

---

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

---

9:00-9:40 **Keynote Speech**

**Ricardo Baeza-Yates: The Web as a Semantic Source**

Chairperson: Jan Odijk

*Ministres*

9:40-9:45 *5-minute Break*

9:45-11 :25 **4 Oral Sessions and 3 Poster Sessions in Parallel**

**O34 - Emotions**

Chairperson: Claire Cardie

*Ministres*

9:45-10:05 Resources for Persuasion

*Marco Guerini, Carlo Strapparava and Oliviero Stock*

10:05-10:25 Semi-automatic Building Method for a Multidimensional Affect Dictionary for a New Language

*Guillaume Pitel and Gregory Grefenstette*

10:25-10:45 Coding Emotional Events in Audiovisual Corpora

*Laurence Devillers and Jean-Claude Martin*

10:45-11:05 Annotating Expressions of Opinion and Emotion in the Italian Content Annotation Bank

*Andrea Esuli, Fabrizio Sebastiani and Ilaria Urciuoli*

11:05-11:25 Adjectives in the Dutch Semantic Lexical Database CORNETTO

*Isa Maks, Piek Vossen, Roxane Segers and Hennie van der Vliet*

**O35 - Semantics and Semantic Annotation**

Chairperson: Stelios Piperidis

*Ambassadeurs*

9:45-10:05 Detecting Errors in Semantic Annotation

*Markus Dickinson and Chong Min Lee*

10:05-10:25 Corpus Co-Occurrence, Dictionary and Wikipedia Entries as Resources for Semantic Relatedness Information

*Michael Roth and Sabine Schulte im Walde*

10:25-10:45 Ontology Learning and Semantic Annotation: a Necessary Symbiosis

*Emiliano Giovannetti, Simone Marchi, Simonetta Montemagni and Roberto Bartolini*

10:45-11:05 Semantically Annotated Snapshot of the English Wikipedia

*Jordi Atserias, Hugo Zaragoza, Massimiliano Ciaramita and Giuseppe Attardi*

11:05-11:25 Annotating Students' Understanding of Science Concepts

*Rodney D. Nielsen, Wayne Ward, James Martin and Martha Palmer*

**O36 - Evaluation Methodologies**

Chairperson: Keith Miller

*Fez 1*

9:45-10:05 Relation between Agreement Measures on Human Labeling and Machine Learning Performance: Results from an Art History Domain

*Rebecca Passonneau, Tom Lippincott, Tae Yano and Judith Klavans*

10:05-10:25 The Construction and Evaluation of Word Space Models

*Yves Peirsman, Simon De Deyne, Kris Heylen and Dirk Geeraerts*

10:25-10:45 Annotation of Nuggets and Relevance in GALE Distillation Evaluation

*Olga Babko-Malaya*

10:45-11:05 Statistical Evaluation of Information Distillation Systems

*J.V. White, D. Hunter and J.D. Goldstein*

11:05-11:25 Automatic Learning and Evaluation of User-Centered Objective Functions for Dialogue System Optimisation

*Verena Rieser and Oliver Lemon*

**O37 - Lexicons, Corpora and Acquisition**

Chairperson: Eva Hajičová

*Fez 2*

9:45-10:05 Building the Valency Lexicon of Arabic Verbs

*Viktor Bielický and Otakar Smrž*

10:05-10:25 Combining Terminology Resources and Statistical Methods for Entity Recognition: an Evaluation

*Angus Roberts, Robert Gaizasukas, Mark Hepple and Yikun Guo*

- 10:25-10:45 A Suite to Compile and Analyze an LSP Corpus  
*Rogelio Nazar, Jorge Vivaldi and Teresa Cabré*
- 10:45-11:05 Causal Relation Extraction  
*Eduardo Blanco, Nuria Castell and Dan Moldovan*
- 11:05-11:25 Learning Morphology with Morfette  
*Grzegorz Chrupala, Georgiana Dinu and Josef van Genabith*
- 9:45-11:25 **Poster Sessions Menara**

## **P22 - Speech Corpus in Various Environments**

Chairperson: Christoph Draxler

STC-TIMIT: Generation of a Single-channel Telephone Corpus

*Nicolas Morales, Javier Tejedor, Javier Garrido, Jose Colas and Doroteo T. Toledano*

LILA: Cellular Telephone Speech Databases from Asia

*Eric Sanders, Asuncion Moreno, Herbert Troup, Lynette Melnar, Nurit Dekel, Breanna Gillies and Niklas Paulsson*

JURISDIC: Polish Speech Database for Taking Dictation of Legal Texts

*Grazyna Demenko, Stefan Grocholewski, Katarzyna Klessa, Jerzy Ogórkiewicz, Agnieszka Wagner, Marek Lange, Daniel Śledziński and Natalia Cylwik*

A Multi-sensor Speech Database with Applications towards Robust Speech Processing in hostile Environments

*Tomas Dekens, Yorgos Patsis, Werner Verhelst, Frédéric Beaugendre and François Capman*

The LECTRA Corpus - Classroom Lecture Transcriptions in European Portuguese

*Isabel Trancoso, Rui Martins, Helena Moniz, Ana Isabel Mata and M. Céu Viana*

ALC: Alcohol Language Corpus

*Florian Schiel, Christian Heinrich, Sabine Barfußler and Thomas Gilg*

Design of a Multimodal Database for Research on Automatic Detection of Severe Apnoea Cases

*Rubén Fernández, Luis A. Hernández, Eduardo López, José Alcázar, Guillermo Portillo and Doroteo T. Toledano*

Test Collections for Spoken Document Retrieval from Lecture Audio Data

*Tomoyosi Akiba, Kiyooki Aikawa, Yoshiaki Itoh, Tatsuya Kawahara, Hiroaki Nanjo, Hiromitsu Nishizaki, Norihito Yasuda, Yoichi Yamashita and Katunobu Itou*

In-car Speech Data Collection along with Various Multimodal Signals

*Akira Ozaki, Sunao Hara, Takashi Kusakawa, Chiyomi Miyajima, Takanori Nishino, Norihide Kitaoka, Katunobu Itou and Kazuya Takeda*

Developing Corpus of Japanese Classroom Lecture Speech Contents

*Masatoshi Tsuchiya, Satoru Kogure, Hiromitsu Nishizaki, Kengo Ohta and Seiichi Nakagawa*

The ATCOSIM Corpus of Non-Prompted Clean Air Traffic Control Speech

*Konrad Hofbauer, Stefan Petrik and Horst Hering*

The MoveOn Motorcycle Speech Corpus

*Thomas Winkler, Theodoros Kostoulas, Richard Adderley, Christian Bonkowski, Todor Ganchev, Joachim Köhler and Nikos Fakotakis*

Audio Database in Support of Potential Threat and Crisis Situation Management

*Stavros Ntalampiras, Ilyas Potamitis, Todor Ganchev and Nikos Fakotakis*

CallSurf: Automatic Transcription, Indexing and Structuration of Call Center Conversational Speech for Knowledge Extraction and Query by Content

*Martine Garnier-Rizet, Gilles Adda, Frederik Cailliau, Sylvie Guillemin-Lanne, Claire Waast-Richard, Lori Lamel, Stephan Vanni and Claire Waast-Richard*

New Telephone Speech Databases for French: a Children Database and an optimized Adult Corpus

*Djamel Mostefa and Arnaud Vallee*

## **P23 - Speech Corpus in Various Languages**

Chairperson: Briony Williams

Design and Data Collection for Spoken Polish Dialogs Database

*Krzysztof Marasek and Ryszard Gubrynowicz*

CORP-ORAL: Spontaneous Speech Corpus for European Portuguese

*Tiago Freitas and Fabíola Santos*

From Human Communication to Intelligent User Interfaces: Corpora of Spoken Estonian

*Tiit Hennoste, Olga Gerassimenko, Riina Kasterpalu, Mare Koit, Andriela Rääbis and Krista Strandson*

The Pronouncing Dictionary of Austrian German (AGPD) and the Austrian Phonetic Database (ADABA): Report on a large Phonetic Resources Database of the three Major Varieties of German

*Rudolf Muhr*

German Today: a really extensive Corpus of Spoken Standard German

*Caren Brinckmann, Stefan Kleiner, Ralf Knöbl and Nina Berend*

Corpus and Voices for Catalan Speech Synthesis

*Antonio Bonafonte, Jordi Adell, Ignasi Esquerra, Silvia Gallego, Asunción Moreno and Javier Pérez*

Developments of “Lëtzebuergesch” Resources for Automatic Speech Processing and Linguistic Studies

*Martine Adda-Decker, Thomas Pellegrini, Eric Bilinski and Gilles Adda*

## **P24 - Speech Corpus Design Methodology and Tools**

Chairperson: Robrecht Comeyne

Speech Errors on Frequently Observed Homophones in French: Perceptual Evaluation vs Automatic Classification  
*Rena Nemoto, Ioana Vasilescu and Martine Adda-Decker*

Creation of Learner Corpus and Its Application to Speech Recognition  
*Hiroki Yamazaki, Keisuke Kitamura, Takashi Harada and Seiichi Yamamoto*

Automatic Rich Annotation of Large Corpus of Conversational transcribed speech: the Chunking Task of the EPAC Project  
*Jean-Yves Antoine, Abdenour Mokrane and Nathalie Friburger*

Manual vs Assisted Transcription of Prepared and Spontaneous Speech  
*Thierry Bazillon, Yannick Estève and Daniel Luzzati*

Developing a Phonemic and Syllabic Frequency Inventory for Spontaneous Spoken Castilian Spanish and their Comparison to Text-Based Inventories

*Antonio Moreno Sandoval, Doroteo Torre Toledano, Raúl de la Torre, Marta Garrote and José M. Guirao*

Phone Segmentation Tool with Integrated Pronunciation Lexicon and Czech Phonetically Labelled Reference Database.  
*Petr Pollák, Jan Volín and Radek Skarnitzl*

Estimating Word Phonosemantics  
*Victoria Bobicev and Tatiana Zidraşco*

memasysco: XML schema based metadata management system for speech corpora  
*Joachim Gasch, Caren Brinckmann and Sylvia Dickgießer*

Comparing Set-Covering Strategies for Optimal Corpus Design  
*Jonathan Chevelu, Nelly Barbot, Olivier Boeffard and Arnaud Delhay*

Automatic Phoneme Segmentation with Relaxed Textual Constraints  
*Pierre Lanchantin, Andrew C. Morris, Xavier Rodet and Christophe Veaux*

IrcamCorpusTools: an Extensible Platform for Spoken Corpora Exploitation  
*Christophe Veaux, Gregory Beller and Xavier Rodet*

Linguistic Resources for Reconstructing Spontaneous Speech Text  
*Erin Fitzgerald and Frederick Jelinek*

Spock: a Spoken Corpus Client  
*Maarten Janssen and Tiago Freitas*

On the Durational Reduction of Repeated Mentions: Recency and Speaker Effects  
*Viktor Tron*

11:25-11:45 Coffee break

11:45-13:25 **4 Oral Sessions and 3 Poster Sessions in Parallel**

### **O38 - Ontologies**

Chairperson: Piek Vossen

#### **Ministres**

11:45-12:05 Corpus Exploitation from Wikipedia for Ontology Construction  
*Gaoying Cui, Qin Lu, Wenjie Li and Yirong Chen*

12:05-12:25 Development and Alignment of a Domain-Specific Ontology for Question Answering  
*Shiyan Ou, Viktor Pekar, Constantin Orăsan, Christian Spurk and Matteo Negri*

12:25-12:45 Unsupervised and Domain Independent Ontology Learning: Combining Heterogeneous Sources of Evidence  
*David Manzano-Macho, Asunción Gómez-Pérez and Daniel Borrajo*

12:45-13:05 L-ISA: Learning Domain Specific Isa-Relations from the Web  
*Alessandra Potrich and Emanuele Pianta*

13:05-13:25 A Common Ground for Virtual Humans: Using an Ontology in a Natural Language Oriented Virtual Human Architecture  
*Arno Hartholt, Thomas Russ, David Traum, Eduard Hovy and Susan Robinson*

### **O39 - Multilingual Resources**

Chairperson: Zygmunt Vetulani

#### **Ambassadeurs**

11:45-12:05 Using the Multilingual Central Repository for Graph-Based Word Sense Disambiguation  
*Eneko Agirre and Aitor Soroa*

12:05-12:25 A Japanese-English Technical Lexicon for Translation and Language Research  
*Fredric Gey, David Evans and Noriko Kando*

12:25-12:45 Mutual Bilingual Terminology Extraction  
*Le An Ha, Gabriela Fernandez, Ruslan Mitkov and Gloria Corpas*

12:45-13:05 Building a Golden Collection of Parallel Multi-Language Word Alignment  
*Joao Graca, Joana Paulo Pardal, Luisa Coheur and Diamantino Caseiro*

13:05-13:25 The QALL-ME Benchmark: a Multilingual Resource of Annotated Spoken Requests for Question Answering  
*Elena Cabrio, Milen Kouylekov, Bernardo Magnini, Matteo Negri, Laura Hasler, Constantin Orasan, David Tomas, Jose Luis Vicedo, Guenter Neumann and Corinna Weber*

## **O40 - Tools for Corpus Construction and Annotation**

Chairperson: Martin Wynne

### **Fez 1**

- 11:45-12:05 Tools & Resources for Visualising Conversational-Speech Interaction  
*Nick Campbell*
- 12:05-12:25 A Web Browser Extension for Growing-up Ontological Knowledge from Traditional Web Content  
*Maria Teresa Pazienza, Marco Pennacchiotti and Armando Stellato*
- 12:25-12:45 A Development Environment for Configurable Meta-Annotators in a Pipelined NLP Architecture  
*Youssef Drissi, Branimir Boguraev, David Ferrucci, Paul Keyser and Anthony Levas*
- 12:45-13:05 Ontology-Based XQuery'ing of XML-Encoded Language Resources on Multiple Annotation Layers  
*Georg Rehm, Richard Eckart, Christian Chiarcos and Johannes Dellert*
- 13:05-13:25 A Lightweight and Efficient Tool for Cleaning Web Pages  
*Stefan Evert*

## **O41 - Speech Varieties**

Chairperson: Martine Adda-Decker

### **Fez 2**

- 11:45-12:05 Borrowing Language Resources for Development of Automatic Speech Recognition for Low- and Middle-Density Languages  
*Lynette Melnar and Chen Liu*
- 12:05-12:25 Corpus Analysis of Spoken Smart-Home Interactions with Older Users  
*Sebastian Möller, Florian Gödde and Maria Wolters*
- 12:25-12:45 A Fully Annotated Corpus for Studying the Effect of Cognitive Ageing on Users' Interactions with Spoken Dialogue Systems  
*Kallirroi Georgila, Maria Wolters, Vasilis Karaiskos, Melissa Kronenthal, Robert Logie, Neil Mayo, Johanna Moore and Matt Watson*
- 12:45-13:05 Recording Speech of Children, Non-Natives and Elderly People for HLT Applications: the JASMIN-CGN Corpus.  
*Catia Cucchiarini, Joris Driesen, Hugo Van hamme and Eric Sanders*
- 13:05-13:25 F0 of Adolescent Speakers - first Results for the German Ph@ttSessionz Database  
*Christoph Draxler, Florian Schiel and Tania Ellbogen*
- 11:45-13:25 **Poster Sessions Menara**

## **P25 - Morphology and Morphosyntax**

Chairperson: Marko Tadic

- A Question Answering System for German. Experiments with Morphological Linguistic Resources  
*Florian Koehler, Hinrich Schuetze and Michaela Atterer*
- Lexical Resources for Automatic Translation of Constructed Neologisms: the Case Study of Relational Adjectives  
*Bruno Cartoni*
- A Proper Approach to Japanese Morphological Analysis: Dictionary, Model, and Evaluation  
*Yasuharu Den, Junpei Nakamura, Toshinobu Ogiso and Hideki Ogura*
- Word-Based or Morpheme-Based? Annotation Strategies for Modern Hebrew Clitics  
*Reut Tsarfaty and Yoav Goldberg*
- Experimental Fast-Tracking of Morphological Analysers for Nguni Languages  
*Sonja Bosch, Laurette Pretorius, Kholisa Podile and Axel Fleisch*
- Generating a Morphological Lexicon of Organization Entity Names  
*Nikola Ljubetic, Tomislava Lauc and Damir Boras*
- Designing and Evaluating a Russian Tagset  
*Serge Sharoff, Mikhail Kopotev, Tomaz Erjavec, Anna Feldman and Dagmar Divjak*
- Czech MWE Database  
*Karel Pala, Lukáš Svoboda and Pavel Šmerk*
- Identification of Naturally Occurring Numerical Expressions in Arabic  
*Nizar Habash and Ryan Roth*
- A Dependency Parser for Thai  
*Shisanu Tongchim, Randolph Altmeyer, Virach Sornlertlamvanich and Hitoshi Isahara*
- A Hybrid Morphology-Based POS Tagger for Persian  
*Mehrnoush Shamsfard and Hakimeh Fadaee*
- A Common Parts-of-Speech Tagset Framework for Indian Languages  
*Baskaran Sankaran, Kalika Bali, Monojit Choudhury, Tanmoy Bhattacharya, Pushpak Bhattacharyya, Girish Nath Jha, S. Rajendran, K. Saravanan, L. Sobha and K.V. Subbarao*

## **P26 - Semantics, Semantic Resources and Semantic Annotation**

Chairperson: Costanza Navarretta

- Lexical Resources for Semantics Extraction  
*Rajat Mohanty and Pushpak Bhattacharyya*
- Evolutionary Basic Notions for a Thematic Representation of General Knowledge  
*Alain Joubert and Mathieu Lafourcade*

The Extended Architecture of Hantology for Japan Kanji  
*Ya-Min Chou, Chu-Ren Huang and Jia-Fei Hong*

Language Resources for Semantic Document Annotation and Crosslingual Retrieval  
*Petya Osenova, Kiril Simov and Eelco Mossel*

Using a Probabilistic Model of Context to Detect Word Obfuscation  
*Sanaz Jabbari, Ben Allison and Louise Guthrie*

Frame Information Transfer from English to Italian  
*Sara Tonelli and Emanuele Pianta*

Towards Spanish Verbs' Selectional Preferences Automatic Acquisition: Semantic Annotation of the SenSem Corpus  
*Jordi Carrera, Irene Castellón, Salvador Climent and Marta Coll-Florit*

Using Lexical Acquisition to Enrich a Predicate Argument Reusable Database  
*Paula Cristina Vaz, David Martins de Matos and Nuno J. Mamede*

Language Resources for Studying Argument  
*Chris Reed, Raquel Mochales Palau, Glenn Rowe and Marie-Francine Moens*

A Linguistic Resource for Discovering Event Structures and Resolving Event Coreference  
*Cosmin Bejan and Sanda Harabagiu*

Lexicon, Grammar, and Multilinguality in the Japanese FrameNet  
*Kyoko Hirose Ohara*

More Semantic Links in the SIMPLE-CLIPS Database  
*Nilda Ruimy and Antonio Toral*

Simple-Clips Ongoing Research: more Information with Less Data by Implementing Inheritance  
*Riccardo Del Gratta, Nilda Ruimy and Antonio Toral*

Linguistically Light Lexical Extensions for Ontologies  
*Brian Davis, Siegfried Handschuh, Alexander Trousov, John Judge and Mikhail Sogrin*

### **P27 - Temporal Annotation**

Chairperson: Dan Cristea

Automatic Identification of Temporal Information in Tourism Web Pages  
*Stéphanie Weiser, Philippe Laublet and Jean-Luc Minel*

Tapping Huge Temporally Indexed Textual Resources with WCTAnalyze  
*Sebastian Gottwald, Matthias Richter, Gerhard Heyer and Gerik Scheuermann*

Spatiotemporal Annotation Using MiniSTEx: how to deal with Alternative, Foreign, Vague and/or Obsolete Names?  
*Ineke Schuurman*

An Empirical Approach to a Preliminary Successful Identification and Resolution of Temporal Expressions in Spanish News Corpora  
*María Teresa Vicente-Díez, Doaa Samy and Paloma Martínez*

Annotation of WordNet Verbs with TimeML Event Classes  
*Georgiana Puscasu and Verginica Barbu Mititelu*

13:25-14:40 Lunch break

14:40-16:00 **1 Panel, 3 Oral Sessions and 1 Poster Session in Parallel**

14:40-16:00 **Panel: Why Language Resources are not “free” today - production, distribution, maintenance issues - The role and mission of Data centers**

Moderators: Khalid Choukri and Christopher Cieri

*Ministres*

### **O42 - Multimodal Dialogue**

Chairperson: Kristiina Jokinen

*Ambassadeurs*

14:40-15:00 Dialogue, Speech and Images: the Companions Project Data Set  
*Yorick Wilks, David Benyon, Christopher Brewster, Pavel Ircing and Oli Mival*

15:00-15:20 Creating and Using a Correlated Corpus to Glean Communicative Commonalities  
*Jade Goldstein-Stewart, Kerri Goodwin, Roberta Sabin and Ransom Winder*

15:20-15:40 Information Extraction Tools and Methods for Understanding Dialogue in a Companion  
*Roberta Catizone, Alexiei Dingli, Hugo Pinto and Yorick Wilks*

15:40-16:00 Production in a Multimodal Corpus: how Speakers Communicate Complex Actions  
*Carlos Gómez Gallo, T. Florian Jaeger, James Allen and Mary Swift*

### **O43 – Semantic Resources**

Chairperson: Gregory Grefenstette

*Fez 1*

14:40-15:00 Towards Formal Interpretation of Semantic Annotation  
*Harry Bunt and Chwhynny Overbeeke*

15:00-15:20 Towards a Vector Space Model for FrameNet-like Resources  
*Marco Pennacchiotti, Diego De Cao, Paolo Marocco and Roberto Basili*

- 15:20-15:40 KnoFusius: a New Knowledge Fusion System for Interpretation of Gene Expression Data  
*Pavel Smrz*
- 15:40-16:00 Modelling Word Similarity: an Evaluation of Automatic Synonymy Extraction Algorithms.  
*Kris Heylen, Yves Peirsman, Dirk Geeraerts and Dirk Speelman*

#### **O44 - Corpora and Evaluation Resources**

Chairperson: Takenobu Tokunaga

**Fez 2**

- 14:40-15:00 Children's Oral Reading Corpus (CHOREC): Description and Assessment of Annotator Agreement  
*Leen Cleuren, Jacques Duchateau, Pol Ghesquière and Hugo Van hamme*
- 15:00-15:20 A Bilingual Corpus of Inter-linked Events  
*Tommaso Caselli, Nancy Ide and Roberto Bartolini*
- 15:20-15:40 New Resources for Document Classification, Analysis and Translation Technologies  
*Stephanie Strassel, Lauren Friedman, Safa Ismael and Linda Brandschain*
- 15:40-16:00 Approximating Learning Curves for Active-Learning-Driven Annotation  
*Tomanek Katrin and Udo Hahn*
- 14:40-16:00 **Poster Sessions Asni Room**

#### **P28 - Multilinguality and Machine Translation**

Chairperson: Elliott Macklovitch

Automatic Translation of Biomedical Terms by Supervised Machine Learning

*Vincent Claveau*

Rapid Deployment of a New METIS Language Pair: Catalan-English

*Toni Badia, Maite Melero and Oriol Valentin*

Evaluation of a Machine Translation System for Low Resource Languages: METIS-II

*Vincent Vandeghinste, Peter Dirix, Ineke Schuurman, Stella Markantonatou, Sokratis Sofianopoulos, Marina Vassiliou, Olga Yannoutsou, Toni Badia, Maite Melero, Gemma Boleda, Michael Carl and Paul Schmidt*

Using Reordering in Statistical Machine Translation based on Alignment Block Classification

*Marta R. Costa-jussà, José A. R. Fonollosa and Enric Monte*

Evaluation of Linguistics-Based Translation

*Janne Bondi Johannessen, Torbjørn Nordgård and Lars Nygaard*

Word Alignment Annotation in a Japanese-Chinese Parallel Corpus

*Yujie Zhang, Zhulong Wang, Kiyotaka Uchimoto, Qing Ma and Hitoshi Isahara*

Selection of Japanese-English Equivalents by Integrating High-quality Corpora and Huge Amounts of Web Data

*Qing Ma, Koichi Nakao, Masaki Murata and Hitoshi Isahara*

Swedish-Turkish Parallel Treebank

*Beáta Megyesi, Bengt Dahlqvist, Eva Pettersson Eva and Joakim Nivre*

Sentence Alignment in DPC: Maximizing Precision, Minimizing Human Effort

*Julia Trushkina, Lieve Macken and Hans Paulussen*

Automatic Construction of a Japanese-Chinese Dictionary via English

*Hiroyuki Kaji, Shin'ichi Tamamura and Dashtseren Erdenebat*

Identification of Comparable Argument-Head Relations in Parallel Corpora

*Kathrin Spreyer, Jonas Kuhn and Bettina Schrader*

Corpus-Based Tools for Computer-Assisted Acquisition of Reading Abilities in Cognate Languages

*Svitlana Kurella, Serge Sharoff and Anthony Hartley*

Synchronizing Translated Movie Subtitles

*Jörg Tiedemann*

Constructing a Corpus that Indicates Patterns of Modification between Draft and Final Translations by Human Translators

*Takeshi Abekawa and Kyo Kageura*

Building a Bilingual Representation of the Roget Thesaurus for French to English Machine Translation

*Violaine Prince and Jacques Chauché*

Generating Bilingual Dictionaries by Transitivity

*Luka Nerima and Eric Wehrli*

Holy Moses! Leveraging Existing Tools and Resources for Entity Translation

*Jean Tavernier, Rosa Cowan and Michelle Vanni*

Linguistic Structure and Bilingual Informants Help Induce Machine Translation of Lesser-Resourced Languages

*Christian Monson, Ariadna Font Llitjós, Vamsi Ambati, Lori Levin, Alon Lavie, Alison Alvarez, Roberto Aranovich, Jaime Carbonell, Robert Frederking, Erik Peterson and Katharina Probst*

Creating Sentence-Aligned Parallel Text Corpora from a Large Archive of Potential Parallel Text using BITS and Champollion

*Kazuaki Maeda, Xiaoyi Ma and Stephanie Strassel*

Application of Resource-based Machine Translation to Real Business Scenes

*Hitoshi Isahara, Masao Utiyama, Eiko Yamamoto, Akira Terada and Yasunori Abe*

Building a Multilingual Lexical Resource for Named Entity Disambiguation, Translation and Transliteration

*Wolodja Wentland, Johannes Knopp, Carina Silberer and Matthias Hartung*

Translation-oriented Word Sense Induction Based on Parallel Corpora

*Marianna Apidianaki*

Smarty - Extendable Framework for Bilingual and Multilingual Comprehension Assistants

*Todor Arnaudov and Ruslan Mitkov*

Parallel Creation of Gigaword Corpora for Medium Density Languages - an Interim Report

*Péter Halácsy, András Kornai, Péter Németh and Dániel Varga*

MTriage: Web-enabled Software for the Creation, Machine Translation, and Annotation of Smart Documents

*Reginald Hobbs, Jamal Laoudi and Clare Voss*

Exploitation of an Arabic Language Resource for Machine Translation Evaluation: using Buckwalter-based Lookup Tool to Augment CMU Alignment Algorithm

*Clare Voss, Jamal Laoudi and Jeffrey Micher*

A Trainable Tokenizer, solution for multilingual texts and compound expression tokenization

*Oana Frunza*

Low-Density Language Bootstrapping: the Case of Tajiki Persian

*Karine Megerdumian and Dan Parvaz*

16:00-16:20 *Coffee break*

16:20-17:20 **4 Oral Sessions and 2 Poster Sessions in Parallel**

#### **O45 - Lexicons**

Chairperson: Monica Monachini

##### ***Ministres***

16:20-16:40 Lexicon Schemas and Related Data Models: when Standards Meet Users

*Thorsten Trippel, Michael Maxwell, Greville Corbett, Cambell Prince, Christopher Manning, Stephen Grimes and Steve Moran*

16:40-17:00 LexSchem: a Large Subcategorization Lexicon for French Verbs

*Cédric Messiant, Thierry Poibeau and Anna Korhonen*

17:00-17:20 Arabic WordNet: Semi-automatic Extensions using Bayesian Inference

*Horacio Rodriguez, David Farwell, Javi Farreres, Manuel Bertran, Musa Alkhalifa and M. Antonia Marti*

#### **O46 - Evaluation Methodologies**

Chairperson: Anthony Hartley

##### ***Ambassadeurs***

16:20-16:40 Subjective Evaluation of an Emotional Speech Database for Basque

*Iñaki Sainz, Ibon Saratzaga, Eva Navas, Inmaculada Hernández, Jon Sanchez, Iker Luengo and Igor Odriozola*

16:40-17:00 How to Compare Treebanks

*Sandra Kübler, Wolfgang Maier, Ines Rehbein and Yannick Versley*

17:00-17:20 The InFile Project: a Crosslingual Filtering Systems Evaluation Campaign

*Romarc Besançon, Stéphane Chaudiron, Djamel Mostefa, Ismaïl Timimi and Khalid Choukri*

#### **O47 - Authoring Tools and Corpora**

Chairperson: Nizar Habash

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

16:20-16:40 DIAC+: a Professional Diacritics Recovering System

*Dan Tufiş and Alexandru Ceauşu*

16:40-17:00 Annotating an Arabic Learner Corpus for Error

*Ghazi Abuhakema, Reem Faraj, Anna Feldman and Eileen Fitzpatrick*

17:00-17:20 All, and only, the Errors: more Complete and Consistent Spelling and OCR-Error Correction Evaluation

*Martin Reynaert*

#### **O48 - TV and Video Processing**

Chairperson: Michael Kipp

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

16:20-16:40 Using Movie Subtitles for Creating a Large-Scale Bilingual Corpora

*Einav Itamar and Alon Itai*

16:40-17:00 The IFADV Corpus: a Free Dialog Video Corpus

*R.J.J.H. van Son, Wieneke Wesseling, Eric Sanders and Henk van den Heuvel*

17:00-17:20 WOZ Acoustic Data Collection for Interactive TV

*Alessio Brutti, Luca Cristoforetti, Walter Kellermann, Lutz Marquardt and Maurizio Omologo*

16:20-17:20 **Poster Sessions Menara**

#### **P29 - Semantic Resources and their Elicitation**

Chairperson: Christopher Brewster

Enriching GermaNet with verb-noun relations - a case study of lexical acquisition

*Lothar Lemnitzer, Holger Wunsch and Piklu Gupta*

What's in a Colour? Studying and Contrasting Colours with COMPARA  
*Diana Santos, Maria do Rosário Silva and Susana Inácio*

A Multilingual Database of Polarity Items  
*Beata Trawinski and Jan-Philipp Soehn*

Integrating Metaphor Information into RDF/OWL EuroWordNet  
*Ernesto William De Luca and Birte Lönneker-Rodman*

Comparing Dependency and Constituent Syntax for Frame-semantic Analysis  
*Richard Johansson and Pierre Nugues*

AnCora-Verb: A Lexical Resource for the Semantic Annotation of Corpora  
*Juan Aparicio, Mariona Taulé and M. Antònia Martí*

Geo-WordNet: Automatic Georeferencing of WordNet  
*Davide Buscaldi and Paolo Rosso*

Domain-Specific English-To-Spanish Translation of FrameNet  
*Mario Crespo Miguel and Paul Buitelaar*

Enriching Frame Semantic Resources with Dependency Graphs  
*Hagen Fürstenau*

The Automatic Mapping of Princeton WordNet Lexical-Conceptual Relations onto the Brazilian Portuguese WordNet Database  
*Bento Carlos Dias-da-Silva, Ariani Di Felippo and Maria das Graças Volpe Nunes*

Semantic Role Labeling Tools Trained on the Cast3LB-CoNNL-SemRol Corpus  
*Roser Morante*

Building a Greek corpus for Textual Entailment  
*Evi Marzelou, Maria Zourari, Voula Giouli and Stelios Piperidis*

Extraction of Attribute Concepts from Japanese Adjectives  
*Kyoko Kanzaki, Francis Bond, Noriko Tomuro and Hitoshi Isahara*

Mapping Events and Abstract Entities from PAROLE-SIMPLE-CLIPS to ItalWordNet  
*Adriana Roventini and Nilda Ruimy*

Supersense Tagger for Italian  
*Davide Picca, Alfio Massimiliano Gliozzo and Massimiliano Ciaramita*

Clustering of Terms from Translation Dictionaries and Synonyms Lists to Automatically Build more Structured Linguistic Resources  
*Maria Teresa Pazienza and Armando Stellato*

Linguistic Description and Automatic Extraction of Definitions from German Court Decisions  
*Stephan Walter*

Hungarian Word-Sense Disambiguated Corpus  
*Veronika Vincze, György Szarvas, Attila Almási, Dóra Szauter, Róbert Ormándi, Richárd Farkas, Csaba Hatvani and János Csirik*

Semantic Annotation Layer in Russian National Corpus: Lexical Classes of Nouns and Adjectives  
*Olga N. Lashevskaja and Olga Yu. Shemanaeva*

A Compact Arabic Lexical Semantics Language Resource Based on the Theory of Semantic Fields  
*Mohamed Attia, Mohsen Rashwan, Ahmed Ragheb, Mohamed Al-Badrashiny and Husein Al-Basoumy*

Pragmatic Annotation of Discourse Markers in a Multilingual Parallel Corpus (Arabic- Spanish-English)  
*Doaa Samy and Ana González-Ledesma*

Building an Annotated Corpus for Text Summarization and Question Answering  
*Patcharee Varasai, Chaveevan Pechsiri, Thanu Sukvari, Vee Satayamas and Asanee Kawtrakul*

### **P30 -Sentiment and Opinion Analysis**

Chairperson: Diana Inkpen

A Multi-Lingual Dictionary of Dirty Words  
*Jonas Sjöbergh and Kenji Araki*

What is poorly Said is a Little Funny  
*Jonas Sjöbergh and Kenji Araki*

Building Affective Lexicons from Specific Corpora for Automatic Sentiment Analysis  
*Yves Bestgen*

Opinion Annotation in On-line Chinese Product Reviews  
*Ruifeng Xu, Yunqing Xia, Kam-Fai Wong and Wenjie Li*

Fine-grained Opinion Topic and Polarity Identification  
*Xiwen Cheng and Feiyu Xu*

Sentiment Analysis Based on Probabilistic Models Using Inter-Sentence Information  
*Kugatsu Sadamitsu, Satoshi Sekine and Mikio Yamamoto*

Valentino: A Tool for Valence Shifting of Natural Language Texts  
*Marco Guerini, Carlo Strapparava and Oliviero Stock*

17:20-17:25 *5-minute Break*

17:25-17:55

**Closing Session by Programme Committee *Ministres***

21:00

**Gala Dinner**