

# Génération automatique des représentations ontologiques<sup>1</sup>

Johannes Heinecke

France Télécom, Division R&D, TECH/EASY/Langues Naturelles  
johannes.heinecke@francetelecom.com

## Résumé

Depuis la conception du Web sémantique une tâche importante se pose au niveau de traitement automatique du langage : rendre accessible le contenu existant du Web dit classique aux traitements et raisonnements ontologiques. Comme la plupart du contenu est composé de textes, on a besoin de générer des représentations ontologiques de ces informations textuelles. Dans notre article nous proposons une méthode afin d'automatiser cette traduction en utilisant des ontologies et une analyse syntaxico-sémantique profonde.

**Mots-clés :** analyse syntaxico-sémantique, ontologies, représentations ontologiques, Web sémantique.

## Abstract

Since the conception of the semantic Web an important challenge faces natural language processing: i.e. making the existing textual contents of the « classic » Web available for ontology based applications, reasoning and finally for the semantic Web. Since the majority of Web sites contain textual information, the creation of ontological representations of these will be useful if not necessary. In our article we propose a method to automate this task by using domain ontologies and a deep syntactic-semantic analysis.

**Keywords:** syntactic and semantic analysis, ontologies, ontological representations, semantic Web.

## 1. Introduction

Dans cet article nous décrivons une approche de TALN choisie pour le projet aceMedia<sup>2</sup>. Malgré le fait que aceMedia n'est pas un projet du Web sémantique, les idées étudiées dans le cadre du traitement automatique du langage réalisées peuvent être généralisées pour le Web sémantique. L'approche décrite dans ce papier a été mise en œuvre et implementée pour aceMedia. Les ontologies de domaines d'aceMedia contiennent environ 300 de classes et relations.

Comme déjà décrit pour le Web sémantique par Buitelaar & Declerck, 2003, le lien entre les langues naturelles et la formalisation du monde sous forme d'ontologies devrait être le plus automatisé possible. En plus, cette automatisation devrait être complètement indépendante d'une langue spécifique ; il faut donc recourir à une analyse sémantique qui utilise des concepts d'un thésaurus qui est soit multilingue soit indépendant d'une langue. L'indépendance à une langue

---

<sup>1</sup> Ce travail a été soutenu par la Commission Européenne sous le contract FP6-001765 aceMedia (<http://www.acemedia.org/>).

<sup>2</sup> Le projet aceMedia a pour but d'organiser des contenus multimedia et d'y accéder plus facilement. Des ontologies sont utilisées pour les métadonnées associées. Le TAL joue un rôle pour la création des métadonnées à partir des annotations textuelles des contenus et pour analyser les requêtes d'utilisateurs (<http://www.acemedia.org/>).

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:tennis="http://www.acemedia.org/ontos/tennis#"
  <tennis:Player rdf:about="#V5">
    <tennis:hasName>Federer</tennis:hasName>
    <tennis:hasFirstName>André</tennis:hasFirstName>
  </tennis:Player>
  <tennis:Trophy rdf:about="#V4">
    <tennis:isWonBy rdf:resource="#V5"/>
  </tennis:Trophy>
</rdf:RDF>

```

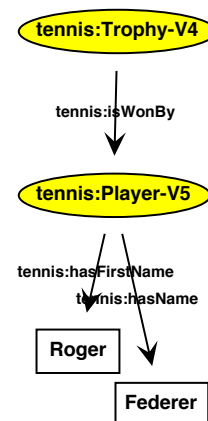


Figure 1. Représentation ontologique (RDF et sa visualisation graphique)

spécifique permet aux utilisateurs d'un tel système de l'utiliser dans une langue autre que celle des contenus indexés (cf. le système multilingue MKBEEM dans Heinecke & Toumani, 2003 et Lehtola *et al.*, 2003).

Nous arrivons donc à trois besoins pour le traitement automatique du langage pour le Web sémantique. il faut pouvoir :

- annoter automatiquement des contenus (textuels) avec des expressions basées sur des ontologies (de domaine) ;
- les créateurs des pages Web ne devraient pas être obligés de fournir des annotations ontologiques de leurs pages (en plus, ils ne possèdent pas toujours les compétences nécessaires) ;
- la représentation ontologique devrait aller plus loin qu'une simple liste de mots-clé, mais aussi représenter les relations entre classes (ontologiques)<sup>3</sup>.

Nous proposons trois tâches afin de satisfaire ces besoins. Une solution pour le premier besoin est présentée en détail dans Heinecke, 2006 ; nous décrivons les deux autres dans la suite de cet article :

- aligner des ontologies de domaines sur les données linguistiques (c'est-à-dire sur un thésaurus sémantique) ;
- transformer des énoncés en langue naturelle en une représentation ontologique (par exemple RDF/RDFS ou OWL<sup>4</sup>). L'exemple de la figure 1 montre le résultat souhaité lorsqu'on traite un énoncé nominal comme *le trophée a été gagné par Roger Federer* ;
- générer des requêtes ontologiques en utilisant un langage comme SPARQL<sup>5</sup>.

Les besoins d'une telle approche ne sont pas négligeables : nous possédons heureusement des lexiques complets pour le français (environ 85 000 lemmes et leurs formes fléchies) et l'anglais (75 000 lemmes) ainsi que des ensembles de règles syntaxiques (1 000 règles pour le français et 400 pour l'anglais) qui permettent de faire une analyse en dépendance (figure 2).

<sup>3</sup> Nous employons les termes *classe* et *relation* seulement dans le contexte d'ontologies. Par contre le terme *concept* (*sémantique*) désigne exclusivement le sens des mots ou locutions.

<sup>4</sup> Resource Description Framework/RDF Schema <http://www.w3.org/RDF/> et <http://www.w3.org/TR/rdf-schema/>, OWL Web Ontology Language, <http://www.w3.org/TR/owl-features/>. D'autres langages sont par exemple DARPA Agent Markup Language (DAML), <http://www.daml.org/>, Ontology Inference Layer (OIL), <http://www.ontoknowledge.org/oil/> et CARIN (Goasdoué *et al.*, 2000).

<sup>5</sup> <http://www.w3.org/TR/rdf-sparql-query/>, d'autres langages de requêtes sont par exemple RDQL <http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/> ou encore FLOGIC (Kifer *et al.*, 1995).

Pour la création des représentations sémantiques, indépendantes de la langue, nous avons un thésaurus multilingue qui est lié avec les entrées lexicales du français et de l'anglais<sup>6</sup> et qui contient pour l'instant environ 100 000 concepts sémantiques (Guimier de Neef *et al.*, 2002)<sup>7</sup>. Finalement on a besoins des règles sémantiques afin de construire les représentations sémantiques à partir des arbres de dépendance.

### 1.1. Autres travaux du domaine

Le besoin de transformer des textes en langue naturelle vers des représentations ontologiques a été déjà reconnu par d'autres auteurs, notamment dans le cadre de l'annotation sémantique pour le Web sémantique. L'annotation manuelle du contenu textuel n'est pas une solution à retenir, car il y a trop de textes à annoter, sans parler du risque d'erreurs. Il faut aussi prendre en compte que les ontologies évoluent. Donc une annotation manuelle serait invalidée après chaque modification incompatible avec l'annotation actuelle. Par contre, bien évidemment même une annotation automatique peut être erronée (Erdmann *et al.*, 2000). Plusieurs publications récentes proposent des travaux dans cette direction et donnent aussi un état de l'art détaillé (Dill *et al.*, 2003 ; Handschuh & Staab, 2003 ; Buitelaar & Ramaka, 2005).

Dill *et al.*, 2003 proposent une approche intéressante qui est bien au-delà de la détection de mots-clé qui est néanmoins basée sur les approches statistiques : après une phase de détection des concepts ils appliquent un apprentissage afin d'annoter les pages Web. Leur algorithme permet ensuite de désambiguïser si nécessaire, spécialement pour les noms propres (des personnes).

Buitelaar & Declerck, 2003 vont plus dans notre direction en voyant la nécessité d'établir des relations (ontologiques) entre les concepts détectés. Leur conclusion propose également une analyse en dépendance. Malgré les différences de ces approches, le but est toujours de « comprendre » ou au moins de simuler la compréhension de texte et de formaliser pour des applications basées sur ontologies. Cette compréhension est aussi utile pour la génération (Fröhlich & Riet, 1998).

## 2. Mécanisme

### 2.1. Mapping entre ontologies et données linguistiques sémantiques

Afin de produire les représentations ontologiques en utilisant les classes et relations d'une ontologie ou plusieurs ontologies un pas préparatoire est nécessaire qui lie les entrées de notre thésaurus sémantique avec les éléments ontologiques. Ce mapping devrait être automatisé le plus possible.

Nous avons exploité les noms des classes et relations ontologiques ainsi que leur voisinage taxonomique (sous-classes, sous-relations, etc.). Dans le cas où le nom d'une classe n'est pas un mot ou une phrase (anglaise<sup>8</sup>) nous l'avons remplacé par un expression anglaise.

Ensuite nous avons analysé chaque nom de classe ontologique avec les noms de ces sous-classes immédiates. L'analyseur prend en compte le contexte sémantique (en utilisant le thésaurus lié

---

<sup>6</sup> Les liens avec les autres langues pour lesquelles nous avons des lexiques (allemand, arabe, chinois et espagnol) sont établis seulement partiellement ; les travaux pour les compléter sont en cours.

<sup>7</sup> Le site <http://www.pagesjaunes.fr> utilise nos lexiques français et anglais ainsi que le thésaurus cités ci-dessus.

<sup>8</sup> Toutes nos ontologies utilisent des termes anglais.

avec le lexique) :  $\langle tennis:Player \rangle^9$  *player*,  $\langle tennis:Player \rangle$  *tennis-player*,  $\langle tennis:Judge \rangle$  *judge*,  $\langle tennis:Coach \rangle$  *coach*,  $\langle tennis:Umpire \rangle$  *umpire*,  $\langle tennis:Spectator \rangle$  *spectator*,  $\langle tennis:Assistant \rangle$  *assistant*. On obtient une liste des prédicats pour chaque classe ontologique :

```
tennis:Player
  sport_activity.sport(agent); sport_activity.tennis(agent, situation); ...
...
```

D'une façon similaire nous créons des phrases à partir des relations ontologiques en utilisant leurs classes en position domaine et portée (*range*). Par exemple l'analyse de *the*  $\langle tennis:Player \rangle$  *tennis-player*  $\langle tennis:isWonBy \rangle$  *wins the*  $\langle tennis:Final \rangle$  *final* nous rend le prédicat du thésaurus qui correspond le mieux à la relation ontologique *tennis:isWonBy* :

```
tennis:isWonBy
  relational_event.gagner(situation, theme, experiercer)
```

Finalement nous créons des règles de transformation sémantique pour traiter les « noms de classes ontologiques complexes ». C'est-à-dire les classes qui ont une locution non-lexicalisée comme nom, comme *tennis:ExhibitionMatch*. Le fait que le terme *exhibition match* n'apparaît ni dans le lexique, ni dans le thésaurus nous oblige à créer une règle qui remplacera les prédicats pour *exhibition* et *match* par un prédicat composé *composed\_ExhibitionMatch* et qui est lié avec la classe ontologique.

Pour pouvoir exploiter la taxonomie de l'ontologie « mappée » sur nos données sémantiques, elle est rajoutée à la taxonomie de notre thésaurus.

## 2.2. Analyse syntaxico-sémantique

On peut imaginer plusieurs façons de générer des représentations ontologiques à partir des énoncés ou même des textes en langue naturelle. Une façon simple serait l'extraction des mots-clé. Par contre, cette méthode a deux inconvénients : son incapacité à rendre un service multilingue et la perte des relations entre actants (c'est-à-dire des objets de la réalité) qui sont normalement exprimées par la structure syntaxique. C'est une des raisons pour lesquelles nous avons choisi une approche d'analyse profonde. Elle nous permet de lier seulement les données sémantiques avec les classes et relations ontologiques.

Pour l'instant on passe par trois étapes en utilisant la boîte d'outils linguistiques Tilt de France Télécom. D'abord nous effectuons une analyse syntaxique en appliquant une grammaire de dépendance (Tesnière, 1959). Par exemple un énoncé comme *le trophée a été gagné par Roger Federer* aboutit à un arbre de dépendance (figure 2).

Outre l'avantage d'être capable d'exploiter les relations syntaxiques dans la deuxième étape, notre analyseur syntaxique est robuste. Dans le cas d'un mot inconnu (par exemple à cause d'une erreur typographique) l'analyseur possède un mécanisme puissant pour corriger le mot inconnu en fonction du lexique, de la phonétique du mot trouvé et aussi en fonction de sa position syntaxique. Si malgré toutes les corrections possibles, l'analyseur n'arrive pas à établir

<sup>9</sup> Les classes entre  $\langle \rangle$  sont ignorées par l'analyseur syntaxique. Elles sont néanmoins nécessaires pour pouvoir associer les résultats d'analyse linguistique avec la classe donnée. Nous abrégons l'espace de nom [http://www.acemedia.org/ontos/tennis#par tennis:](http://www.acemedia.org/ontos/tennis#par_tennis:)

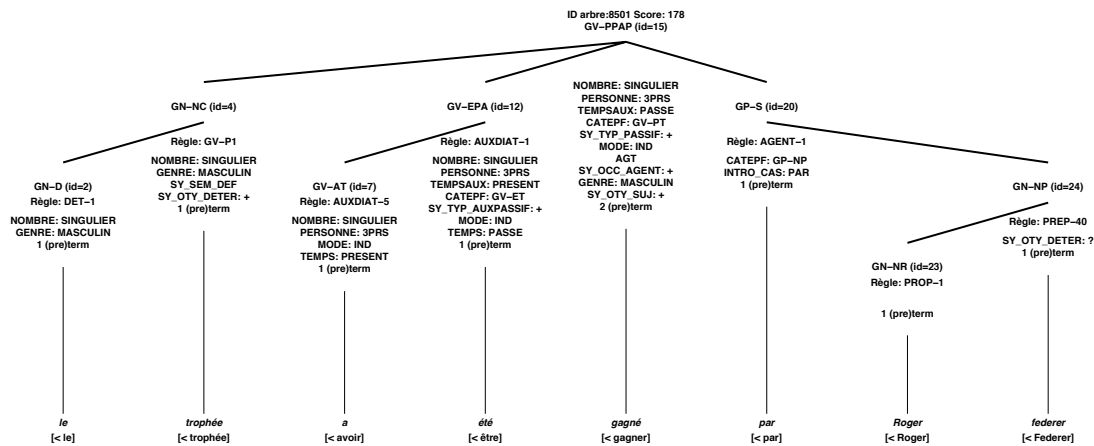


Figure 2. Arbre de dépendance (simplifié)

```

relational_event.gagner(situation=u12748, theme=x12749, experiercer=x12750)
NP~Player(class=x12750, name=x12754, firstname=x12755)
name~Federer(ref=x12754)
firstname~Roger(ref=x12755)
entité.trophée(ref=x12749)
def.artDef(detd=x12749)

```

Figure 3. Graphe sémantique

une relation syntaxique entre deux parties de la phrase, il produit une solution avec des arbres partiels. Dans le pire des cas on obtient donc un « arbre » syntaxique partiel pour chaque mot (chaque arbre contiendrait un seul mot).

À partir des arbres syntaxiques, on crée des représentations sémantiques. Ces représentations ou graphes sémantiques sont construits à partir des prédicats (du thésaurus) associés à chaque terminal et des règles de construction qui lient les prédicats en fonction de leurs arguments et des relations syntaxiques liant les terminaux. En outre les traits morpho-syntaxiques permettent de faire émerger des prédicats supplémentaires (nombre, temps sémantiques, etc.).

Les représentations sont influencées par les travaux menés par Kamp & Reyle, 1993. Elles contiennent des prédicats<sup>10</sup> qui sont complètement indépendants de la langue analysée. C'est-à-dire que deux énoncés avec un contenu identique dans deux langues résultent en des représentations sémantiques identiques. C'est ainsi le cas pour la diathèse (voix active *vs.* passive), les temps sémantiques, etc. Pour notre tâche nous n'avons pas besoin des différences exprimées par ces phénomènes linguistiques. En conséquence ils n'apparaissent pas dans la représentation sémantique. La représentation sémantique obtenue se trouve dans les figures 3 et 4 sous forme d'un graphe sémantique orienté.

Une fois arrivé à une telle représentation sémantique, qui ignore les informations jugées peu importantes<sup>11</sup>, on peut construire une représentation ontologique qui n'utilise que le vocabulaire ontologique (classes et relations). Cette transformation est faite grâce à des règles créées lors de l'alignement entre les données sémantiques et les ontologies (Heinecke, 2006) :

<sup>10</sup> Nous définissons le terme « prédicat » comme le concept sémantique du thésaurus plus ses arguments. Note : les prédicats associés aux terminaux de l'arbre syntaxique ne sont pas affichés dans figure 2.

<sup>11</sup> Actuellement les graphes sémantiques ne contiennent aucune information (prédicat) sur le temps sémantique, le focus, le thème/rhème, l'aspect, etc.

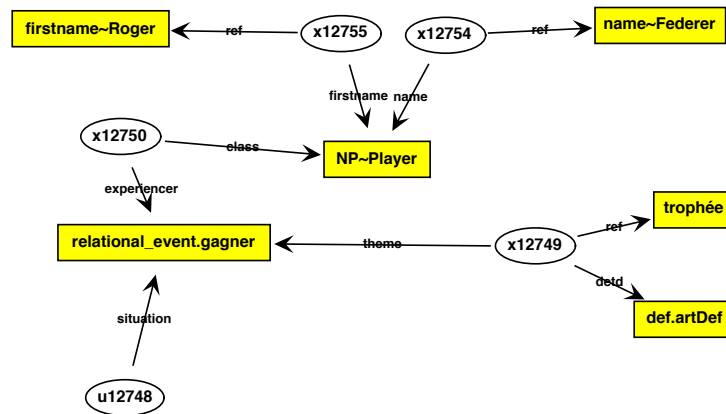


Figure 4. Graphe sémantique (sous forme graphique)

prédicat:sport\_activity.sport + argument:agent  
 ⇒ tennis:Player  
 prédicat:entité.trophée + argument:ref  
 ⇒ tennis:Trophy  
 prédicat:relational\_event.gagner + argument:theme + argument:experience  
 ⇒ tennis:isWonBy  
 prédicat:composed\_ExhibitionMatch + argument:sit  
 ⇒ tennis:ExhibitionMatch

Nous avons déjà montré le résultat dans la figure 1.

À cause du thésaurus très grands, une explosion combinatoire lors de la création des représentations sémantiques risque d'empêcher un résultat utilisable. Pour limiter le nombre de résultats et pour les trier qualitativement, le choix d'un prédicat (si le terminal en a plusieurs) est contrôlé par la structure thématique du thésaurus qui privilégie les prédicats du même thème (ou domaine ou macrodomaine). Finalement les graphes non-connexes sont ignorés s'il y a des graphes connexes.

### 2.3. Correction ontologique

Il nous reste néanmoins à corriger un problème que nous avons caché jusqu'à maintenant. Malgré le fait que les représentations créées ainsi n'utilisent que les classes et relations définies dans les ontologies, la cohérence des relations ontologiques utilisées avec les définitions de leurs domaines et portées (en anglais *range*) n'est pas encore assurée. Il est donc possible que les énoncés analysés parlent des choses qui « n'existent pas » dans le modèle du monde tel que défini dans les ontologies.

Par exemple, l'exemple de cet article, *le trophée a été gagné par Federer*, exige qu'il existe une relation ontologique associée au prédicat sémantique du verbe *gagner* (*relational\_event.gagner*) qui puisse lier un joueur et un trophée. Dans notre cas, l'ontologie du domaine « tennis » ne permet pas une telle relation car l'ontologie impose :

- qu'on ne puisse gagner que des *tennis:Finals* (gagné par des *tennis:Players*, c'est-à-dire la relation *tennis:isWonBy* exige comme domaine une instance de *tennis:Finals* (ou d'une sous-classe. La classe ontologique *tennis:Trophy* n'en fait pas partie) ;
- que des *tennis:Trophy* soient attribués par des *tennis:Tournaments* ;
- que des *tennis:Tournaments* contiennent des *tennis:Finals*.

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:tennis="http://www.acemedia.org/ontos/tennis#">
  <tennis:Player rdf:about="#V12750">
    <tennis:hasName>Federer</tennis:hasName>
    <tennis:hasFirstName>Roger</tennis:hasFirstName>
  </tennis:Player>
  <tennis:Trophy rdf:about="#V12749"/>
  <tennis:Finale rdf:about="#V12749_1">
    <tennis:isWonBy rdf:resource="#V12750"/>
  </tennis:Finale>
</rdf:RDF>

```

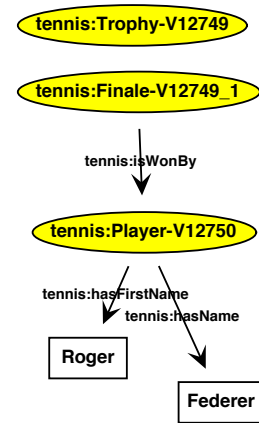


Figure 5. relation ontologique corrigée

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:tennis="http://www.acemedia.org/ontos/tennis#">
  <tennis:Player rdf:about="#V12750">
    <tennis:hasName>Federer</tennis:hasName>
    <tennis:hasFirstName>Roger</tennis:hasFirstName>
  </tennis:Player>
  <tennis:Trophy rdf:about="#V12749"/>
  <tennis:Finale rdf:about="#V12749_1">
    <tennis:isWonBy rdf:resource="#V12750"/>
  </tennis:Finale>
  <tennis:Tournament rdf:about="#N1">
    <tennis:awards rdf:resource="#V12749"/>
    <tennis:hasFinale rdf:resource="#V12749_1"/>
  </tennis:Tournament>

```

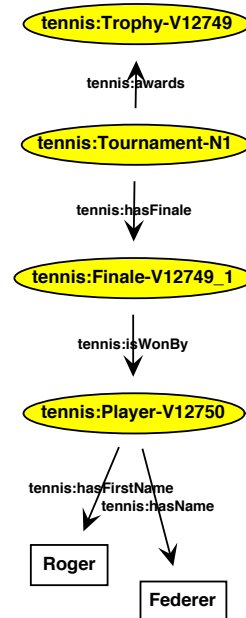


Figure 6. Représentation ontologique reliée

Afin d'obtenir une représentation ontologique qui est cohérente avec les ontologies du domaine nous appliquons un mécanisme de « correction ontologique ». Ce mécanisme est contrôlé par les ontologies. Dans un premier temps nous coupons la classe incohérente de la relation ontologique (pour le domaine et pour la portée) et nous la remplaçons par la classe spécifiée pour cette position dans l'ontologie. Dans notre exemple c'est la classe *tennis:Trophy* qui ne peut pas être le domaine de *tennis:isWonBy*. On obtient donc l'expression de la figure 5.

Finalement nous essayons d'établir un nouveau lien entre la classe découpée et la classe insérée en utilisant les relations définies dans l'ontologie. Bien sûr ce n'est pas toujours possible. Mais dans notre exemple des relations et classes ontologiques permettent d'arriver à la représentation finale et ontologiquement cohérente (figure 6).

Il nous reste le problème des énoncés hors domaine. C'est-à-dire si une phrase à analyser contient des mots qui ne correspondent à aucune classe ni relation définie dans l'ontologie. Par exemple *orange juice* n'existe pas dans notre exemple d'ontologie (qui ne couvre que le tennis). Au lieu de rendre une représentation complètement vide il est possible de rajouter dans

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:midlevel="http://www.acemia.org/ontos/midlevel#">
  <midlevel:unknown rdf:about="#D45317">
    <midlevel:isOfType>boisson.jus</midlevel:isOfType>
  </midlevel:unknown>
  <midlevel:unknown rdf:about="#D45318">
    <midlevel:isOfType>fruit.orange</midlevel:isOfType>
  </midlevel:unknown>
</rdf:RDF>

```

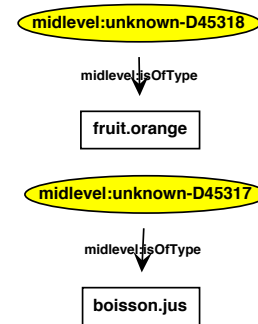


Figure 7. Représentation ontologique des concepts sans correspondant ontologique

l'ontologie (préalablement) une classe spécial *midlevel:unknown* et une relation spéciale *midlevel:isOfType* dont le domaine est la classe *midlevel:unknown* et sa portée est une simple chaîne. L'exemple *orange juice* aboutit à l'expression donnée en figure 7.

### 3. Conclusion et travaux futurs envisagés

L'article a montré comment une analyse syntaxico-sémantique profonde est appliquée afin de générer des représentations ontologiques pour un traitement ontologique (Web sémantique), stockage (par exemple RDF) et pour extraire des informations d'une base de connaissance (dans notre cas avec des requêtes en SPARQL).

Nous avons montré qu'il est important de dépasser le « key word spotting » afin de profiter des structures syntaxiques qui souvent expriment les relations ontologiques. La robustesse de cette approche vient du fait que dans le cas où l'analyse syntaxique de la phrase complète échoue, on peut recourir à une découpage en chunks avant d'être obligé de traiter la phrase comme une liste des mots isolés. Grâce à l'analyse profonde et au fait que les correspondances avec les classes et relations ontologiques sont établies sur le niveau sémantique et pas lexical, on peut traiter plusieurs langues (dans la mesure où on possède les lexiques et grammaires appropriés) en obtenant les mêmes représentations ontologiques pour des contenus textuels identiques. L'inconvénient est surtout l'explosion combinatoire lors de la création des graphes sémantiques et le filtrage qui peut supprimer par erreur la bonne solution.

Pour le futur nous envisageons de poursuivre trois pistes : le traitement des entités nommées pour les noms propres<sup>12</sup> ; le traitement « super-phrasal », c'est-à-dire la combinaison intelligente des représentations ontologiques issues de deux phrases consécutives (par exemple à l'aide du traitement des anaphores) ; et encore l'interprétation en fonction du contexte, de la phrase en entrée au lieu de sa simple représentation sémantique. Cela est important afin d'être capable d'utiliser correctement les relations ontologiques dont la portée est une chaîne, un nombre ou une valeur booléenne. Par exemple, l'ontologie du domaine tennis sur laquelle nous nous sommes appuyés dans cet article, contient des relations comme *tennis:isCovered* avec une valeur booléenne comme portée. Donc un énoncé comme *le court principal de Roland Garros a un toit* devrait contenir l'expression RDF suivante :

<sup>12</sup> Actuellement environ 2000 noms de personnes et noms de lieux se trouvent dans le lexique.



<tennis:Stadium rdf:about="#D45365">  
<tennis:isCovered>false</tennis:isCovered>  
</tennis:Stadium>

L'interprétation pourrait aussi aider à traiter les « usages créatifs » de la langue (par exemple métonymie, *pars pro toto*, contenant pour le conteneur, etc.).

Une autre piste à poursuivre serait de tester si une ontologie donnée est utilisable pour un texte à analyser. On pourrait compter le nombre des classes et relations qui apparaissent dans les représentations ontologiques générées et calculer le taux des classes du type *midlevel:unknown* et *midlevel:isOfType* parmi les classes et relations. Si ce taux est trop élevé on peut en tirer la conclusion que le domaine du texte est trop éloigné du domaine de l'ontologie utilisée et choisir une autre ontologie ou enrichir cette ontologie.

## Remerciements

Nous remercions les collègues et partenaires du projet IST aceMedia pour leur collaboration et pour nous avoir fourni les ontologies des domaines que nous utilisons, notamment l'ontologie du domaine « tennis ». Nous aimerions également remercier les trois relecteurs anonymes pour leurs remarques, suggestions et corrections.

## Références

- BUITELAAR P. & DECLERCK T. (2003). Linguistic Annotation for the Semantic Web. In S. HANDSCHUH & S. STAAB, Eds., *Annotation for the Semantic Web*. IOS Press.
- BUITELAAR P. & RAMAKA S. (2005). Unsupervised Ontology-based Semantic Tagging for Knowledge Markup. In *Learning in Web Search. 22nd International Conference on Machine Learning, Bonn*.
- DILL S., EIRON N., GIBSON D., GRUHA R., JHINGRAN A., GRUHL D., KANUNGO T., RAJAGOPALAN S., TOMKINS A., TOMLIN J. A. & ZIEN J. Y. (2003). SemTag and Seeker : Bootstrapping the semantic web via automated semantic annotation. In *Proceedings of the Twelfth International WWW Conference, Budapest, Hungary 2003* : ACM Press.
- ERDMANN M., MAEDCHE A., SCHNURR H.-P. & STAAB S. (2000). From Manual to Semi-automatic Semantic Annotation. About Ontology-based Text Annotation Tools. In P. BUITELAAR & K. HASIDA, Eds., *Proceedings of the CoLing 2000 Workshop on Semantic Annotation and Intelligent Content*.
- FRÖHLICH M. O. & RIET R. P. V. D. (1998). Using Multiple Ontologies in a Framework for Natural Language Generation. In A. GÓMEZ-PÉREZ & V. BENJAMINS, Eds., *Proceedings of the Workshop on Applications of Ontologies and Problem-Solving Methods, held in conjunction with ECAI 1998, Brighton*, p. 67–77.
- GOASDOUÉ F., LATTES V. & ROUSSET M.-C. (2000). The Use of CARIN Language and Algorithms for Information Integration. The Piciel Project. *International Journal of Cooperative Information Systems*, **9** :4, 383–401.
- GUIMIER DE NEEF E., BOUALEM M., CHARDENON C., FILOCHE P. & VINESSE J. (2002). Natural language processing software tools and linguistic data developed by France Télécom R&D. In *Indo European Conference on Multilingual Technologies, Pune, India*.
- HANDSCHUH S. & STAAB S. (2003). *Annotation for the Semantic Web*. IOS Press.

- HEINECKE J. (2006). Mapping ontological classes and relations onto linguistic data. France Télécom internal research paper.
- HEINECKE J. & TOUMANI F. (2003). A Natural Language Mediation System for E-Commerce applications. An ontology-based approach. In *Proceedings of Workshop Human Language Technology for the Semantic Web and Web Services. International Semantic Web Conference (ISWC), Sanibel Island, Florida, 20-23 October 2003*, p. 39–50.
- KAMP H. & REYLE U. (1993). *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Studies in Linguistics and Philosophy 42. Dordrecht : Kluwer.
- KIFER M., LAUSEN G. & WU J. (1995). Logical foundations of object oriented and frame-base languages. *Journal of the ACM*, **42 :4**, 741–843.
- LEHTOLA A., HEINECKE J. & BOUNSAYTHIP C. (2003). Intelligent Human Language Query Processing in MKBEEM. In C. STEPHANIDIS, Ed., *Universal Access in HCI : Inclusive Design in the Information Society. Proceedings of the 2nd International Conference on Universal Access in Human-Computer Interaction, Crete, Greece, 22-27 June 2003*, Proceedings of HCI International 2003 : Vol 4, p. 750–754, Mahwah, NJ : Lawrence Erlbaum Associates.
- TESNIÈRE L. (1959). *Éléments de syntaxe structurale*. Paris : Klincksieck.