicatif-Verbe support

Tableau 1. Distribution des événements médiatiques selon leurs structures linguistiques

Notre évaluation fait appel aux critères suivants :

- emploi événementiel ou non d'un nom ou d'un groupe nominal. Par exemple le mot 'mariage' n'a pas un emploi événementiel dans « le mariage du champagne et des fromages », alors qu'il le possède dans « célébrer un mariage » ;
- appartenance ou non du nom prédicatif à une classe de la typologie de Lee (2001). Par exemple, le nom prédicatif *manifestation* (au sens de manifestation culturelle) apparaissant dans la phrase « *La manifestation* intitulée "Pierre Bonnard, scènes d'intérieur" montre 80 toiles, dessins et aquarelles réalisés entre 1923 et 1947 (année de sa mort), pour la plupart dans cette bourgade qui surplombe Cannes et où il avait acquis une maison avec son épouse et muse, Marthe » doit être distingué du nom prédicatif *manifestation* (au sens de descente dans la rue) apparaissant dans la phrase « *Le Tibet* reste un sujet extrêmement sensible en Chine, Pékin étant régulièrement accusée par les responsables tibétains en exil, dont leur

chef spirituel, le dalaï lama, et les militants de leur cause de réprimer férocement toute manifestation favorable à l'indépendance du territoire himalayen ».

- savoir si un événement est considéré comme médiatique ou pas reste subjectif (et lié à une communauté donnée). Dans le cadre de ce travail, nous proposons de mener une évaluation de notre annotation en fonction de ce qui est désigné comme étant un événement médiatique pour les journalistes de l'AFP.

Les résultats de cette évaluation sont présentés dans le tableau 2.

	Corpus d'apprentissage		Corpus test	
	rappel	précision	rappel	précision
Nom prédicatif à usage événement médiatique ET bien classifié	60,91%	89,44%	50,00%	87,55%

Tableau 2.

Le taux de précision élevé montre que les noms prédicatifs « décontextualisés » sont généralement utilisés dans un usage événementiel et que la classification fondée sur la typologie de (Lee, 2001) est pertinente pour la situation et le contexte d'utilisation. En revanche, le taux de rappel bas montre que les noms prédicatifs « décontextualisés » qui ont été recensés dans (Lee, 2001) ne couvrent pas la (ou les) thématique (s) du corpus de textes qui a été annoté. La prise en compte des verbes supports et opérateurs (généraux et appropriés) dans un travail ultérieur devrait permettre une amélioration de ce taux de rappel.

6 Limites et perspectives

Ce travail nous a permis de constater que les événements médiatiques se présentent sous plusieurs formes et plusieurs « structures linguistiques » dans les textes. Un nom prédicatif peut être accompagné d'un verbe support ou d'un verbe opérateur (prédicatif) mais ce n'est pas forcément toujours le cas. Nous faisons le même constat que celui établi dans (Laporte et al, 2008) : il existe un certain nombre de noms prédicatifs qui n'occurrent pas avec un verbe support. Nous pourrions même ajouter qu'ils n'occurrent pas nécessairement non plus avec des verbes opérateurs. Pourtant ils représentent bien des événements dits médiatiques (cf. éléments de la structure (a) dans la figure 2). Ce travail nous a également permis de constater que, même à travers des noms prédicatifs d'événements « décontextualisés », la classification de (Lee, 2001) était adaptée à l'analyse et au repérage des événements médiatiques. Notre travail présente cependant en l'état actuel encore beaucoup de limites. Tout d'abord, les différentes formes linguistiques des événements médiatiques que nous avons identifiées n'ont pas été définies de manière formelle. Le problème est complexe. Il vient en partie du fait que nous avons éprouvé une difficulté à cerner des critères qui permettent de différencier par exemple si un élément est prédicat ou argument, support ou opérateur dans des expressions longues et complexes provenant de textes réels (dépêches d'agences) et non construits comme dans le cadre du travail de (Lee, 2001). Ensuite, ayant travaillé sur un corpus de taille réduite, les cinq structures linguistiques que nous avons identifiées ne sont certainement pas suffisantes pour représenter les différentes formes linguistiques des événements médiatiques. Ces structures linguistiques pourront donc être modifiées ou enrichies avec un travail sur des corpus plus volumineux. Enfin, la structuration des connaissances que nous avons brièvement montré a été limitée aux noms prédicatifs d'événements « décontextualisés », ce qui ne reflète pas la richesse de la typologie proposée dans (Lee, 2001). De manière plus générale, ce travail nous a amenés à considérer dans toute sa complexité l'identification de ce qui fonctionne comme élément prédicatif et/ou comme argument dans une unité phrastique. Nos travaux futurs visent à aborder de manière plus précise et détaillée ce fonctionnement dans une perspective TAL.

Remerciements

Ce projet est financé par le projet ANR RMM2.

Références

- Arnulphy B., Tannier X. & Vilnat A. (2010). Les entités nommées événement et les verbes de cause-conséquence. In Actes TALN 2010, Montréal, Canada.
- Bittar A. (2008). Annotation des informations temporelles dans des textes en français. In RECITAL 2008, Avignon, France.
- Bittar A. (2009a). Annotation of events and temporal expressions in french texts. In Proceedings of the Third Linguistic Annotation Workshop, p. 48–51: ACL and AFNLP.
- Bittar A. (2009b). Intégration des constructions à verbe support dans TimeML. In TALN 2009, Senlis France.
- Bittar A. (2010). Construction d'un TimeBank du français : un corpus de réfé annoté selon la norme ISO-TimeML. Thèse de doctorat, Université Paris Diderot.
- Constant M. (2003). Grammaires locales pour l'analyse automatique de textes: Méthodes de construction et outils de gestion. PhD thesis, Université de Marne-la-Vallée.
- Constant M. & Tolone E. (2010). A generic tool to generate a lexicon for NLP from Lexicon-Grammar tables. In M. D. Gioia, Ed., Actes du 27e Colloque international sur le lexique et la grammaire (L'Aquila, 10-13 septembre 2008). Seconde partie, volume 1 of Lingue d'Europa e del Mediterraneo, Grammatica comparata, p. 79–93. Aracne. ISBN 978-88-548-3166-7.
- Giry-Schneider J (1987) Les prédicats nominaux en français. Les phrases simples à verbe support, Genève, Droz.
- Gross M. (1975). Méthodes en syntaxe. Paris : Hermann.
- Gross M. (1981). Les bases empiriques de la notion de prédicat sémantique. p. 7–52: Langages.
- Harris Z. S. (1976), Notes du cours de syntaxe, trad .par M.Gross, Paris, Le Seuil.
- Laporte E., Ranchhod E. & Yannacopoulou A. (2008). Syntactic variation of support verb constructions. *Linguisticae Investigationes*, 31(2), 173–185.
- LINGUISTIC DATA CONSORTIUM (2005). ACE English Annotation Guidelines for Events, Version 5.4.3 2005.07.01, http://www ldc.upenn.edu/Projects/ACE/docs/English-Events-Guidelines_v5.4.3.pdf
- Lecolle M (2009). « Changement de sens du toponyme en discours : de Outreau “ville” à Outreau “fiasco judiciaire” », Les Carnets du Cediscor 11, p. 90-106, Presses Sorbonne Nouvelle, Paris, 2009.
- Lee S.-H. (2001). Les classes d'objets d'événements. Pour une typologie sémantique des noms prédicatifs d'événements. Thèse de doctorat, Université Paris 13.
- Parent G., Gagnon M. & Muller P. (2008). Annotation d'expressions temporelles et d'événements en français. In Actes de TALN.
- Paumier S. (2008). Unitex 2.0 User Manual.
- Pesant D. L. & Mathieu-Colas M. (1998). Introduction aux classes d'objets. *Langages*, 131, 6–33.
- Prestini-Christophe M. (2006). Une nouvelle grille de lecture : l'événement. *Pensée plurielle*, 13, 81–90.
- Pustejovsky J., Castano J., Ingria R., Saurí R., Gauzauskas R., Setzer A. et Katz G. (2005). TimeML: Robust specification of event and temporal expressions in text. Actes de IWCS-5, Fifth International Workshop on Computational Semantics.
- Sagot B. & Tolone E. (2009). Exploitation des tables du Lexique-Grammaire pour l'analyse syntaxique automatique. *Arena Romanistica, Journal of Romance Studies*, 4, 302–312. ISSN 1473-3536.
- Saurí R., Knippen R., Verhagen M. & Pustejovsky J. (2005). Evita: a robust event recognizer for qa systems. In HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, p. 700–707, Morristown, NJ, USA: Association for Computational Linguistics.
- Schmid H. (1994). Probabilistic part-of-speech tagging using decision trees. In Proceedings of International Conference on New Methods in Language Processing.
- Van de Velde D. (2006). Grammaire des événements. Presses Universitaires du Septentrion.