

## **Atelier 2**

### **Acquisition et modélisation de relations sémantiques**

La constitution de ressources terminologiques et ontologiques repose en majeure partie sur l'exploration et l'exploitation des corpus textuels. L'information première fournie par ces corpus concerne tout d'abord le repérage de termes : listes de candidats termes extraits qui seront ensuite sélectionnés et validés par les experts. De manière de plus en plus généralisée, les corpus sont également exploités en vue d'en extraire les relations. De nombreux travaux ont étudié ainsi les relations sémantiques habituellement proposées par les ressources terminologiques (hyponymie, méronymie, synonymie). Dans cet atelier, nous aborderons plus particulièrement les relations sémantiques dites transversales, spécifiques au domaine étudié (localisation, cause, ...).

Il est à noter également que le travail nécessaire à la validation, formalisation, modélisation et exploitation des relations extraites reste néanmoins majoritairement manuel. Notre atelier propose une réflexion sur l'automatisation de l'acquisition de relations terminologiques : ce processus allant de leur extraction vers leur modélisation, avec une attention particulière pour l'articulation entre ces deux aspects. Ainsi, la problématique de cet atelier se trouve à la charnière entre les travaux en linguistique et terminologie et les travaux en modélisation et ontologie.

L'étude linguistique des relations sémantiques concerne souvent leur détection dans les corpus, et ensuite leur extraction et description. La maîtrise de différents types d'approches pour l'extraction de relations (marqueurs, patrons lexico-syntaxiques, règles d'association, apprentissage supervisé, analyse syntaxique, ...) est ainsi un moment clé pour leur étude, car chaque approche pourra proposer des informations différentes et spécifiques sur ces relations. Par exemple, les relations n-aires (Nguyen, Gaio & Sallaberry) ne se fondent pas sur les mêmes éléments et ne reçoivent pas la même description que les relations binaires (Aussenac-Gilles & Hernandez; Després & Nobécourt). Chacune des approches exploitées y contribue par conséquent d'une manière qui lui est propre.

Quant à la modélisation, souvent fondée sur les informations obtenues à l'issue de l'étape linguistique, elle conduit à une formalisation des relations. Comme nous l'avons noté, les approches linguistiques exploitées n'ont pas le même impact sur la modélisation des relations qu'elles proposent : elles peuvent être plus ou moins explicites quant à la nature de ces relations. Dans les travaux présentés, l'analyse syntaxique (Godbert & Royauté) ou les marqueurs et patrons (Nguyen, Gaio & Sallaberry; Aussenac-Gilles & Hernandez; Després & Nobécourt) permettent ainsi de détecter des structures prédicatives et les dépendances syntaxiques : les informations exploitées pour la modélisation des relations sémantiques au sein d'une ressource terminologique ou ontologique. Les corpus peuvent également fournir les informations utiles pour l'enrichissement ou une réorganisation formelle d'une ressource sémantique existante (Declerck & Lendvai). Enfin, l'importance d'un outil d'assistance à la modélisation est montrée au travers la présentation de la plateforme Dafoe (Aussenac-Gilles & Hernandez).

La thématique de recherche qui se trouve au centre de cet atelier n'est pas spécifique à un domaine de spécialité donné : elle est pertinente pour modéliser différentes activités. Ainsi, les travaux présentés se situent dans différents domaines spécialisés : électricité (Aussenac-Gilles & Hernandez), biologie (Godbert & Royauté), médecine et radiologie (Declerck & Lendvai), accidentologie routière (Després & Nobécourt), et le domaine de relations spatiales

et géographiques (Nguyen, Gaio & Sallaberry).

Les thèmes abordés par les papiers retenus concernent donc l'extraction de relations sémantiques particulières qui sont reliées à des concepts de domaine. La recherche de relations peut être facilitée par l'usage d'ontologies ou d'approches linguistiques particulières. Par ailleurs, les ontologies peuvent être enrichies par des relations sémantiques identifiées dans les corpus. A travers son thème, cet atelier s'adresse à plusieurs communautés scientifiques : terminologues, linguistes, ingénieurs de la connaissance, ontologues, etc.

Au cours de cet atelier, cinq papiers seront présentés et discutés :

1. *Nathalie Aussenac-Gilles & Nathalie Hernandez* - Du linguistique au conceptuel : identification de relations conceptuelles à partir de textes
2. *Elisabeth Godbert & Jean Royauté* - Exploring predicate-arguments structures in texts to relate biological entities
3. *Sylvie Després & Jérôme Nobecourt* - Construction d'ontologies à partir de textes : exploitation des relations verbales
4. *Van Tien Nguyen & Mauro Gaio & Christian Sallaberry* - Recherche de relations spatio-temporelles : une méthode basée sur l'analyse de corpus textuels
5. *Thierry Declerck & Piroska Lendvai* - Extraction de concepts et relations sémantiques à partir des labels d'ontologies

Avant de souhaiter aux participants, un atelier fertile en idées et en discussions, nous tenons à remercier les auteurs pour leur contribution et les membres du comité de programme pour leurs relectures constructives. Nous tenons également à remercier le comité d'organisation de la conférence TIA 2009.

Sylvie Després (LIM&BIO Univ. Paris13) & Natalia Grabar (SPIM, INSERM; HEGP, AP-HP)

#### **Comité de programme de l'atelier :**

- Cesar Antonio AGUILAR (Autonomous University of Queretaro, Mexique)
- Nathalie AUSSENAC-GILLES (IRIT, France)
- Caroline BARRIERE (Université d'Ottawa, Canada)
- Marc van CAMPENHOUDT (Université de Bruxelles, Belgique)
- Olivier BODENREIDER (National Library of Medicine, NCBI, USA)
- Mauro GAIO (LIUPPA, Université de Pau-Pays de l'Adour, France)
- Thierry HAMON (LIM&BIO, Université Paris 13 France)
- Agata JACKIEWICZ (Université Paris Sorbonne, France)
- Marie-Paule JACQUES Université de Strasbourg, France)
- Elizabeth MARSHMAN (Université d'Ottawa, Canada)
- Véronique MALAISE (Vrije Universiteit Amsterdam, Pays-Bas)
- Josette REBEYROLLE (Université Toulouse 2, France)