

---

## 28<sup>th</sup> May 2008

---

10:00-10:40 **Opening Session *Ministres***

10:45-11:45 **Antonio Zampolli Prize Talk *Ministres***

11:45-12:05 *Coffee break*

12:05-13:25 **4 Oral Sessions and 3 Poster Sessions in Parallel**

### **O1 - Information Extraction and Question Answering**

Chairperson: Bernardo Magnini

#### *Ministres*

12:05-12:25 Unsupervised Relation Extraction From Web Documents  
*Kathrin Eichler, Holmer Hensen and Günter Neumann*

12:25-12:45 Combining Multiple Models for Speech Information Retrieval  
*Muath Alzghool and Diana Inkpen*

12:45-13:05 Event Detection and Summarization in Weblogs with Temporal Collocations  
*Chun-Yuan Teng and Hsin-Hsi Chen*

13:05-13:25 The Usage of Various Lexical Resources to Improve Performance of Web Search Engines  
*Cvetana Krstev, Ranka Stanković, Duško Vitas, Ivan Obradović*

### **O2 - LRs: Infrastructures, Projects, Centers**

Chairperson: Chiu-yu Tseng

#### *Ambassadeurs*

12:05-12:25 The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics  
*Steven Bird, Robert Dale, Bonnie Dorr, Bryan Gibson, Mark Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev and Yee Fan Tan*

12:25-12:45 The Linguistic Data Consortium Member Survey: Purpose, Execution and Results  
*Marian Reed, Denise DiPersio and Christopher Cieri*

12:45-13:05 Language-Sites: Accessing and Presenting Language Resources via Geographic Information Systems  
*Dieter van Uytvanck, Alex Dukers, Jacquellijn Ringersma and Paul Trilsbeek*

13:05-13:25 CLARIN: Common Language Resources and Technology Infrastructure  
*Tamás Váradi, Steven Krauwer, Peter Wittenburg, Martin Wynne and Kimmo Koskenniemi*

### **O3 - Corpus, Lexicon and Evaluation**

Chairperson: Nuria Bel

#### *Fez 1*

12:05-12:25 Evaluating Dialogue Act Tagging with Naive and Expert Annotators  
*Jeroen Geertzen, Volha Petukhova and Harry Bunt*

12:25-12:45 Validating the Quality of Full Morphological Annotation  
*Drahomira Johanka Spoustová, Pavel Pecina, Jan Hajič and Miroslav Spousta*

12:45-13:05 Evaluating a German Sketch Grammar: A Case Study on Noun Phrase Case  
*Kremena Ivanova, Ulrich Heid, Sabine Schulte im Walde, Adam Kilgarriff and Jan Pomikalek*

13:05-13:25 Evaluating Complement-Modifier Distinctions in a Semantically Annotated Corpus  
*Mark McConville and Myroslava O. Dzikovska*

### **O4 - Multiparty and non-Verbal Communication**

Chairperson: Amy Isard

#### *Fez 2*

12:05-12:25 The PIT Corpus of German Multi-Party Dialogues  
*Petra-Maria Strauß, Holger Hoffmann, Wolfgang Minker, Heiko Neumann, Günther Palm, Stefan Scherer, Harald Traue and Ulrich Weidenbacher*

12:25-12:45 Annotation and Analysis of Overlapping Speech in Political Interviews  
*Martine Adda-Decker, Claude Barras, Gilles Adda, Patrick Paroubek, Philippe Boula de Mareüil and Benoit Habert*

12:45-13:05 Data Collection for the CHIL CLEAR 2007 Evaluation Campaign  
*Nicolas Moreau, Djamel Mostefa, Rainer Stiefelhagen, Susanne Burger and Khalid Choukri*

13:05-13:25 A Comparative Cross-Domain Study of the Occurrence of Laughter in Meeting and Seminar Corpora  
*Susanne Burger, Kornel Laskowski and Matthias Woelfel*

12:05-13:25 **Poster Sessions *Asni Room***

## **P1 - Corpus Construction and Annotation**

Chairperson: Justus Roux

Process Model for Composing High-quality Text Corpora

*Mikko Loumela*

AnCora: Multilevel Annotated Corpora for Catalan and Spanish

*Mariona Taulé, M. Antònia Martí and Marta Recasens*

The U.S. Policy Agenda Legislation Corpus Volume 1 - a Language Resource from 1947 - 1998

*Stephen Purpura, John Wilkerson and Dustin Hillard*

Unsupervised Resource Creation for Textual Inference Applications

*Jeremy Bensley and Andrew Hickl*

A Simple Method for Tagset Comparison

*Markus Dickinson and Charles Jochim*

From D-Coi to SoNaR: A reference corpus for Dutch

*Nelleke Oostdijk, Martin Reynaert, Paola Monachesi, Gertjan Van Noord, Roeland Ordelman, Ineke Schuurman and Vincent Vandeghiste*

Relationships between Nursing Conversations and Activities

*Hiroimi Itoh Ozaku, Akinori Abe, Kaoru Sagara and Kiyoshi Kogure*

Management of Large Annotation Projects Involving Multiple Human Judges: a Case Study of GALE Machine Translation Post-editing

*Meghan Lammie Glenn, Stephanie Strassel, Lauren Friedman and Haejoong Lee*

Bootstrapping Language Description: the case of Mpiemo (Bantu A, Central African Republic)

*Harald Hammarström, Christina Thornell, Malin Petzel and Torbjörn Westerlund*

Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus

*Satoshi Sato, Suguru Matsuyoshi and Yohsuke Kondoh*

## **P2 - LRs for Specific Domains: Bio-Medicine and Chemistry**

Chairperson: Beatrice Alex

Building a Bio-Event Annotated Corpus for the Acquisition of Semantic Frames from Biomedical Corpora

*Paul Thompson, Philip Cotter, John McNaught, Sophia Ananiadou, Simonetta Montemagni, Andrea Trabucco and Giulia Venturi*

Language Resources and Chemical Informatics

*C.J. Rupp, Ann Copestake, Peter Corbett, Peter Murray-Rust, Advait Siddharthan, Simone Teufel and Benjamin Waldron*

Semantic Annotations for Biology: a Corpus Development Initiative at the Jena University Language & Information Engineering (JULIE) Lab

*Udo Hahn, Elena Beisswanger, Ekaterina Buyko, Michael Poprat, Katrin Tomanek and Joachim Wermter*

A lexicon for biology and bioinformatics: the BOOTStrep experience.

*Valeria Quochi, Monica Monachini, Riccardo Del Gratta and Nicoletta Calzolari*

Dependency-Based Relation Mining for Biomedical Literature

*Fabio Rinaldi, Gerold Schneider, Karel Kaljurand and Michael Hess*

MeSH@: from a Controlled Vocabulary to a Processable Resource

*Dimitrios Kokkinakis*

A Semantically Annotated Swedish Medical Corpus

*Dimitrios Kokkinakis*

Learning Patterns for Building Resources about Semantic Relations in the Medical Domain

*Mehdi Embarek and Olivier Ferret*

## **P3 - Syntactically Annotated Resources and Related Tools**

Chairperson: Tomaz Erjavec

Automatic Extraction of Subcategorization Frames for Italian

*Dino Ienco, Serena Villata and Cristina Bosco*

Parallel Multi-Theory Annotations of Syntactic Structure

*Jerid Francom and Mans Hulden*

Tagging a Hebrew Corpus: the Case of Participles

*Meni Adler, Yael Netzer, Yoav Goldberg, David Gabay and Michael Elhadad*

Unsupervised Parts-of-Speech Induction for Bengali

*Joydeep Nath, Monojit Choudhury, Animesh Mukherjee, Christian Biemann and Niloy Ganguly*

Tagging Spanish Texts: the Problem of Problem of "SE"

*Guadalupe Aguado de Cea, Javier Puche and José Ángel Ramos*

Does Netgraph Fit Prague Dependency Treebank?

*Jiří Mirovský*

The Kalshnikov 691 Dependency Bank

*Tomas By*

Treebank-Based Acquisition of LFG Parsing Resources for French

*Natalie Schluter and Josef van Genabith*

Chooser: a Multi-Task Annotation Tool  
*Svetla Koeva, Borislav Rizov and Svetlozara Leseva*

BOEMIE Ontology-Based Text Annotation Tool  
*Pavlina Fragkou, Georgios Petasis, Aris Theodorakos, Vangelis Karkaletsis, Constantine D. Spyropoulos*

Minding the Source: Automatic Tagging of Reported Speech in Newspaper Articles  
*Ralf Krestel, Sabine Bergler and René Witte*

13:25-14:40 Lunch break

14:40-16:20 **4 Oral Sessions and 4 Poster Sessions in Parallel**

### **O5 - Spatio-Temporal Annotation**

Chairperson: Jerry Hobbs

#### ***Ministres***

14:40-15:00 SpatialML: Annotation Scheme, Corpora, and Tools  
*Inderjeet Mani, Janet Hitzeman, Justin Richer, Dave Harris, Rob Quimby and Ben Wellner*

15:00-15:20 Building a Corpus of Temporal-Causal Structure  
*Steven Bethard, William Corvey, Sara Klengenstein and James H. Martin*

15:20-15:40 Computational Models for Event Type Classification in Context  
*Alessandra Zarcone and Alessandro Lenci*

15:40-16:00 GMT to +2 or how can TimeML be used in Romanian  
*Corina Forascu*

16:00-16:20 Annotating “tense” in a Tense-less Language  
*Nianwen Xue, Hua Zhong and Kai-Yun Chen*

### **O6 - Syntax and Parsing**

Chairperson: Sadao Kurohashi

#### ***Ambassadeurs***

14:40-15:00 Subdomain Sensitive Statistical Parsing using Raw Corpora  
*Barbara Plank and Khalil Sima'an*

15:00-15:20 Developing a TT-MCTAG for German with an RCG-based Parser  
*Laura Kallmeyer, Timm Lichte, Wolfgang Maier, Yannick Parmentier and Johannes Dellert*

15:20-15:40 Some Fine Points of Hybrid Natural Language Parsing  
*Peter Adolphs, Stephan Oepen, Ulrich Callmeier, Berthold Crysmann, Dan Flickinger and Bernd Kiefer*

15:40-16:00 Evaluating and Extending the Coverage of HPSG Grammars: A Case Study for German  
*Jeremy Nicholson, Valia Kordoni, Yi Zhang, Timothy Baldwin and Rebecca Dridan*

16:00-16:20 Robust Parsing with a Large HPSG Grammar  
*Yi Zhang and Valia Kordoni*

### **O7 - Document Classification**

Chairperson: Chris Brew

#### ***Fez 1***

14:40-15:00 Modeling Document Dynamics: an Evolutionary Approach  
*Jahna Otterbacher and Dragomir Radev*

15:00-15:20 Semantic Vectors: a Scalable Open Source Package and Online Technology Management Application  
*Dominic Widdows and Kathleen Ferraro*

15:20-15:40 Revealing Relations between Open and Closed Answers in Questionnaires through Text Clustering Evaluation  
*Magnus Rosell and Sumithra Velupillai*

15:40-16:00 Personae: a Corpus for Author and Personality Prediction from Text  
*Kim Luyckx and Walter Daelemans*

16:00-16:20 Using the Complexity of the Distribution of Lexical Elements as a Feature in Authorship Attribution  
*Leanne Spracklin, Diana Inkpen and Amiya Nayak*

### **O8 - Multimodal Annotation Tools**

Chairperson: Jean-Claude Martin

#### ***Fez 2***

14:40-15:00 An Exchange Format for Multimodal Annotations  
*Thomas Schmidt, Susan Duncan, Oliver Ehmer, Jeffrey Hoyt, Michael Kipp, Dan Loehr, Magnus Magnusson, Travis Rose and Han Sloetjes*

15:00-15:20 SCARE: a Situated Corpus with Annotated Referring Expressions  
*Laura Stoia, Darla Magdalena Shockley, Donna K. Byron and Eric Fosler-Lussier*

15:20-15:40 Annotation by Category: ELAN and ISO DCR  
*Han Sloetjes and Peter Wittenburg*

15:40-16:00 A Common Multimedia Annotation Framework for Cross Linking Cultural Heritage Digital Collections  
*Hennie Brugman, Véronique Malaisé and Laura Hollink*

16:00-16:20 Creating and Exploiting Multimodal Annotated Corpora  
*Philippe Blache, Roxane Bertrand and Gaëlle Ferré*

14:40-16:20 **Poster Sessions Menara**

#### **P4 - Named Entity Recognition, Information Extraction and Document Classification**

Chairperson: Andrei Popescu-Belis

KYOTO: a System for Mining, Structuring and Distributing Knowledge across Languages and Cultures  
*Piek Vossen, Eneko Agirre, Nicoletta Calzolari, Christiane Fellbaum, Shu-kai Hsieh, Chu-Ren Huang, Hitoshi Isahara, Kyoko Kanzaki, Andrea Marchetti, Monica Monachini, Federico Neri, Remo Raffaelli, German Rigau, Maurizio Tesconi and Joop VanGent*

Extracting and Querying Relations in Scientific Papers on Language Technology  
*Ulrich Schäfer, Hans Uszkoreit, Christian Federmann, Torsten Marek and Yajing Zhang*

Named Entity Relation Mining using Wikipedia  
*Adrian Iftene and Alexandra Balahur-Dobrescu*

Named Entity Recognition for Digitised Historical Texts  
*Claire Grover, Sharon Givon, Richard Tobin and Julian Ball*

Entity Translation and Alignment in the ACE-07 ET Task  
*Zhiyi Song and Stephanie Strassel*

Automated Subject Induction from Query Keywords through Wikipedia Categories and Subject Headings  
*Yoji Kiyota, Noriyuki Tamura, Satoshi Sakai, Hiroshi Nakagawa and Hidetaka Masuda*

Using Random Indexing to improve Singular Value Decomposition for Latent Semantic Analysis  
*Linus Sellberg and Arne Jönsson*

#### **P5 - Multi-Word Expressions**

Chairperson: Simonetta Montemagni

Harvesting Multi-Word Expressions from Parallel Corpora  
*Špela Vintar and Darja Fišer*

Integration of a Multilingual Keyword Extractor in a Document Management System  
*Andrea Agili, Marco Fabbri, Alessandro Panunzi and Manuel Zini*

Dictionary of Multiword Expressions for Translation into highly Inflected Languages  
*Daiga Dekšne, Raivis Skadins and Inguna Skadina*

Verb-Noun Collocation SyntLex Dictionary: Corpus-Based Approach  
*Grazyna Vetulani, Zygmunt Vetulani and Tomasz Obrębski*

Targeting Chinese Nominal Compounds in Corpora  
*Weiruo Qu, Christoph Ringlstetter and Randy Goebel*

Using Semantically Annotated Corpora to Build Collocation Resources  
*Margarita Alonso Ramos, Owen Rambow and Leo Wanner*

#### **P6 - Ontologies and Knowledge**

Chairperson: Massimo Poesio

Eksairesis: A Domain-Adaptable System for Ontology Building from Unstructured Text  
*Katia Lida Kermanidis, Aristomenis Thanopoulos, Manolis Maragoudakis and Nikos Fakotakis*

Conceptual Modeling of Ontology-based Linguistic Resources with a Focus on Semantic Relations  
*Francisco Alvarez Montero, Antonio Vaquero Sanchez and Fernando Saenz Perez*

Ontology Search with the OntoSelect Ontology Library  
*Paul Buitelaar and Thomas Eigner*

A Framework for Multilingual Ontology Mapping  
*Cássia Trojahn, Paulo Quaresima and Renata Vieira*

Acquiring a Taxonomy from the German Wikipedia  
*Laura Kassner, Vivi Nastase and Michael Strube*

LMM: an OWL-DL MetaModel to Represent Heterogeneous Lexical Knowledge  
*Davide Picca, Alfio Massimiliano Gliozzo and Aldo Gangemi*

Development of the Japanese WordNet  
*Hitoshi Isahara, Fransis Bond, Kiyotaka Uchimoto, Masao Utiyama and Kyoko Kanzaki*

Lexical Ontology Extraction using Terminology Analysis: Automating Video Annotation  
*Neil Newbold, Bogdan Vrusias and Lee Gillam*

Workbench with Authoring Tools for Collaborative Multi-lingual Ontological Knowledge Construction and Maintenance  
*Mukda Suktarachan, Dussadee Thamvijit, Daoyos Noikongka, Panita Yongyuth, Puwarat Pavaputanont Na Mahasarakham, Asanee Kawtrakul, Asanee Kawtrakul and Margherita Sini*

Towards Semi Automatic Construction of a Lexical Ontology for Persian  
*Mehrnoush Shamsfard*

Mapping Roget's Thesaurus and WordNet to French  
*Gerard de Melo and Gerhard Weikum*

Representation of Atypical Entities in Ontologies  
*Christophe Jouis and Julien Bourdaillet*

Extracting Concrete Senses of Lexicon through Measurement of Conceptual Similarity in Ontologies  
*Siaw-Fong Chung, Laurent Prévot, Mingwei Xu, Kathleen Ahrens, Shu-Kai Hsieh and Chu-Ren Huang*

A Contextual Dynamic Network Model for WSD Using Associative Concept Dictionary  
*Jun Okamoto, Kiyoko Uchiyama and Shun Ishizaki*

A Semantic Memory for Incremental Ontology Population  
*Berenike Loos and Lasse Schwarten*

### **P7 - Term Identification/Extraction and Terminological Databases**

Chairperson: Maria Gavrilidou

Turning a Term Extractor into a new Domain: first Experiences  
*Jorge Vivaldi, Anna Joan and Mercè Lorente*

Similar Term Discovery using Web Search  
*Peter Anick, Vijay Murthi and Shaji Sebastian*

Temporal Aspects of Terminology for Automatic Term Recognition: Case Study on Women's Studies Terms  
*Junko Kubo, Keita Tsuji and Shigeo Sugimoto*

A Comparative Evaluation of Term Recognition Algorithms  
*Ziqi Zhang, Jose Iria, Christopher Brewster and Fabio Ciravegna*

Learning-based Detection of Scientific Terms in Patient Information  
*Veronique Hoste, Els Lefever, Klaar Vanopstal and Isabelle Delaere*

WNTERM: Enriching the MCR with a Terminological Dictionary  
*Eli Pociello, Antton Gurrutxaga, Eneko Agirre, Izaskun Aldezabal and German Rigau*

Encoding Terms from a Scientific Domain in a Terminological Database: Methodology and Criteria  
*Rita Marinelli, Melissa Tiberi and Remo Bindi*

16:20-16:40 Coffee break

16:40-18:20 **4 Oral Sessions and 1 Poster Session in Parallel**

### **O9 - Lexicon, Corpus and Semantics**

Chairperson: Claire Gardent

#### *Ministres*

16:40-17:00 The Encoding of Lexical Implications in VerbNet Predicates of Change of Locations  
*Annie Zaenen, Daniel Bobrow and Cleo Condoravdi*

17:00-17:20 FATE: a FrameNet-Annotated Corpus for Textual Entailment  
*Aljoscha Burchardt and Marco Pennacchiotti*

17:20-17:40 Projecting Propbank Roles onto the CCGbank  
*Stephen Boxwell and Michael White*

17:40-18:00 Integrating Lexical Units, Synsets and Ontology in the Cornetto Database  
*Piek Vossen, Isa Maks, Roxane Segers and Hennie VanderVliet*

18:00-18:20 Complete and Consistent Annotation of WordNet using the Top Concept Ontology  
*Javier Álvez, Jordi Atserias, Jordi Carrera, Salvador Climent, Egoitz Laparra, Antoni Oliver and German Rigau*

### **O10 - Multimodal and Speech Data over the Web**

Chairperson: Siegfried Kunzmann

#### *Ambassadeurs*

16:40-17:00 A Conceptual Approach to Web Image Retrieval  
*Adrian Popescu and Gregory Grefenstette*

17:00-17:20 On the Use of Web Resources and Natural Language Processing Techniques to Improve Automatic Speech Recognition Systems  
*Gwénolé Lecorvé, Guillaume Gravier and Pascale Sébillot*

17:20-17:40 Local Methods for On-Demand Out-of-Vocabulary Word Retrieval  
*Stanislas Oger, Georges Linares and Frederic Bechet*

17:40-18:00 Exploring and Enriching a Language Resource Archive via the Web  
*Marc Kemps-Snijders, Alex Klassmann, Claus Zinn, Peter Berck, Albert Russel and Peter Wittenburg*

18:00-18:20 Talking and Looking: the SmartWeb Multimodal Interaction Corpus  
*Florian Schiel and Hannes Mögele*

### **O11 - Coreference and Discourse**

Chairperson: Rebecca Passonneau

#### *Fez I*

16:40-17:00 In Contrast - A Complex Discourse Connective  
*Erhard Hinrichs and Monica Lau*

17:00-17:20 Towards a Reference Corpus of Web Genres for the Evaluation of Genre Identification Systems  
*Georg Rehm, Marina Santini, Alexander Mehler, Pavel Braslavski, Rüdiger Gleim, Andrea Stubbe, Svetlana Symonenko, Mirko Tavosanis and Vedrana Vidulin*

- 17:20-17:40 Error Analysis for Learning-based Coreference Resolution  
*Olga Uryupina*
- 17:40-18:00 From Sentence to Discourse: Building an Annotation Scheme for Discourse Based on Prague Dependency Treebank  
*Lucie Mladová, Šárka Zikánová and Eva Hajičová*
- 18:00-18:20 A Corpus for Cross-Document Co-reference  
*David Day, Janet Hitzeman, Michael Wick, Keith Crouch and Massimo Poesio*

### **O12 - Named Entity Recognition**

Chairperson: Satoshi Sekine

*Fez 2*

- 16:40-17:00 Named Entity WordNet  
*Antonio Toral, Rafael Muñoz and Monica Monachini*
- 17:00-17:20 Is this NE tagger getting old?  
*Cristina Mota and Ralph Grishman*
- 17:20-17:40 Improving NER in Arabic Using a Morphological Tagger  
*Benjamin Farber, Dayne Freitag, Nizar Habash and Owen Rambow*
- 17:40-18:00 Identifying Foreign Person Names in Chinese Text  
*Stephan Busemann and Yajing Zhang*
- 18:00-18:20 Low-Complexity Heuristics for Deriving Fine-Grained Classes of Named Entities from Web Textual Data  
*Marius Pasca*
- 16:40-18:20 **Poster Sessions Asni Room**

### **P8 - Information Extraction, Question Answering and Document Classification**

Chairperson: Barbara Di Eugenio

- An Evaluation Resource for Geographic Information Retrieval  
*Thomas Mandl, Fredric Gey, Giorgio Di Nunzio, Nicola Ferro, Mark Sanderson, Diana Santos and Christa Womser-Hacker*
- Bilingual Text Classification using the IBM 1 Translation Model  
*Jorge Civera and Alfons Juan-Ciscar*
- Ping-pong Document Clustering using NMF and Linkage-Based Refinement  
*Hiroyuki Shinnou and Minoru Sasaki*
- Spectral Clustering for a Large Data Set by Reducing the Similarity Matrix Size  
*Hiroyuki Shinnou and Minoru Sasaki*
- A Text-based Query Interface to OWL Ontologies  
*Danica Damjanovic, Valentin Tablan and Kalina Bontcheva*
- A Research on Automatic Chinese Catchword Extraction  
*Han Ren, Donghong Ji and Lei Han*
- ParsCit: an Open-source CRF Reference String Parsing Package  
*Isaac Council, C Lee Giles and Min-Yen Kan*
- Automatic Acquisition of Usage Information for Language Resources  
*Shunsuke Kozawa, Hitomi Tohyama, Kiyotaka Uchimoto and Shigeki Matsubara*
- Cost-Sensitive Learning in Answer Extraction  
*Michael Wiegand, Jochen L. Leidner and Dietrich Klakow*
- Definition Extraction Using a Sequential Combination of Baseline Grammars and Machine Learning Classifiers  
*Lukasz Degórski, Michał Marcińczuk and Adam Przepiórkowski*
- Yet another Platform for Extracting Knowledge from Corpora  
*Francesca Fallucchi and Fabio Massimo Zanzotto*
- A Framework for Identity Resolution and Merging for Multi-source Information Extraction  
*Milena Yankova, Horacio Saggion and Hamish Cunningham*
- Experiments to Investigate the Connection between Case Distribution and Topical Relevance of Search Terms in an Information Retrieval Setting  
*Jussi Karlgren, Hercules Dalianis and Bart Jongejan*
- Identifying Strategic Information from Scientific Articles through Sentence Classification  
*Fidelia Ibekwe-SanJuan, Chaomei Chen and Roberto Pinho*
- Keywords, k-NN and Neural Networks: a Support for Hierarchical Categorization of Texts in Brazilian Portuguese  
*Susana Azeredo, Silvia Moraes and Vera Lima*
- Automatic Extraction of Textual Elements from News Web Pages  
*Hossam Ibrahim, Kareem Darwish and Abdel-Rahim Madany*
- Extraction of Informative Expressions from Domain-specific Documents  
*Eiko Yamamoto, Hitoshi Isahara, Akira Terada and Yasunori Abe*
- Connecting Text Mining and Pathways using the PathText Resource  
*Rune Sætre, Brian Kemper, Kanae Oda, Naoaki Okazaki, Yukiko Matsuoka, Norihiro Kikuchi, Hiroaki Kitano, Yoshimasa Tsuruoka, Sophia Ananiadou and Jun'ichi Tsujii*
- Detecting Co-Derivative Documents in Large Text Collections  
*Jan Pomikálek and Pavel Rychlý*

Extraction and Evaluation of Keywords from Learning Objects: a Multilingual Approach  
*Lothar Lemnitzer and Paola Monachesi*

Exploiting the Role of Position Feature in Chinese Relation Extraction  
*Peng Zhang, Wenjie Li, Furu Wei, Qin Lu and Yuexian Hou*

Authorship Attribution of E-Mail: Comparing Classifiers over a New Corpus for Evaluation  
*Ben Allison and Louise Guthrie*

Creating a Research Collection of Question Answer Sentence Pairs with Amazon's Mechanical Turk  
*Michael Kaisser and John Lowe*

Adaptation of Relation Extraction Rules to New Domains  
*Feiyu Xu, Hans Uszkoreit, Hong Li and Niko Felger*

Boosting Precision and Recall of Hyponymy Relation Acquisition from Hierarchical Layouts in Wikipedia  
*Asuka Sumida, Naoki Yoshinaga and Kentaro Torisawa*

Parameters for Topic Boundary Detection in Multi-Party Dialogues  
*Margot Mieskes and Michael Strube*

Semantic Press  
*Eugenio Picchi, Eva Sassolini, Sebastiana Cucurullo, Francesca Bertagna and Paola Baroni*

An Approach to Modeling Heterogeneous Resources for Information Extraction  
*Lei Xia and Jose Iria*

On Classifying Coherent/Incoherent Romanian Short Texts  
*Anca Dinu*

Characterization of Scientific and Popular Science Discourse in French, Japanese and Russian  
*Lorraine Goeuriot, Natalia Grabar and Béatrice Daille*

Converting Romanized Persian to the Arabic Writing Systems  
*Jalal Maleki and Lars Ahrenberg*

Unsupervised Learning-based Anomalous Arabic Text Detection  
*Nasser Abouzakhar, Ben Allison and Louise Guthrie*

Condensing Sentences for Subtitle Generation  
*Prokopis Prokopidis, Vassia Karra, Aggeliki Papagianopoulou and Stelios Piperidis*

Making Text Resources Accessible to the Reader: the Case of Patent Claims  
*Simon Mille and Leo Wanner*

18:20-18:25 *5-minute Break*

18:25-19:45 **1 Panel, 2 Oral Sessions and 4 Poster Sessions in Parallel**

18:25-19:45 **Panel: Funding Agencies in Support to Language Processing: from the Regional to the Multinational Scale**

Moderators: Joseph Mariani and Nicoletta Calzolari

*Ministres*

### **O13 - Parallel and Multilingual Resources**

Chairperson: Timothy Baldwin

*Ambassadeurs*

18:25-18:45 Annotation Guidelines for Chinese-Korean Word Alignment  
*Jin-Ji Li, Dong-Il Kim and Jong-Hyeok Lee*

18:45-19:05 CzEng 0.7: Parallel Corpus with Community-Supplied Translations  
*Ondřej Bojar, Miroslav Janiček, Zdeněk Žabokrtský, Pavel Česka and Peter Beňa*

19:05-19:25 Toward Active Learning in Data Selection: Automatic Discovery of Language Features During Elicitation  
*Jonathan Clark, Robert Frederking and Lori Levin*

19:25-19:45 Babylon Parallel Text Builder: Gathering Parallel Texts for Low-Density Languages  
*Michael Mohler and Rada Mihalcea*

### **O14 - Evaluation Tools and Methodologies**

Chairperson: Mark Przybocki

*Fez I*

18:25-18:45 SECTra\_w.1: an Online Collaborative System for Evaluating, Post-editing and Presenting MT Translation Corpora  
*Cong-phap Huynh, Christian Boitet and Hervé Blanchon*

18:45-19:05 Adjudicator Agreement and System Rankings for Person Name Search  
*Mark Arehart, Chris Wolf and Keith J. Miller*

19:05-19:25 Evaluating Summaries Automatically - A system Proposal  
*Paulo C F de Oliveira, Edson Wilson Torrens, Alexandre Cidral, Sidney Schossland and Evandro Bittencourt*

19:25-19:45 Do we Still Need Gold Standards for Evaluation?  
*Thierry Poibeau and Cédric Messiant*

**P9 - Authoring Tools and Related Resources**

Chairperson: Sonja Bosch

Exploiting Lexical Resources for Disambiguating CJK and Arabic Orthographic Variants  
*Jack Halpern*

Automatic Document Quality Control  
*Neil Newbold and Lee Gillam*

OpenCCG Workbench and Visualization Tool  
*Thepchai Supnithi, Suchinder Singh, Taneth Ruangrajitpakorn, Prachya Boonkwan and Monthika Boriboon*

Using the Web as a Linguistic Resource to Automatically Correct Lexico-Syntactic Errors  
*Mathieu Hermet, Alain Désilets and Stan Szpakowicz*

I saw TREE trees in the park: How to Correct Real-Word Spelling Mistakes  
*Davide Fossati and Barbara Di Eugenio*

User-Centred Design of Error Correction Tools  
*Marti Quixal, Toni Badia, Francesc Benavent, Jose R. Boulosa, Judith Domingo, Bernat Grau, Guillem Massó and Oriol Valentin*

Professor or Screaming Beast? Detecting Anomalous Words in Chinese  
*Wei Liu, Ben Allison and Louise Guthrie*

Spelling Correction: from Two-Level Morphology to Open Source  
*Iñaki Alegria, Klara Ceberio, Nerea Ezeiza, Aitor Soroa and Gregorio Hernandez*

Automatic Rewriting of Patient Record Narratives  
*Catalina Hallett and David Hardcastle*

**P10 - Coreference and Discourse**

Chairperson: Inderjeet Mani

BART: A modular toolkit for coreference resolution  
*Yannick Versley, Simone Ponzetto, Massimo Poesio, Vladimir Eidelman, Alan Jern, Jason Smith, Xiaofeng Yang and Alessandro Moschitti*

ANAWIKI: Creating Anaphorically Annotated Resources through Web Cooperation  
*Massimo Poesio, Udo Kruschwitz and Chamberlain Jon*

Influence of Text Type and Text Length on Anaphoric Annotation  
*Daniela Goecke, Maik Stührenberg and Andreas Witt*

Annotating Abstract Pronominal Anaphora in the DAD Project  
*Costanza Navarretta and Sussi Olsen*

Deriving Rhetorical Complexity Data from the RST-DT Corpus  
*Sandra Williams and Richard Power*

Knowledge-based Coreference Resolution for Hungarian  
*Márton Miháلتz*

The Italian Particle “ne”: Corpus Construction and Analysis  
*Malvina Nissim and Sara Perboni*

**P11 - Tools, Systems, Applications**

Chairperson: Emanuele Pianta

Introducing DRS (The Digital Replay System): a Tool for the Future of Corpus Linguistic Research and Analysis  
*Dawn Knight and Paul Tennent*

An Inverted Index for Storing and Retrieving Grammatical Dependencies  
*Michaela Atterer and Hinrich Schütze*

MaltEval: an Evaluation and Visualization Tool for Dependency Parsing  
*Jens Nilsson and Joakim Nivre*

New Functions of FrameSQL for Multilingual FrameNets  
*Hiroaki Sato*

Division of Example Sentences Based on the Meaning of a Target Word Using Semi-Supervised Clustering  
*Hiroyuki Shinnou and Minoru Sasaki*

The Japanese FrameNet Software Tools  
*Hiroaki Saito, Shunta Kuboya, Takaaki Sone, Hayato Tagami and Kyoko Ohara*

JMWNL: an Extensible Multilingual Library for Accessing Wordnets in Different Languages  
*Maria Teresa Pazienza, Armando Stellato and Alexandra Tudorache*

Benchmarking Textual Annotation Tools for the Semantic Web  
*Diana Maynard*

Authorship Identification of Romanian Texts with Controversial Paternity  
*Liviu Dinu, Marius Popescu and Anca Dinu*



## **P12 - Lexical Resources and Tools**

Chairperson: Hitoshi Isahara

Ensuring Semantic Interoperability on Lexical Resources

*Marc Kemps-Snijders, Claus Zinn, Jacquelyn Ringersma and Menzo Windhouwer*

Exploring and Navigating: Tools for GermaNet

*Marc Finthammer and Irene Cramer*

A Knowledge-Modeling Approach for Multilingual Regulus Lexica

*Marianne Santaholma and Nikos Chatzichrisafis*

ODL: an Object Description Language for Lexical Information

*Michael Rosner*

How to Evaluate and Raise the Quality in a Collaborative Lexicographic Approach

*Dan Cristea, Corina Forascu, Marius Raschip and Michael Zock*

Merging a Syntactic Resource with a WordNet: a Feasibility Study of a Merge between STO and DanNet

*Bolette Sandford Pedersen, Anna Braasch, Lina Henriksen, Sussi Olsen and Claus Povlsen*

Hydra: a Modal Logic Tool for Wordnet Development, Validation and Exploration

*Borislav Rizov*

---

21:00

**Welcome Reception**