

Compte rendu de la 8^e Conférence internationale Terminologie et Intelligence Artificielle (TIA'09)

18-20 novembre 2009

<http://www.irit.fr/TIA09/>

La 8^e édition de la conférence internationale TIA'09 (Terminologie et Intelligence Artificielle) s'est déroulée du 18 au 20 novembre 2009 à l'Institut de Recherche en Informatique de Toulouse.

Présidentes du Comité de Programme :

M.-Cl. L'Homme (OLST, U. de Montréal), S. Szulman (LIPN, U. Paris 13)

Présidentes du Comité d'Organisation :

N. Aussenac-Gilles (IRIT, CNRS et U. Toulouse 3), A. Condamines (CLLE-ERSS, CNRS et U. Toulouse 2)

Les actes de la conférence et des ateliers sont désormais en ligne : <http://www.irit.fr/TIA09/>

Ils seront prochainement disponibles sur le site de diffusion d'actes d'ateliers internationaux <http://CEUR-WS.org/>

Comité de Programme : G. Aguado (Universidad Politécnica de Madrid, Espagne), A. Alcina (Universitat Jaume-I, Espagne), N. Aussenac-Gilles (IRIT, France), C. Barrière (NRC, Canada), O. Bodenreider (National Library of Medicine, USA), M.-T. Cabré (Universitat Pompeu Fabra, Spain), M. Van Campenhoudt (Université de Bruxelles, Belgique), F. Cerbah (Dassault-aviation, France), J. Charlet (AP-HP & INSERM, France), A. Condamines (CLLE-ERSS, France), J. Cussens (University of York, UK), L. Da Sylva (EBSI, Montréal, Canada), V. Delavigne (Institut national du cancer, France), P. Drouin (OLST, Montréal, Canada), P. Faber (Universidad de Granada, Espagne), U. Heid (Universität Stuttgart, Allemagne), F. Ibekwe-SanJuan (Université Lyon-3, France) K. Kageura (University of Tokyo, Japan), A. Nazarenko (LIPN, Université Paris 13), P. Sébillot (IRISA, France), K. Takeuchi (Okayama University, Japan), R. Temmerman (Erasmushogeschool, Belgique), Y. Toussaint (LORIA, France), P. Zweigenbaum (LIMSI-CNRS & CRIM-INALCO, France)

Vers une internationalisation de la conférence TIA

Cette conférence, éminemment interdisciplinaire puisqu'elle réunit des chercheurs en ingénierie des connaissances, traitement automatique des langues, linguistique de corpus et terminologie s'est particulièrement ouverte à l'international cette année. Deux éléments chiffrés rendent compte de cette ouverture. D'une part, sur les 63 inscrits, on compte 26 participants d'autres pays que la France, soit plus de 40%. Ces collègues provenaient de 10 pays (par ordre décroissant du nombre de participants) : Espagne (10), Canada (3), Italie (3), Allemagne (2), Chine (2), Croatie (2), Belgique (1), Inde (1), UK (1), USA (1). Cette diversité a également eu comme corollaire une représentation conséquente des travaux portant sur des langues variées.

D'autre part, sur les 13 communications plénières retenues, 6 ont été faites par des collègues étrangers. L'appel du comité de programme a été entendu puisque 32 articles ont été soumis au comité de programme.

Notons qu'on a pu compter jusqu'à 80 personnes présentes dans la salle, des collègues des deux laboratoires organisateurs (l'IRIT et CLLÉ-ERSS (Equipe de Recherche en Syntaxe et Sémantique) étant venus assister ponctuellement à des présentations.

Thématiques de recherche exposées

Poursuivant la tradition engagée dès 1995, TIA 2009 a été une tribune pour tous les travaux fondamentaux, méthodologiques et applicatifs gravitant autour des questions liées aux travaux de recherche en ingénierie des connaissances dans lesquels la linguistique textuelle joue un rôle important. Un accent particulier a été mis sur la recherche multidisciplinaire faisant converger les champs d'intérêts de l'ingénierie des connaissances, de la linguistique, de la terminologie, des sciences de l'information, du traitement automatique des langues, etc. Ainsi, dans la sélection des articles, nous avons tenu à assurer la diversité des interventions et leur intérêt sur les liens entre le texte et la représentation des connaissances dans des supports formels, ainsi que sur les études et expériences concrètes permettant de mieux comprendre le rôle des termes et des textes dans la transmission des connaissances spécialisées.

Cette année, le comité de programme de la conférence a focalisé l'appel à communication sur les termes, leurs réalisations dans les textes ainsi que sur les structures terminologiques. Les termes sont utilisés dans de nombreuses applications comme pivots de la structuration des connaissances : les ontologies, les thesaurus documentaires et autres ressources terminologiques plus classiques comme les dictionnaires spécialisés et les bases de données terminologiques. Entre les termes, on dégage un nombre plus ou moins élevé de relations sémantiques. Or, si toutes les applications citées ci-dessus se caractérisent par la recherche d'une certaine stabilité, la nature linguistique des termes et des relations dans lesquelles ils entrent soulève de nombreuses questions. Quelles relations représenter en fonction de l'application ? Comment représenter les termes ainsi que les relations dans lesquelles ils se trouvent ? Comment lier les relations représentées dans les ressources terminologiques et leurs réalisations linguistiques dans les textes ?

En sollicitant également des propositions de communications portant sur des théories sémantiques ou la terminologie par rapport à la linguistique textuelle et aux ontologies (notamment, les approches de linguistique théorique visant à caractériser le terme ou les structures terminologiques), le comité de programme a souhaité poursuivre une réflexion essentielle autour de questions fondamentales.

La conférence a été aussi un lieu de discussion sur des questions méthodologiques et théoriques portant sur la représentation des termes et des relations sémantiques ou conceptuelles dans des structures de données spécifiques (ontologies, thesaurus, etc.) ; la représentation des termes et des relations dans les applications multilingues ; la confrontation de terminologies, des ressources terminologiques et des ressources ontologiques provenant de langues, de communautés, de périodes différentes ; l'étude de problèmes théoriques et techniques liés à la constitution automatique ou manuelle de terminologies à partir de corpus mono ou multilingues ; l'étude des liens entre ontologies et terminologies ou thesaurus (exploitation de terminologies pour la constitution et la structuration d'ontologies). Les discussions portant sur l'automatisation partielle ou complète de l'acquisition ou de la représentation des connaissances ont été également denses, plus précisément, les méthodes de structuration automatique de terminologies ; l'identification de relations entre termes, la répartition en domaines, l'apport de la formalisation ontologique à l'explication de la sémantique de thesaurus et de terminologies). Les questions de l'évaluation et de la validation de terminologies, les difficultés liées à la constitution de terminologies multilingues, la réutilisabilité, la standardisation, la comparaison et la fusion de ressources terminologiques ou ontologiques, ont également été au cœur des préoccupations des intervenants (conférenciers ou participants). Enfin, le contexte applicatif des travaux visant la modélisation des connaissances a été mis en évidence : les applications exploitant des ressources terminologiques, comme le web sémantique, la gestion de flux documentaires, la recherche d'information, la classification de documents, la veille technologique, les question-réponse, etc.

Les sessions ont porté cette année sur les thématiques suivantes :

- Annotation sémantique
- Terminologie et applications à des domaines spécifiques
- Terminologie et applications
- Corpus et réseaux terminologiques
- Corpus et ontologies.

Conférences invitées et ateliers

Deux conférenciers invités ont présenté des travaux plus généraux pouvant intéresser le plus grand nombre :

- D. Maynard (Univ. Of Sheffield, UK): Bridging the Gap between Terminology and Linguistics. Cette conférencière a décrit la plateforme GATE et ses composants et la diversité de ses utilisations tout en explicitant ses fondements.
- C. Subirats (Univ. Autonoma de Barcelona (Spain) and Berkeley (USA)): Spanish FrameNet. An online Lexical Resource and its Application to NLP. C. Subirats a fait un bilan de l'élaboration de la version espagnole de la ressource lexicale FrameNet, de sa méthodologie et de ses applications au traitement automatique des langues.

Deux ateliers thématiques ont eu lieu le 20 novembre en parallèle. Ils ont attiré chacun environ 30 personnes, donnant lieu à des actes à diffusion restreinte, et disponibles en ligne sur le site de la conférence :

- Du thème au terme. Emergence et lexicalisation des connaissances (resp. : M. Slodzian et M. Valette). Cet atelier s'est intéressé aux travaux permettant de mieux comprendre le processus d'apparition de nouveaux concepts. Un compte rendu détaillé est proposé dans la partie suivante. Les actes sont disponibles sur le site de la conférence : <http://www.irit.fr/TIA09/thekey/tdmatelier1.htm>
- Acquisition et modélisation de relations sémantiques (resp. : S. Després et N. Grabar). L'atelier s'est focalisé sur des aspects essentiels de l'analyse des relations sémantiques dans les corpus spécialisés, c'est-à-dire comment les repérer dans les textes et comment les représenter dans différents types d'applications. Un compte rendu détaillé est proposé dans la partie suivante. Les actes sont disponibles sur le site de la conférence : <http://www.irit.fr/TIA09/thekey/tdmatelier2.htm>

Remerciements

Les participants ont exprimé leur satisfaction tant en ce qui concerne l'aspect scientifique que l'aspect organisationnel. La promenade dans Toulouse ainsi que la visite de la fondation Bemberg et le dîner du jeudi soir ont aussi emporté tous les suffrages.

Organismes ayant accepté de subventionner TIA'09, ce qui a permis de maintenir des tarifs d'inscriptions raisonnables et adaptés au budget des chercheurs en sciences du langage :

- Région Midi-Pyrénées,
- DGLFLF
- CFH (Facteurs Humains, Ingénierie linguistique)
- INRIA
- Université Paul Sabatier
- IRIT
- CLLE-ERSS

De plus, la conférence a été parrainée par deux associations scientifiques : l'AFIA et le GDR-I3.

Atelier « Du thème au terme : Emergence et lexicalisation des connaissances »

L'atelier, organisé par Monique Slodzian (ERTIM – INALCO, Paris) et Mathieu Valette (ATILF – CNRS, Nancy) a impliqué treize auteurs dont neuf sont intervenus à Toulouse, et plus de 20 participants. Les auteurs étaient Aurélie Picton (OLST, Montréal) ; Valentina Dragos et Marie-Christine Jaulent (SPIM – INSERM, Paris) ; Evelyne Bourion et Nadia Makouar (ERTIM – INALCO, Paris) ; Coralie Reutenauer, Evelyne Jacquy (ATILF – CNRS, Nancy) et Michelle Lecolle (CELTED, Metz) ; Armelle Boussidan, Sylvain Lupone et Sabine Ploux (L2C2 – CNRS, Lyon).

La terminologie textuelle s'est particulièrement consacrée à l'extraction de candidats termes dans les textes pour les expertiser et les valider comme concepts. Elle s'est en revanche peu intéressée à l'émergence de ces concepts dans les textes. Cependant, elle a promu l'idée que les textes sont les lieux de production des termes. L'objectif de cet atelier, qui a réuni pour l'essentiel des linguistes sensibilisés à la problématique et non des terminologues, a été d'illustrer l'hypothèse qu'avant d'accéder au statut de signes dont les signifiés sont normés (les termes), les concepts émergents se manifestent dans les textes sous des formes hétérogènes et instables que l'on pourrait assimiler, à des thèmes sémantiques ayant valeur de *préconcepts*. L'enjeu est donc d'une part de décrire et de modéliser ce processus émergentiste pour lui donner une place dans une théorie de la terminologie et, d'autre part, de fournir des outils d'identification et de détection pour la veille et la constitution de terminologies.

A. Picton (OLST, Montréal) a présenté une méthode de repérage des traces d'innovation dans des textes relevant du domaine des études spatiales basée sur quatre indices linguistiques (empreintes de fréquence, identification de contextes riches en connaissances évolutives, coexistence de variantes de termes, et identification des dépendances syntaxiques). Ses propositions permettent notamment une meilleure prise en compte de la diachronie en terminologie textuelle. V. Dragos et M.-C. Jaulent (SPIM – INSERM, Paris) ont présenté un corpus de textes médicaux traitant du *tabac*. Elles l'étudient de façon à circonscrire le réseau thématique selon le point de vue porté sur le concept de tabac selon qu'il est un facteur négatif explicite, implicite ou indirect, facteur neutre, etc.). Les auteurs établissent des corrélations entre genres textuels et sélection des thèmes. E. Bourion et N. Makouar (ERTIM – INALCO, Paris) ont tenté quant à elles d'isoler les différentes acceptions et les réseaux cooccurentiels du thème de la *finance islamique* hérité de plusieurs domaines (économie, philosophie, droit, religion) dans un contexte fortement multilingue (arabe, anglais et français). Elles étudient le processus d'assimilation et de lexicalisation d'un concept polymorphe dans le domaine au contraire bien structuré de la finance. C. Reutenauer, E. Jacquy, M. Valette (ATILF – CNRS, Nancy) et M. Lecolle (CELTED, Metz) cherchent à faire émerger de façon semi-automatique l'évolution de sens du nom propre *Outreau* à partir d'un corpus diachronique annoté en traits sémantiques. L'analyse permet d'observer une évolution diachronique de la signification du mot établie en termes de groupements de trait, évolution validée et corroborée par la lecture humaine. Enfin, A. Boussidan, S. Lupone, S. Ploux (L2C2 – CNRS, Lyon) se focalisent sur le mot *malbouffe* dans un corpus journalistique (*Le Monde*, 1997-2001) et examinent ses cooccurrences afin d'en étudier le réseau sémantique, ses glissements et sa stabilisation. Les auteurs esquissent également une réflexion sur la morphologie de ce mot au destin singulier.

Les interventions ont donné lieu à des débats nourris où linguistes, chercheurs en ingénierie des connaissances et terminologues ont eu le loisir d'exprimer leurs points de vue. La preuve a été apportée que ces recherches à l'articulation de la linguistique et de l'IC présentent un potentiel en termes de veille et de production de nouvelles méthodes de production des connaissances qu'il convient maintenant de faire fructifier.

Atelier « Acquisition et modélisation de relations sémantiques »

L'atelier « Acquisition et modélisation des relations sémantiques », organisé par Sylvie Després (LIM&BIO Université Paris 13) et Natalia Grabar (UMRS872 eq20 Université Paris 5 et HEGP) a eu pour problématique principale une réflexion approfondie sur l'articulation entre l'étude linguistique des relations sémantiques et leur formalisation, au sein d'un processus automatique de construction de ressources termino-ontologiques (RTO).

La sélection des papiers a été assurée par un comité de relecture international constitué de 13 chercheurs (Cesar Antonio Aguilar - Mexique, Nathalie Aussenac-Gilles – France, Caroline Barriere - Canada, Marc van Campenhoudt - Belgique, Olivier Bodenreider - États-Unis, Mauro Gaio - France, Thierry Hamon – France, Agata Jackiewicz - France, Marie-Paule Jacques - France, Élisabeth Marshman - Canada, Véronique Malaisé - Pays-Bas, Josette Rebeyrolle - France).

Cinq papiers ont été sélectionnés pour une présentation orale.

- Nathalie Aussenac & Nathalie Hernandez - **Du linguistique au conceptuel : identification de relations conceptuelles à partir de textes**

- Elisabeth Godbert & Jean Royauté - **Exploring predicate-arguments structures in texts to relate biological entities**

- Van Tien Nguyen & Mauro Gaio & Christian Sallaberry - **Recherche de relations spatio-temporelles : une éthode basée sur l'analyse de corpus textuels**

Nathalie Aussenac & Nathalie Hernandez - **Du linguistique au conceptuel : identification de relations conceptuelles à partir de textes**

- Sylvie Després & Jérôme Nobecourt - **Construction d'ontologies à partir de textes : exploitation des relations verbales**

Les différentes présentations ont mis en évidence les points d'articulation entre l'acquisition et la modélisation des relations sémantiques. L'enrichissement mutuel de ces deux étapes a également été considéré comme un point crucial dans les différentes présentations. Les méthodologies exploitant les patrons lexico-syntaxiques et l'analyse syntaxique montrent des résultats prometteurs. Les domaines scientifiques et techniques d'application sont nombreux : électricité, biologie, médecine et radiologie, accidentologie routière et le domaine de relations spatiales et géographiques. Cet éventail varié montre, si cela était nécessaire, que l'intérêt de la thématique de recherche abordée dans l'atelier n'est pas limité à un domaine applicatif particulier mais est commun à tout domaine et activité à modéliser. Par ailleurs, les pistes de recherche et les résultats déjà obtenus ne peuvent qu'encourager les chercheurs à poursuivre les travaux sur cette thématique.

Des frais d'inscription modiques ont permis de rassembler entre 25 et 30 chercheurs français ou venant d'autres pays et continents. A l'image de TIA, plusieurs communautés scientifiques étaient représentées : terminologues, linguistes, ingénieurs de la connaissance, ontologues, etc. Grâce à la qualité des travaux présentés et à l'intérêt que les participants ont porté à cette thématique de recherche, l'atelier a constitué un lieu d'échanges scientifiques stimulants. Des contacts et des collaborations ont pu être noués au cours de cet atelier.

Les prochaines rencontres « Terminologie et Intelligence Artificielle » (TIA'11) devraient se dérouler à Paris.