

# From Texts to Ontologies: Non-taxonomical Relation Extraction

Marie Chagnoux  
IRIT - University of Toulouse  
118 route de Narbonne  
F- 31062 Toulouse Cedex 9  
chagnoux@irit.fr

Nathalie Hernandez  
IRIT - University of Toulouse  
118 route de Narbonne  
F- 31062 Toulouse Cedex 9  
hernande@irit.fr

Nathalie Aussenac  
IRIT - University of Toulouse  
118 route de Narbonne  
F- 31062 Toulouse Cedex 9  
aussenac@irit.fr

## ABSTRACT

We propose a methodological approach which aims at creating and acquiring patterns from texts in order to enrich lightweight ontologies with non taxonomical relations.

Our approach relies on an existing ontology, a morpho-syntactically tagged corpus and an evolutive pattern bank. Our goal is twofold : (i) the morpho-syntactic patterns stored in the base are used to identify new relations between the concepts of the ontology (ii) new patterns are defined according to the syntactic contexts in the corpus of the terms that are concept labels. These patterns are added to the base and can be reused on a new corpus. At the end of the process, the initial ontology is enriched with new relations extracted from the corpus. The originality of our approach is that we guide the pattern definition process by suggesting abstract forms of patterns according to the context of co-occurring terms.

## Categories and Subject Descriptors

I.4 [Computing Methodologies]: Knowledge Representation Formalisms and Methods

## Keywords

Ontology building, morpho-syntactic patterns, semantic relations

## 1. INTRODUCTION

L'extraction de relations à partir de textes intervient de deux manières complémentaires en ingénierie des ontologies. Elle sert à l'identification automatique de relations entre instances de concepts ; ceci relève du peuplement d'ontologies. Elle peut correspondre aussi à l'extraction de relations ou de *properties* entre des classes de concepts et participe alors de la construction d'ontologies ou, dans une perspective d'automatisation, de l'apprentissage d'ontologies.

D'une part, on peut distinguer les récents travaux en peuplement d'ontologie qui s'appuient sur des approches com-

binées de statistiques et de linguistique [16] [22] [17] [8]. Les bons résultats qu'ils obtiennent proviennent du caractère fortement contraint des ressources utilisées pour la tâche : les concepts et les relations sont déjà définis dans l'ontologie; des documents annotés manuellement sont disponibles ainsi que des listes d'étiquettes d'instances de concepts. Les relations recherchées sont exprimées par un nombre réduit de variations phrastiques listées *a priori*. Un certain nombre d'approches sophistiquées ont été proposées pour l'extraction d'instances, comme celles utilisées par des outils de recherche d'Entités Nommées. Brewster et Ciravegna [6] proposent ainsi des algorithmes d'apprentissage pour l'extraction de relations par patrons. Leur système peut définir un ou plusieurs patrons à partir d'un ensemble d'expressions linguistiques qui ont été identifiées comme contenant une relation entre des instances de concepts. L'apprentissage est particulièrement pertinent dans les domaines où les textes ont un contenu comportant d'importantes régularités de formulation. Ainsi, par exemple, de très nombreux travaux [7] ont été menés dans le domaine biomédical où les textes ont l'avantage de présenter des structures fortement redondantes qui peuvent être facilement abstraites dans des patrons reproductibles.

D'autre part, la construction d'ontologies soulève des problèmes plus complexes puisque ni les concepts ni les relations ne sont connus : les deux types d'éléments doivent justement être identifiés. Si nombre de travaux ont été consacrés à l'extraction de termes en tant qu'étiquettes de concepts et à l'extraction de relations taxonomiques [5], l'extraction de relations non-taxonomiques demeure une gageure [12]. L'absence d'ancrage initial la rend difficile à appréhender par des méthodes d'apprentissage. Dans ce contexte, la projection de patrons linguistiques sur des textes est une méthode plus robuste à condition qu'un humain supervise le processus. A partir des idées de Hearst, nous avons proposé un premier outil, Caméléon, qui implémente cette projection de patrons sur des corpus pour identifier des relations et des concepts pour la construction d'ontologies [18]. Une seconde version de Caméléon a permis de prendre en compte des textes taggés syntaxiquement. Néanmoins, l'utilisation de Caméléon dans différents domaines a confirmé qu'ajuster manuellement des patrons et filtrer les phrases correspondantes pour identifier des relations conceptuelles étaient coûteux et fastidieux [3]. Cet aspect est particulièrement renforcé pour les relations dépendantes d'un domaine et nous avons montré que les approches par patrons était plus efficaces pour la recherche de relations taxonomiques dans des documents pédagogiques [14].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JFO 2008 December 1-2, 2008, Lyon, France

Copyright 2008 ACM 978-1-60558-373-0/08/0003 ...\$5.00.

Notre objectif est de partir de ces expériences en extraction de relations à base de patrons et en construction d'ontologies avec Caméléon pour préserver la robustesse de l'extraction tout en améliorant le processus d'acquisition. Nous proposons donc un cadre méthodologique qui assiste davantage l'utilisateur pendant les tâches d'identification de patrons et de relations quand une hiérarchie de concepts et les termes afférents sont disponibles. En conservant les hypothèses développées dans Caméléon, cette méthodologie met l'accent sur la réutilisation et l'adaptation de patrons. Son originalité réside dans son efficacité à identifier des relations entre des termes liés dans une phrase en fonction de patrons existants. La principale évolution est la suggestion automatique de patrons propres au corpus : le système suggère des patrons abstraits du contexte où les termes référant à des étiquettes de concepts co-occurrent.

Pour présenter ce cadre méthodologique, dans une première partie, nous introduirons les approches par patrons en insistant sur leur apport à la construction d'ontologies. Nous détaillerons les principales problématiques dans la seconde partie avant de présenter notre cadre applicatif, notre algorithme et son implémentation dans la dernière partie.

## 2. LES APPROCHES PAR PATRONS : DÉFINITION ET PROBLÉMATIQUE

En linguistique, les approches par patrons sont utilisées pour associer des régularités structurelles à des informations sémantiques. Hearst fut la première à les utiliser dans le contexte de l'extraction d'information afin d'extraire des relations linguistiques [13]. Elle proposa d'identifier des ensembles de patrons lexico-syntaxiques qui soient facilement reconnaissables, qui apparaissent fréquemment et qui dépassent les frontières du genre textuel. Ces patrons indiquent sans ambiguïté la relation lexicale visée ("*identify a set of lexico-syntactic patterns that are easily recognisable, that occur frequently and across text genre boundaries, and that indisputably indicate the lexical relation of interest*"). Dans la méthode qu'elle définit, l'identification des patrons susceptibles d'encoder une certaine relation lexicale ou sémantique suppose une interprétation de leurs occurrences. Depuis, de nombreux chercheurs ont utilisé ce cadre pour proposer des modèles d'acquisition et d'extraction de connaissances ou améliorer la compréhension des textes [10]. Ils ont également évalué ses limites, la principale citée étant le coût d'acquisition des patrons. Pour y répondre, des algorithmes d'apprentissage ont été combinés avec des techniques de TAL afin de définir automatiquement des patrons à partir de corpus annotés.

### 2.1 La définition de patrons

Les approches par patrons dédiées à l'extraction de relations sémantiques reposent sur le postulat selon lequel une relation sémantique entre termes telle que l'hyponymie ou la méronymie peut *a priori* être inférée à partir d'analyses textuelles. Elle est alors abstraite dans un patron qui correspond potentiellement à toutes les réalisations langagières associées. Par exemple, [13] montre à partir de l'exemple 1 que le lecteur peut inférer qu'un "*Bambara ndang*" est une sorte de "*bow lute*" sans même savoir ce que sont un Bambara ndang et un bow lute.

- (1) "*The bow lute, such as the Bambara ndang, is plucked [...]*"

De fait, la relation hyponymique peut être encodée par la construction: "un terme suivi par *such as* et un autre terme". Cette relation hyponymique peut être abstraite au sein du patron  $P^{hyponym}_1$  proposé par 2<sup>1</sup>.

- (2)  $X \text{ such as } Y$ , où  $X$  et  $Y$  sont des syntagmes nominaux

L'extraction de relations sémantiques avec des approches par patrons s'organise en trois étapes : une *phase d'acquisition* pendant laquelle les patrons sont créés, une *phase de projection* sur corpus pendant laquelle les patrons servent à extraire des phrases contenant potentiellement des relations lexicales entre des couples de termes et une *phase de modélisation* pendant laquelle chaque relation lexicale permet de définir des couples de concepts associés par une relation dans l'ontologie.

Une relation donnée  $REL$  peut être extraite à l'aide de différents patrons dont l'ensemble est décrit par  $\mathcal{P}^{REL} = \{P^{REL}_1, \dots, P^{REL}_n\}$ . Un patron  $P^{REL}_i$  est une construction lexico-syntaxique qui est définie pour extraire une relation sémantique  $REL$  entre deux entités  $X$  and  $Y$ . Cette extraction suppose une relation lexicale implicite intermédiaire entre deux termes  $t1$  associé à  $X$  et  $t2$  associé à  $Y$ . Une relation  $REL$  peut être dénotée par plusieurs étiquettes ou termes notés  $Label(REL)$ . Projeté sur les textes,  $P^{REL}_i$  permet d'extraire des couples de termes  $REL^t(t1, t2)$  liés par la relation  $REL$ .

Ainsi, dans l'exemple proposé par Hearst,  $REL$  correspond à la relation  $HYPONYM$ . Les étiquettes de la relation sont définies comme  $Label(HYPONYM) = \{ "est un", "sorte de" \}$ . Quand le patron  $P^{hyponym}_1$  est projeté sur la phrase 1, le couple de termes extraits est  $Hyponym^t = (Bambara ndang, bow lute)$  où *Bambara ndang* et *bow lute* correspondent respectivement à  $t1$  et  $t2$ .

### 2.2 Création de patrons

Les patrons généralisent des expressions lexico-syntaxiques identifiées dans les textes qui encodent des relations lexicales particulières. Ils peuvent être créés par l'abstraction de toutes les relations lexicales correspondantes dans les textes. Cette abstraction se situe à différents niveaux selon les caractéristiques prises en compte. Ainsi, l'exemple 3 qui contient la relation  $REPORT$  permet de proposer plusieurs types de patrons.

- (3) "*The heads report to the vice president*"

Nous partons du principe que le corpus auquel appartient cet exemple a préalablement été annoté par un analyseur morpho-syntaxique et éventuellement un analyseur sémantique. Les patrons peuvent alors se composer d'éléments hétérogènes : lemmes, catégories grammaticales, rôles syntaxiques, catégories sémantiques et symboles interprétables par le moteur de recherche des patrons. La complexité de ces

<sup>1</sup>Les exemples de patrons sont modélisés selon une syntaxe simplifiée afin de faciliter la lecture et la compréhension. En réalité, il est nécessaire de distinguer d'une part l'expression linguistique d'un patron et d'autre part son implémentation qui la contraint par son formalisme. Tout patron "linguistique" peut être implémenté selon différents langages et méthodes qui pourront par ailleurs influencer la qualité de ses performances. Nous donnons un exemple de patron formalisé avec le langage JAPE dans la section 4.2.

combinaisons dépend de la nature des éléments sélectionnés. Ainsi, plusieurs patrons peuvent être associés à l'exemple 3 comme l'illustrent les exemples 12<sup>2</sup>.

- (4) a.  $P^{REPORT}_1: X \text{ report to the } Y$   
 b.  $P^{REPORT}_2: X \text{ VB IN DT } Y$   
 c.  $P^{REPORT}_3: X \text{ MD VB to DT } Y$   
 d.  $P^{REPORT}_{3a}: X \text{ MD VB[report] to DT } Y$   
 e.  $P^{REPORT}_4: X \text{ MD [report] to the } Y$   
 f.  $P^{REPORT}_5: X \text{ (JJ)? VB to the } Y$   
 g.  $P^{REPORT}_6: X \text{ MD VB to (**)? } Y$

Le patron peut être constitué uniquement d'éléments lexicaux comme dans  $P^{REPORT}_1$  où seuls sont utilisés la ponctuation et les mots sous leur forme originale ou sous leur lemme suivant l'outil utilisé.  $P^{REPORT}_2$  s'appuie sur des annotations syntaxiques où *VB* est un verbe, *IN* une préposition et *DT* un déterminant. Les patrons morpho-syntaxiques prennent en plus en compte les catégories morphologiques et les structures syntaxiques dans lesquelles s'inscrivent les éléments comme l'illustre l'exemple  $P^{REPORT}_3$  où *MD* correspond à une liste sémantique de verbes modaux et *to* signifie que la forme exacte de *to* doit être prise en compte. Les patrons morpho-syntaxiques comme  $P^{REPORT}_{3a}$  peuvent articuler les annotations syntaxiques et les formes lexicales (en général des lemmes).  $P^{REPORT}_4$  illustre la manière dont l'information sémantique peut être exploitée : *[report]* renvoie à une liste sémantique de verbes synonymes de "report" comme "be responsible to".

Il est possible d'étendre la portée de ces éléments en utilisant un ensemble d'opérateurs logiques comme "et", la négation, etc. Par exemple, dans  $P^{REPORT}_5$ , "(JJ)?" est un adverbe optionnel (*JJ* est un adverbe et l'opérateur? indique son caractère optionnel). De plus, la plupart des syntaxes<sup>3</sup> qui permettent d'exprimer des patrons proposent de définir une fenêtre de taille variable comme dans l'exemple  $P^{REPORT}_6$  où un ou deux éléments indéterminés peuvent s'intercaler entre "to" et "Y".

La combinaison entre les catégories et les opérateurs permet une très grande flexibilité : alors que le patron  $P^{REPORT}_1$  peut correspondre uniquement au segment linguistique "The heads report to the vice president",  $P^{REPORT}_5$  pourra s'aligner avec un segment plus complexe comme "The heads will directly report to the executive vice president". Cependant, l'une des plus grandes difficultés de la création de patrons réside dans le compromis entre des structures marginales et trop contraintes (qui conduisent à l'oubli de certaines occurrences et augmentent le silence) et des structures ouvertes trop permissives qui rapportent des phrases non pertinentes.

Un nom peut être étendu en y substituant un nom et un déterminant ; un nom, un déterminant et plusieurs adjectifs ou encore un syntagme nominal complexe comme dans les exemples suivants :

- (5) "the aircraft maker's first assembly facility"  
 (6) "one of China's three biggest airlines in terms of traffic"  
 (7) "the left- and right-hand engine pylons"

<sup>2</sup>Comme précédemment, tous ces exemples sont modélisés selon une syntaxe simplifiée.

<sup>3</sup>Il n'existe pas de syntaxe consensuelle puisque tout système de création et d'intégration de patrons propose sa propre grammaire et son propre langage.

## 2.3 Patrons et enrichissement d'ontologies

Notre objectif étant d'exploiter la projection de patrons pour enrichir une ontologie existante, les patrons ont pour fonction d'extraire des relations entre des concepts présents dans cette ontologie ou des nouveaux concepts qui seront ajoutés à l'ontologie. Cet objectif participe directement à la troisième phase qui consiste à aligner des couples de termes et des couples de concepts appropriés en identifiant la bonne relation sémantique et en enrichissant l'ontologie avec cette relation.

Selon la notation proposée en section 2.1, les couples  $REL^t$  ( $t1, t2$ ) doivent être transformés en couples  $REL^c$  ( $c1, c2$ ) où  $c1$  et  $c2$  sont deux concepts appartenant à l'ontologie. La principale difficulté réside dans le processus d'interprétation. Cette section vise à présenter quelques-unes des problématiques à résoudre à ce stade.

### 2.3.1 Identifier des concepts à partir de termes

L'identification des concepts référencés par les termes d'un couple pose entre autres trois types de difficultés : la polysémie, la métonymie et l'anaphore.

#### La polysémie.

Du point de vue des ontologies, la polysémie est un phénomène moins complexe que d'un point de vue général puisque les ontologies visent justement à rendre compte d'un domaine particulier. On parle de polysémie quand à un terme correspond une étiquette partagée par plusieurs concepts. Le terme peut donc se référer aux différents concepts  $c_j$  ou  $c_k$ . Par exemple, "plant" est, en anglais, polysémique car il peut référer au concept "factory" ou au concept "plant" au sens médical.

#### La métonymie.

La métonymie est un problème récurrent de la construction d'ontologies. Il y a métonymie quand un terme  $t$  ne doit pas être associé à  $c_j$  à qui il semble référer mais à  $c_k$  qui est fortement corrélé à  $c_j$ . Ainsi, la phrase, "the pilot climbed to height 600 m" doit être interprétée comme "l'avion est monté" puisque c'est bien l'avion qui a la propriété de s'élever.

#### L'anaphore.

L'anaphore est sans doute le principal problème à résoudre car il est le plus fréquent dans les textes. Les patrons créés sont contraints par la structure physique et jusqu'ici aucune méthode efficace n'a été proposée pour dépasser les limites de la phrase. Par conséquent, une grande partie de l'information pertinente est souvent ignorée puisque les patrons sont conçus de manière à explorer chaque phrase de manière indépendante. Quand deux termes associés apparaissent dans deux phrases adjacentes -ou pire, non adjacentes-, l'entité  $X$  ou l'entité  $Y$  -ou même les deux- peut faire défaut. Ainsi, par exemple, dans 8, le contenu sémantique de la seconde phrase ne sera pas pris en compte si le patron projeté est  $N \text{ VB[BE] [ADJ]? NP * [to introduce]}$ .

- (8) "The Airbus A310 is a medium to long-range wide-body airliner manufactured by Airbus SAS. It was the second model to be introduced by Airbus after the A300."

### 2.3.2 Alignement de relations sémantiques

Une fois que les concepts référencés par le couple de termes ont été identifiés, il faut enrichir l'ontologie avec la nouvelle relation extraite. Il est souvent délicat de déterminer à quel niveau de l'ontologie la relation doit être ajoutée. La relation peut être pertinente uniquement pour les concepts identifiés à partir des termes ou, au contraire, être généralisable à l'un des concepts parents.

Ainsi, par exemple, supposons qu'une ontologie comprenne les deux hiérarchies suivantes : *CF6-50 engine IS-A CFM engine IS-A engine IS-A Aircraftcomponent* ; *A300 IS-A AirbusAircraft IS-A Aircraft*. Alors, la relation  $POWER^t$  (*American General Electric CF6 - 50 engine, A300*) et  $POWER^t$  (*The Airbus A318 Elite, CFM engine*) extraite pour la relation  $POWER^t$  à partir des exemples 9 et 10 peut mener à plusieurs relations conceptuelles potentielles décrites en 11.

- (9) *“American General Electric CF6-50 engines powered the A300”*
- (10) *“The Airbus A318 Elite will be powered by CFM engines”*
- (11)
  - a.  $POWER^c$  (*CF6-50 engine, A300*)
  - b.  $POWER^c$  (*CFM engine, Airbus A318 Elite*)
  - c.  $POWER^c$  (*engine, AirbusAircraft*)
  - d.  $POWER^c$  (*engine, Aircraft*)
  - e.  $POWER^c$  (*Aircraftcomponent, A300*)
  - f.  $POWER^c$  (*Aircraftcomponent, Aircraft*)

Comme il est difficile d'automatiser la conceptualisation de la relation au bon niveau, la seule option réside dans la validation manuelle par un utilisateur. Deux options sont alors possibles : soit la validation est effectuée étape par étape au fur et à mesure des différentes phases de construction de l'ontologie, soit la validation est seulement opérée quand la vue générale est enfin disponible.

Comme le soulignent [6] et [3], de nombreuses validations intermédiaires peuvent être menées. La validation peut porter sur les phases où un patron est reconnu (résultant de la projection d'un patron, de tous les patrons associés à une relation ou de tous les patrons possibles), sur la relation lexicale potentiellement trouvée dans chaque phrase valide et ensuite de la relation conceptuelle candidate qui peut être inférée de ces relations lexicales.

La validation étape par étape, comme la proposait Caméléon est coûteuse en temps. Le risque est qu'il ne s'agisse que d'interprétations locales contraintes par la seule portée de la phrase si un contexte plus large - comme le paragraphe, une autre relation lexicale ou l'état courant de l'ontologie construite) n'est pas fourni à l'utilisateur. Pour dépasser cette limite, Caméléon permet de valider des relations conceptuelles en fonction de la position des concepts liés dans l'ontologie. Cependant, la validation ne garantit pas la cohérence générale puisqu'à ce stade, l'utilisateur ne peut avoir à l'esprit des critères de structuration généraux de l'ontologie. Le principal avantage de ces nombreuses validations est qu'elles garantissent la création d'une ébauche d'ontologie de qualité. Cette ébauche peut être reconsidérée lors d'une ultime étape - appelée "normalisation" dans la méthode Terminae [1]- à l'aide de critères liés à l'ontologie.

La génération automatique d'ébauches est plus efficace. La validation de l'ébauche peut sembler plus pertinente puis-

qu'il s'agit de vérifier chaque relation candidate, son existence, ses étiquettes et ses connexions avec les concepts existants. Il est possible d'obtenir une ébauche de ce type grâce à des textes contenant un nombre suffisant d'occurrences de patrons afin d'obtenir un modèle suffisamment cohérent. Cependant, si l'utilisateur n'est pas extrêmement rigoureux lors de l'examen, les résultats peuvent être de mauvaise qualité.

## 2.4 Caméléon : partir de l'expérience passée

La méthode présentée dans cet article repose sur Caméléon, un système qui permet d'enrichir des ontologies avec des relations sémantiques obtenues à l'aide de relations extraites des textes grâce à des patrons. Caméléon offre la possibilité de réutiliser, adapter ou créer des patrons pour des textes annotés syntaxiquement. Il permet ensuite de projeter les patrons. Les phrases où ont été identifiés les patrons sont présentées à l'utilisateur pour valider les termes et les relations lexicales identifiés. Enfin, il contient un éditeur d'ontologies pour ajouter des relations conceptuelles à l'ontologie originale.

Les qualités de Caméléon pour l'extraction de relations non-taxonomiques à partir de textes ont fait l'objet d'une récente évaluation dont les résultats ont été présentés dans [2] et [3]. Ces articles ont permis de dégager aussi bien les points forts que les limites de l'approche. Le principal avantage de l'approche par patrons réside dans la robustesse des relations découvertes. Appliquer cette méthode à un ensemble de corpus hétérogènes a montré que la méthode présentait l'intérêt de dégager des ressources indépendantes des domaines et des corpus.

- Premièrement, cette expérience a confirmé le coût de la création de patrons et par conséquent le gain que représentait la réutilisation de patrons déjà existants.
- Deuxièmement, lors de la construction d'ontologies, l'un des principaux problèmes vient de la difficulté qu'il y a à identifier les termes candidats à la relation et à les aligner avec des concepts de l'ontologie.
- Troisièmement, si la réutilisation de patrons est pertinente pour l'identification de relations multi-domaines comme l'hyponymie, elle est parfois peu productive selon le genre textuel du corpus sur lequel elle est appliquée. Ainsi, par exemple, seuls des ouvrages didactiques contiennent des patrons définitoires. L'approche par patrons trouve ses limites lors de la création de patrons propres à un domaine où elle se révèle fastidieuse et amoindrie par le très faible nombre de patrons employés.
- Enfin, l'évaluation de la méthode sur des corpus hétérogènes a montré de grandes disparités de résultats d'un texte à l'autre comme d'une relation à l'autre. Il faut néanmoins préciser que ces résultats ne sont pas liés spécifiquement à Caméléon, mais qu'ils sont partagés par tous les systèmes d'extraction de relations basés sur des approches par patrons.

Nous considérons que les relations conceptuelles enrichissent une taxonomie de concepts et les termes liés qui constituent le noyau d'une ontologie et nous proposons deux méthodes complémentaires pour améliorer ce processus :

- Pendant la projection des patrons, le système recherche tous les termes reconnus dans l'ontologie cooccurant dans les phrases et les considère comme des termes potentiellement liés. A partir de ces relations lexicales, les relations conceptuelles potentielles sont proposées. Il est possible de retarder la validation et de ne la faire porter que sur les relations conceptuelles candidates.
- A partir des cooccurrences de termes ou concepts de l'ontologie, le système guide l'apprentissage de patrons en explorant leur contexte. Dans un premier temps, les textes sont enrichis à l'aide d'annotations sémantiques (les concepts correspondant alors aux occurrences de termes). L'apprentissage prend en compte les annotations syntaxiques et sémantiques pour définir des patrons de haut niveau. Cette étape permet de réduire le coût de la création de patrons et d'améliorer l'identification de concepts car le processus utilise des termes et concepts figés pour déterminer la relation.

### 3. CADRE METHODOLOGIQUE

A partir de ces constats, nous avons fait évoluer l'approche initiale de Caméléon d'une part en ajoutant des ressources existantes et d'autre part en restreignant les relations sémantiques. La plupart des approches que nous avons évoquées s'intéressent à l'identification de concepts. Les patrons sont généralement utilisés pour extraire des couples d'unités lexicales pertinentes. L'approche présentée dans cet article s'en distingue puisque nous cherchons à identifier des patrons et non des concepts en relation : par conséquent, nous supposons que les  $X$  et  $Y$  présents dans nos patrons sont identifiés puisqu'ils correspondent à des concepts appartenant déjà à l'ontologie.

De plus, partant du constat que la plupart des systèmes d'extraction de relations visent à rechercher des relations taxonomiques [21], des relations partie-tout [12], des relations hyperonymiques [19], nous avons choisi d'orienter nos recherches vers l'extraction de relations non-taxonomiques (ou relations transversales). D'autres travaux se sont déjà intéressés à ce problème : [11] s'est penché sur les relations causales, [15] a comparé les relations causales entre le français et l'anglais, [20] a plus particulièrement exploité les possibilités offertes par le Web Sémantique, etc. Mais la plupart du temps, ces travaux restreignent leur typologie à un nombre très limité de relations. Nous pensons que l'approche par patrons peut dépasser ces ambitions à condition de prendre en compte des ressources externes. Ce postulat était déjà un des principes fondamentaux dans la mise en oeuvre de Caméléon où la liste des types de relations recherchés demeure ouverte et peut être adaptée à tout nouveau corpus étudié.

Grâce à l'engouement croissant du Web Sémantique pour les ontologies, non seulement un nombre de plus en plus important d'ontologies sont désormais accessibles sur Internet mais en plus il existe des plateformes comme **Watson**<sup>4</sup> ou **Swoogle**<sup>5</sup> qui permettent de consulter et d'utiliser des ontologies et les éléments de ces ontologies. Dans notre approche, nous proposons d'exploiter les relations - ou propriétés- existant dans ces ontologies en les recherchant à l'aide de Watson.

<sup>4</sup>Watson [9], <http://watson.kmi.open.ac.uk/Overview.html>

<sup>5</sup>Swoogle, <http://swoogle.umbc.edu/>, crawl Internet pour rechercher des documents du Web Sémantique écrits en RDF

Pour résumer les différents éléments développés plus haut, les avantages comme les inconvénients des approches par patrons sont désormais bien identifiés : soit les patrons sont construits manuellement et coûteux, soit ils sont obtenus par des approches d'apprentissage et sont alors moins pertinents. Nous pensons donc qu'il est possible de collecter et d'accumuler des patrons afin de les réutiliser, et proposons un outil qui utilise une base de patrons tout en offrant un cadre méthodologique d'enrichissement interactif de la base au fur et à mesure des utilisations.

### 3.1 Ressources initiales

Notre approche prend en entrée : (i) une liste de patrons; (ii) une ontologie légère; (iii) un corpus.

#### 3.1.1 La base de patrons

La première étape méthodologique consiste en la réutilisation de patrons : nous avons adapté les patrons implémentés dans Caméléon<sup>6</sup> et utilisé d'autres travaux comme [4] et [15] pour alimenter une base de patrons où ils ont été organisés selon le type de la relation  $REL$  qu'ils dénotent.

#### 3.1.2 L'ontologie

Notre objectif étant d'identifier les concepts auxquels les couples de termes extraits grâce aux patrons font référence, la qualité de la tâche est fortement dépendante des composants lexicaux de l'ontologie. Nous ne considérons dans notre approche que des ontologies légères composées de deux niveaux sémiotiques. Le niveau lexical ( $L$ ) comprend tous les termes ou étiquettes désignant les concepts ou les relations. Le niveau conceptuel défini dans la structure ( $S$ ) de l'ontologie contient les concepts et les relations sémantiques construites à partir des relations conceptuelles qui les lient. Des ontologies dites "lourdes" pourraient être utilisées mais seules les composantes lexicales et structurelles seront exploitées.

La structure de l'ontologie est un tuple  $S := \{C, R, \leq, \partial_R\}$  où :

- $C, R$  sont des ensembles disjoints contenant les concepts et les relations non-taxonomiques
- $\leq : C \times C$  est un ordre partiel sur  $C$ ; il définit la hiérarchie de concepts
- $\partial_R : R \rightarrow C \times C$  est la signature d'une relation associative ou non taxonomique

Le lexique d'une ontologie légère est un tuple  $L : \{L^C, L^R, F, G\}$

- $L^C, L^R$  sont deux ensembles disjoints contenant les étiquettes (ou les termes) associées aux concepts et relations
- $F, G$  sont deux relations de référence, elles permettent d'accéder aux étiquettes ou aux termes associés respectivement aux concepts ou aux relations.

Il faut préciser qu'un concept peut être défini par plusieurs termes et un terme, s'il est ambigu, peut renvoyer à plusieurs concepts.

<sup>6</sup>Caméléon ayant été conçu pour le français, nous les avons adaptés pour l'anglais

**Table 1: Exemple de relations transversales et des patrons associés**

Relation	exemple de patrons
Communicat.	X instruct (DT)? Y
Build	X (MD)? build Y
Build	X (MD)? assemble (DT)? Y
Own	X belonging to Y

Dans notre approche, une ontologie légère constitue l'entrée du système. L'ontologie est constituée d'au moins une hiérarchie de concepts définis par un lexique  $L^0 := \{L^C, F\}$  et une structure  $S^0 := \{C, \leq\}$ . L'objectif de notre approche est d'enrichir d'une part  $S^0$  en identifiant  $R$  et  $\partial_R$  à partir de textes et  $L^0$  en définissant  $G$ .

### 3.1.3 Corpus

Le choix du corpus est un facteur primordial pour le processus d'enrichissement de l'ontologie. Le corpus doit décrire les éléments de connaissance qui seront intégrés à l'ontologie. Dans notre approche, le corpus est extrait de corpus existants et il revient aux experts de vérifier qu'il couvre le domaine entier pour une période représentative.

## 3.2 Un cas d'utilisation : le projet WebContent

Notre travail s'inscrit dans le contexte de WebContent<sup>7</sup>. WebContent est un projet qui vise à construire un environnement logiciel pour l'exploration et l'utilisation des technologies du Web Sémantique pour des applications comme, par exemple, la veille technologique. L'un des domaines d'application est celui de la veille économique dans l'industrie aéronautique. Par conséquent, toutes les ressources citées dans les exemples de cet article sont celles que nous utilisons dans le cadre du projet pour lequel cette application a été développée. Nous présentons ici les ressources utilisées dans ce cadre : la base de patrons, l'ontologie et le corpus.

### 3.2.1 La base de patrons WebContent

La base de patrons présentée plus haut a été enrichie manuellement avec des relations non-taxonomiques tirées de l'analyse du corpus. Le tableau 1 présente quelques exemples de ces relations.

### 3.2.2 L'ontologie Webcontent

Dans le cadre du projet, l'ontologie à enrichir est une hiérarchie de concepts du domaine aéronautique construite manuellement. La figure 1 en présente un extrait.

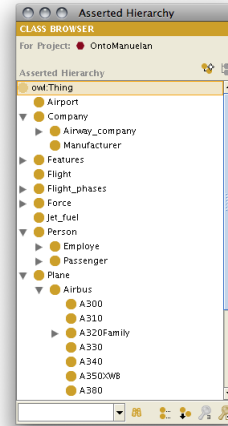
Chaque concept de l'ontologie est associé à une liste d'étiquettes.

### 3.2.3 Le corpus Webcontent

L'extraction de relations entre concepts se base sur l'exploitation de deux types de textes complémentaires. Le corpus, en langue anglaise, est constitué de 513 courtes dépêches de l'AFP<sup>8</sup> et de 14 articles de la *Wikipedia* portant sur des avions particuliers comme l'A300 ou le B737. Cette

<sup>7</sup>(<http://www.webcontent.fr>)

<sup>8</sup>Agence France-Presse



**Figure 1: Extrait de l'ontologie utilisée pour WebContent**

double nature place nos travaux à la fois dans le domaine de l'économie avec les dépêches de l'AFP et dans celui de l'aéronautique avec les articles de la *Wikipedia*.

## 4. ALGORITHME ET IMPLEMENTATION

Le cadre méthodologique propose d'une part d'utiliser et d'enrichir la base de patrons et d'autre part d'enrichir l'ontologie existante à l'aide de nouvelles relations. Dans cette partie, nous décrivons tout d'abord le système en présentant l'algorithme général avant de présenter au fur et à mesure les différentes étapes qui nous permettent d'atteindre ces deux objectifs.

### 4.1 L'algorithme

La figure 2 offre un aperçu du système.

La première partie de l'algorithme porte sur le traitement (des relations et de la découverte de patrons) alors que la seconde décrit les validations de l'utilisateur. La phase de validation est effectuée dans un second temps afin de proposer à l'utilisateur tous les couples extraits pour une relation. Nous pensons que d'une part la tâche en est facilitée et que d'autre part, l'utilisateur a alors une meilleure vue globale des validations à effectuer.

#### Traitement.

Pour toute paire de concepts de l'ontologie  $(c_i, c_j)$ , nous recherchons toutes les phrases qui contiennent  $t_i, t_j$  où  $t_i, t_j$  appartiennent respectivement à l'ensemble de étiquettes associées aux concepts  $c_i, c_j$ . Si l'un des patrons  $P^{REL}_i$  de la base de patrons peut être aligné avec la phrase  $s$ , la relation  $REL$  extraite par le patron est stockée ainsi que le couple de concepts  $REL^c(c_i, c_j)$  et  $s$ . Dans le cas contraire, nous recherchons une relation qui aurait pu être déjà définie par une autre ontologie. Si une telle relation est identifiée, nous stockons la nouvelle relation  $REL_{new}$ , le couple de termes  $REL_{new}^c(c_i, c_j)$  et  $s$ .

#### Proposition et validation de patrons.

Pour toute nouvelle relation  $REL_{new}$  identifiée grâce à des ontologies existantes, nous proposons à l'utilisateur la

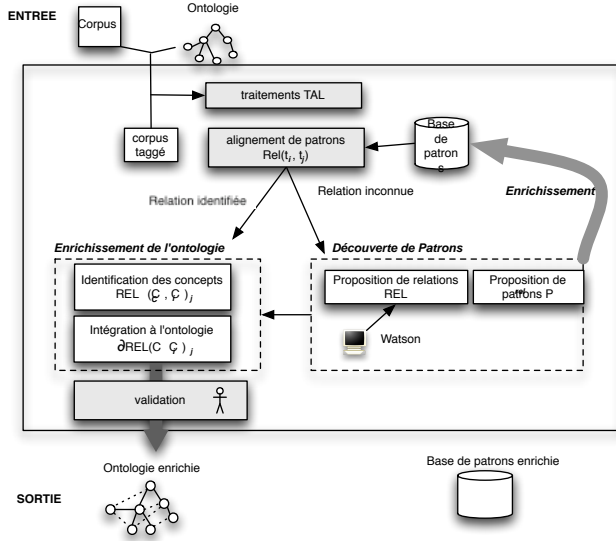


Figure 2: Aperçu du système

relation ainsi que les couples de concepts  $REL_{new}^c(c_i, c_j)$ . L'utilisateur doit alors valider la pertinence de la relation. S'il la valide, le système lui propose alors un ensemble de patrons qui ont été générés en fonction des phrases dans lesquelles les couples ont été identifiés. L'utilisateur peut alors valider un ou plusieurs patrons qui seront versés à la base de patrons.

### Validation de relation.

Pour toute relation détectée ( $REL$  and  $REL_{new}$ ), le système propose à l'utilisateur les étiquettes associées à la relation ainsi que les couples de concepts associés. Celui-ci peut alors décider d'ajouter la relation à l'ontologie, plusieurs possibilités s'offrent à lui :

- soit il ajoute la relation entre  $c_i$  et  $c_j$ ;
- soit il ajoute la relation entre un ancêtre de  $c_i$  et un ancêtre de  $c_j$ ;
- soit il ajoute la relation entre un concept lié à  $c_i$  dans l'ontologie et un concept lié à  $c_j$  ;
- soit il rejette la relation associée au couple.

Afin de faciliter les choix et de guider la décision, l'utilisateur peut accéder aux phrases dans lesquelles les couples ont été trouvés dans le corpus et au contexte de chaque concept de l'ontologie (les étiquettes de concepts et les concepts associés).

L'implémentation des différentes étapes est détaillée dans la partie suivante.

## 4.2 Patrons pour l'enrichissement de l'ontologie

### Pré-traitement.

Table 2: Exemple de patron

```
Rule:BuiltModal
// modality
Priority: 50
(
{Lookup.majorType==OntoWC}{Lookup.majorType=
=ModalConstruction}{Lookup.majorType==BuiltVerb}({Token.
category==DT})? ({Token.category==JJ})? ({Token.category==
CD})? {Lookup.majorType==OntoWC
}):BuildModal -->
:BuildModal.Build={kind="Build", rule="BuildModal"}
```

Table 3: Exemples de relations transversales entre concepts.

	Relation	Concept c1	Concept c2	Exemples
1	Comm.	Pilote	Controller	12a
2	Build	Aircraft Manufacturer	Plant	12b
3	Build	Factory	Plane	12c
4	Own	Airway Company	Plane	12d

Notre système a été implémenté avec GATE, une plateforme qui propose une architecture logicielle d'ingénierie linguistique pour développer et distribuer des traitements de TAL. Cette plate-forme présente l'avantage d'utiliser des techniques d'automates à états finis pour extraire ou annoter de l'information. Chaque texte est tokenisé, segmenté en phrases, taggé avec TreeTagger et annoté avec des informations tirées de l'ontologie.

### Projection de patrons.

Une fois que le texte a été soumis à la chaîne de traitement, les relations sont identifiées grâce à des règles exprimées avec JAPE<sup>9</sup>. JAPE fournit des transducteurs à états finis qui peuvent s'appliquer sur des annotations grâce à des expressions régulières. Pour chaque paire de termes qui correspondent à des étiquettes de concepts, GATE est utilisé pour repérer des relations potentielles. Chaque type de relation est associé à une grammaire JAPE, c'est-à-dire un ensemble de règles(sic) déclenchées en cascade de manière à appliquer une cascade de transducteurs à états finis sur les annotations définies au début de chaque grammaire. Les résultats de TreeTager, lemmes et annotations morpho-syntaxiques, sont utilisés lors de la rédaction des patrons. L'exemple 2 présente une règle qui identifie la relation "BUILT" pour le couple "BUILT<sup>t</sup>(plant, aircraft)" à partir de phrases comme "by 2011 the plant will have the capacity to assemble four aircraft per month". Cette règle implémente une des relations présentées dans le tableau 1.

Le tableau 3 présente des exemples de relations identifiées entre des concepts grâce à des patrons, une fois le processus de désambiguïsation mené à son terme. Nous présentons ensuite en 12 les phrases illustrant les relations proposées.

- (12)
- "the controller instructed the pilot"
  - "Airbus is expected to begin to build plant"
  - "the plant will have the capacity to assemble four aircrafts per month"

<sup>9</sup>Java Annotation Patterns Engine

- d. “the Airbus A320 belonging to Armenian Airlines”

### Identification des concepts.

Afin d’identifier les concepts impliqués dans le couple  $REL^t$  ( $t1, t2$ ), des *gazetteers* annotent les mots du texte en fonction des étiquettes de concepts qui ont été définies dans le lexique de l’ontologie. Pour chaque concept  $c$ , un *gazetteer* est construit en fonction de  $F(c)$ . Par exemple, le *gazetteer* du concept *plane* sera composé des termes *aircraft*, *plane*, *aeroplane*, *airplane*. Cette étape est cruciale car elle détermine les  $X$  et les  $Y$  utilisés dans les patrons et l’ancrage de la proposition de patrons. Pour être sûrs que seuls les concepts les plus spécifiques sont identifiés, les *gazetteers* sont implémentés afin de favoriser la détection des étiquettes correspondant aux phrases les plus longues possibles (i.e. composés du plus grand nombre de mots). Par exemple, dans la phrase “The heads report to the vice president”, les étiquettes de deux concepts sont présentes. On favorisera donc l’étiquette “vice president” plutôt que “president”.

Comme indiqué dans la sous-section précédente, une étiquette peut faire référence à plusieurs concepts. Il est donc nécessaire de désambigüiser les termes des couples : quand l’étiquette présente dans un couple peut se référer à un seul concept, le concept correspondant est identifié. Par exemple, dans le couple  $BUILT^t$  (*plant*, *aircraft*), le terme “aircraft” n’est pas ambigu puisqu’il fait référence au concept “plane”. Cependant, quand l’étiquette correspond à plusieurs concepts, l’utilisateur doit identifier le concept pertinent pendant l’étape de validation.

### Enrichissement de l’ontologie avec les nouvelles relations.

Une fois identifiés les concepts de la relation, nous proposons d’enrichir l’ontologie avec les relations extraites pour les concepts les plus généraux. Ainsi, par exemple, le système extrait la relation “*BUILD*” pour le couple de concepts “*factory*”, “*plane*” (comme présenté plus haut) et le couple “*assemblyLine*”, “*plane*”. Comme le concept “*assemblyLine*” est défini dans l’ontologie comme un sous-concept de “*factory*”, le système propose uniquement d’ajouter dans l’ontologie la relation entre “*factory*” et “*plane*”. Avant chaque ajout de relation à l’ontologie, il est demandé à l’utilisateur de valider la proposition.

## 4.3 Création et validation de nouveaux patrons

Parfois, il est impossible d’identifier une relation entre deux concepts qui co-occurrent dans une même phrase, il faut alors traiter le problème des paires orphelines de concepts. Dans ce cas, nous partons du principe que des ontologies déjà existantes peuvent contenir une ou plusieurs relations déjà identifiées pour lier ces deux concepts. Notre stratégie est de soumettre la relation ainsi que des patrons associés à l’utilisateur.

### Proposition de relations.

Tout d’abord, les paires orphelines sont extraites des textes afin d’être soumises à *Watson*. Cette étape permet d’identifier toutes les relations disponibles entre les deux concepts dans les bases de données répertoriées. Ainsi, par exemple, aucun patron de notre base n’a pu identifier la relation entre les concepts “*Person*” et “*Company*”, alors que le corpus montre

qu’ils apparaissent très fréquemment dans le même contexte comme l’illustrent les exemples ci-dessous.

1. “Airbus Names New Chief Managers for A380, A320 Programs”
2. “Aircraft maker Airbus has named Mario Heinen senior vice president and chief manager of the A380 aircraft program”
3. “Laurence Barron, the vice senior president of Airbus”

Interrogé, *Watson* propose ici la relation “*Work\_for*” qui va pouvoir être soumise à l’utilisateur.

### Proposition de patrons.

En parallèle, le système suggère un ensemble de patrons qui sont susceptibles d’être associés à la nouvelle relation. Ces patrons combinent des éléments hétérogènes puisqu’ils exploitent toutes les ressources disponibles dans *Gate* pour la création de patrons. Ainsi, par exemple, le lemme “name” peut être traité comme une catégorie syntaxique (un verbe), une catégorie sémantique (un verbe lié au choix) ou comme une simple chaîne de caractères<sup>10</sup>. Dans le cas présent, parmi les seize synonymes proposés, il peut sélectionner {*select*; *pick*; *decide on*; *nominate*; *designate*} et ignorer {*give a name*; *baptize*}. Pour chaque élément de la phrase, le système propose soit la catégorie grammaticale, soit la catégorie syntaxique, soit le lemme. Il en résulte une liste de combinaisons mêlant des éléments hétérogènes mais qui correspondent toutes à la réalisation linguistique.

Pour des phrases comme celles des exemples 1 et 2, notre système peut proposer plusieurs combinaisons illustrées par les exemples 1a–1d. Si les Entités Nommées ou toute autre connaissance sémantique supplémentaire sont disponibles dans un *gazetteer* particulier, elles pourront être prises en compte comme dans l’exemple 1e.

1. (a) X name Y  
(b) X name (NP)? (NP)? Y  
(c) X (VB=choose) Y  
(d) X (MD)? (VB=choose) Y  
(e) X name NE\_Person Y (with NE\_Person semantic class of Named Person)

Contrairement aux approches par apprentissage, la méthode que nous proposons permet à l’utilisateur de contrôler intégralement le processus de construction des patrons. Ceci permet de garantir la signification sémantique de ces patrons ainsi que la pertinence des relations conceptuelles identifiées. De plus, le processus de validation manuelle des patrons est accéléré par le système d’acquisition interactif.

## 5. CONCLUSION

L’extraction de relations à l’aide d’approches par patrons apparaît comme un moyen efficace d’enrichir une ontologie à condition que les patrons soient collectés, stockés, adaptés, réutilisés et acquis semi-automatiquement et que les termes liés soient identifiés dans des phrases. Pour mener à bien ce processus, nous proposons dans le prolongement de

<sup>10</sup>Si la catégorie sémantique est choisie, l’utilisateur doit créer un *gazetteer* contenant les instances des relations proposées dans le thesaurus



Caméléon un outil d'extraction de relations qui intègre certains principes d'apprentissage. Ce nouvel outil permet (i) l'identification de termes dans des phrases qui correspondent à des patrons existants; (ii) une aide à la création de nouveaux patrons basées sur des suggestions automatiques mais qui ne relève pas de l'apprentissage.

Cet outil est actuellement en cours d'implémentation et deux aspects restent encore à améliorer. Le principal obstacle à surmonter est lié à la définition de la fenêtre contextuelle à l'intérieur de laquelle les instances de concepts sont identifiées. En effet, une fenêtre trop large entraîne une explosion combinatoire incontrôlable et une fenêtre trop réduite se révèle le plus souvent inefficace pour trouver de nouvelles relations. La seconde difficulté concerne la présentation des nouveaux patrons automatiquement acquis. Actuellement, toutes les combinaisons des éléments linguistiques sont proposées à l'utilisateur. Il apparaît indispensable de réduire le nombre de ces patrons quand ils sont trop nombreux pour une seule et même relation en appliquant des techniques d'apprentissage ou de clustering. Si les techniques d'apprentissage se révèlent infructueuses, il nous semble possible d'ordonner les patrons grâce à des mesures de pertinence. Ces mesures pourraient être établies en comparant la structure des nouveaux patrons à celle des patrons déjà présents dans la base.

## 6. REFERENCES

- [1] N. Aussenac-Gilles, S. Despres, and S. Szulman. The terminae method and platform for ontology engineering from texts. In P. Buitelaar and P. Cimiano, editors, *Bridging the Gap between Text and Knowledge Selected Contributions to Ontology Learning and Population from Text*, pages 199–223. IOS Press, 2008.
- [2] N. Aussenac-Gilles and M.-P. Jacques. Designing and evaluating patterns for ontology enrichment from texts. In S. Verlag, editor, *EKAW 2006, 15th International Conference on Knowledge Engineering and Knowledge Management, Prague, oct. 2006*, pages 158 – 165, Prague, 2006.
- [3] N. Aussenac-Gilles and M.-P. Jacques. Designing and evaluating patterns for relation acquisition from texts with caméléon. *Terminology, special issue on Pattern-Based approaches to Semantic Relations*, 14(1):45 – 73, 2008.
- [4] C. Barrière and A. Agbago. Terminoweb: A software environment for term study in rich contexts. In *International Conference on Terminology, Standardisation and Technology Transfer (TSTT 2006)*, Beijing, Chine, 25 et 26 août 2006.
- [5] C. Brewster. Techniques for automated taxonomy building: Towards ontologies for knowledge management. In *Proc. of the 5th Annual CLUK Research Colloquium, Leeds, 2002*.
- [6] C. Brewster, F. Ciravegna, and Y. Wilks. User-centred ontology learning for knowledge management. In B. Andersson, M. Bergholtz, and P. Johannesson, editors, *NLDB*, volume 2553 of *Lecture Notes in Computer Science*, pages 203–207. Springer, 2002.
- [7] M. Ciaramita, A. Gangemi, E. Ratsch, J. Saric, and I. Rojas. Unsupervised learning of semantic relations between concepts of a molecular biology ontology. In *IJCAI*, pages 659–664, 2005.
- [8] P. Cimiano. *Ontology Learning and Population from Text. Algorithms, evaluation and applications*. Springer, Berlin, 2007.
- [9] M. d'Aquin, C. Baldassarre, L. Gridinoc, S. Angeletou, M. Sabou, and E. Motta. Watson: A gateway for next generation semantic web applications. In *ISWC, Poster session of the International Semantic Web Conference, 2007*.
- [10] M. Finkelstein-Landau and E. Mori. Extracting semantic relationships between terms: Supervised vs. unsupervised methods. In *Proc. of the Int. Workshop on Ontological Engineering on the Global Information Infrastructure, Dagstuhl Castle, Germany*, pages 71–80, 1999.
- [11] R. Girju and D. Moldovan. Text mining for causal relations. In *FLAIRS 2002*, pages 360 – 364, Pensacola Beach, Florida, 14-16 mai 2002 2002.
- [12] R. G. Girju, A. Badulescu, and D. Moldovan. Learning semantic constraints for the automatic discovery of part-whole relations. In *HLT-2003*, 2003.
- [13] M. A. Hearst. Automatic acquisition of hyponyms from large text corpora. In A. Zampolli, editor, *Computational Linguistics (CoLing'1992)*, pages 539 – 545, Nantes, France, 1992.
- [14] M.-P. Jacques and N. Aussenac-Gilles. Variabilité des performances des outils de tal et genre textuel, cas des patrons lexico-syntaxiques. *Traitement Automatique des Langues (TAL), NumÉro spécial Non Thématique*, 47(1):11 – 32, 2006.
- [15] E. Marshman. Towards strategies for processing relationships between multiple relation participants in knowledge patterns: An analysis in english and french. In J. Benjamins, editor, *Terminology*, volume 13:1, pages 1 – 34, 2007.
- [16] A. Schutz and P. Buitelaar. Relext: A tool for relation extraction from text in ontology extension. In Y. G. et al., editor, *ISWC 2005, LNCS 3729*, pages 593 – 606, 2005.
- [17] A. Schutz and P. Buitelaar. Relext: A tool for relation extraction in ontology extension. In *Proceedings of the 4th International Semantic Web Conference*, Galway, Ireland, November 2005.
- [18] P. Séguéla. *Construction de modèles de connaissances par analyse linguistique de relations lexicales dans les documents techniques*. PhD thesis, Thèse de Doctorat en Informatique, Toulouse 3 Paul Sabatier, 2001.
- [19] R. Snow, D. Jurafsky, and A. Y. Ng. Learning syntactic patterns for automatic hypernym discovery. In *Advances in Neural Information Processing Systems 17*, 2005.
- [20] M. Tchalakova, B. Popov, and M. Yankova. Methodology for bootstrapping relation extraction for the semantic web. In *AIMSA 2006*, Varna, Bulgaria, 13-15 sept. 2006.
- [21] P. Velardi, A. Cucchiarelli, and M. Petit. A taxonomy learning method and its application to characterize a scientific web community. *IEEE Transactions on Knowledge and Data Engineering*, 19(2):180 – 191, 2007.
- [22] D. Widdows. Unsupervised methods for developing taxonomies by combining syntactic and statistical informations. In *HLT/NAACL*, pages 276 –283, 2003.