

Cross-Language Knowledge Induction Workshop

International Workshop held as part of

EACL 2006

The 11th Conference of the European Chapter of the Association for
Computational Linguistics

April 3, 2006

Trento, Italy

The conference, the workshop and the tutorials are sponsored by:



Center for the Evaluation of Language and Communication Technologies

Celct
c/o BIC, Via dei Solteri, 38
38100 Trento, Italy
<http://www.celct.it>

XEROX

Research Centre Europe

Xerox Research Centre Europe
6 Chemin de Maupertuis
38240 Meylan, France
<http://www.xrce.xerox.com>



CELI s.r.l.
Corso Moncalieri, 21
10131 Torino, Italy
<http://www.celi.it>

THALES

Thales
45 rue de Villiers
92526 Neuilly-sur-Seine Cedex, France
<http://www.thalesgroup.com>

EACL-2006 is supported by

Trentino S.p.a.  and Metasistem Group 

© April 2006, Association for Computational Linguistics

Order copies of ACL proceedings from:
Priscilla Rasmussen,
Association for Computational Linguistics (ACL),
3 Landmark Center,
East Stroudsburg, PA 18301 USA

Phone +1-570-476-8006
Fax +1-570-476-0860
E-mail: acl@aclweb.org
On-line order form: <http://www.aclweb.org/>

Preface

Knowledge of the behavior of words and text in other languages has recently been used to help solving tasks in a first language. An example of such a task is word-sense disambiguation by using translations in a second language. Another example is verb classification by studying properties of verbs in several languages.

A second modality of knowledge transfer across languages is to take advantage of resources already built for English and for a few other resource-rich languages. These resources have been used to induce knowledge in languages for which few linguistic resources are available. This was made possible by the wider availability of parallel corpora and better alignment methods at paragraph, sentence, and word level. Examples of such knowledge induction tasks are learning morphology, part-of-speech tags and grammatical gender. Cross-language knowledge transfer has also been possible thanks to the development of wordnets aligned to the original Princeton WordNet.

This workshop provides a forum for discussion between leading names and researchers involved in cross-language applications. Topics of interest include, but are not limited to: applications that exploit parallel corpora (learning morphological segmentation; learning part-of-speech; learning grammatical gender; other applications); induction of knowledge from a language for which resources are abundant to another language for which fewer resources are available; using other languages to solve a task in a first language (word-sense disambiguation by using translations in other languages; verb classification by studying verb properties in several languages; other tasks of this kind); identifying and using cognate words between languages; building wordnets by knowledge transfer; and exploiting multi-language wordnets for NLP applications.

We would like to thank all the authors who submitted papers for the hard work that went behind their submissions. We express our deepest gratitude to the committee members for their thorough reviews. We also thank the EACL 2006 organizers for their help with administrative matters.

Diana Inkpen
Carlo Strapparava
Eneko Agirre

Organizing committee

Diana Inkpen (University of Ottawa, Canada)
Carlo Strapparava (ITC-IRST, Povo-Trento, Italy)
Eneko Agirre (University of the Basque Country, Donostia, Spain)

Program Committee

Paul Buitelaar (DFKI, Saarbrücken, Germany)
Silviu Cucerzan (Microsoft Research, US)
Mona Diab (Columbia University, US)
Greg Kondrak (University of Alberta, Canada)
Lluís Màrquez (Universitat Politècnica de Catalunya, Barcelona, Spain)
Joel Martin (National Research Council of Canada)
Rada Mihalcea (University of North Texas, US)
Viviana Nastase (University of Ottawa, Canada)
Ted Pedersen (University of Minnesota, Duluth, US)
Emanuele Pianta (ITC-IRST, Povo-Trento, Italy)
Philip Resnik (University of Maryland, US)
German Rigau (University of the Basque Country, Donostia, Spain)
Laurent Romary (LORIA, Nancy, France)
Michel Simard (National Research Council of Canada)
Suzanne Stevenson (University of Toronto, Canada)
Doina Tatar (Babes-Bolyai University, Cluj-Napoca, Romania)
Amalia Todirascu (Université Marc Bloch, Strasbourg, France)
Dan Tufis (Romanian Academy, Bucharest, Romania)
Nikolai Vazov (University of Sofia, Bulgaria)

Invited Speakers

Mona Diab (Columbia University, US)
Dan Tufis (Romanian Academy, Bucharest, Romania)

Workshop Program

- 8:50 – 9:00 *Welcome and Introduction*
- 9:00 – 10:00 *Invited talk: Mona Diab*
- 10:00 – 10:30 *Multilingual Extension of a Temporal Expression Normalizer using Annotated Corpora*
E. Saquete, P. Martínez-Barco, R. Muñoz, M. Negri, M. Speranza and R. Sprugnoli
- 10:30 – 11:00 Break
- 11:00 – 11:30 *A Framework for Incorporating Alignment Information in Parsing*
Mark Hopkins and Jonas Kuhn
- 11:30 – 12:00 *Induction of Cross-Language Affix and Letter Sequence Correspondence*
Ari Rappoport and Tsahi Levent-Levi
- 12:00 – 12:30 *Improving Name Discrimination: A Language Salad Approach*
Ted Pedersen, Anagha Kulkarni, Roxana Angheluta, Zornitsa Kozareva,
and Tamar Solorio
- 12:30 – 14:30 Lunch
- 14:30 – 15:30 *Invited talk: Dan Tufis*
- 15:30 – 16:00 *Tagging Portuguese with a Spanish Tagger*
Jirka Hana, Anna Feldman, Luiz Amaral, and Chris Brew
- 16:00 – 16:30 Break
- 16:30 – 17:00 *Automatic Generation of Translation Dictionaries Using Intermediary Languages*
Kisuh Ahn and Matthew Frampton
- 17:00 – 17:30 *Word Sense Disambiguation Using Automatically Translated Sense Examples*
Xinglong Wang and David Martinez
- 17:30 – 18:00 *Projecting POS Tags and Syntactic Dependencies from English and French to Polish
in Aligned Corpora*
Sylwia Ozdowska

Table of Contents

<i>Multilingual Extension of a Temporal Expression Normalizer using Annotated Corpora</i> E. Saquete, P. Martnez-Barco, R. Muoz, M. Negri, M. Speranza and R. Sprugnoli	1
<i>A Framework for Incorporating Alignment Information in Parsing</i> Mark Hopkins and Jonas Kuhn	9
<i>Induction of cross-language affix and letter sequence correspondence</i> Ari Rappoport and Tsahi Levent-Levi	17
<i>Improving Name Discrimination: A Language Salad Approach</i> Ted Pedersen, Anagha Kulkarni, Roxana Angheluta, Zornitsa Kozareva and Thamar Solorio	25
<i>Tagging Portuguese with a Spanish Tagger</i> Jirka Hana, Anna Feldman, Luiz Amaral and Chris Brew	33
<i>Automatic Generation of Translation Dictionaries Using Intermediary Languages</i> Kisuh Ahn and Matthew Frampton	41
<i>Word Sense Disambiguation Using Automatically Translated Sense Examples</i> Xinglong Wang and David Martinez	45
<i>Projecting POS tags and syntactic dependencies from English and French to Polish in aligned corpora</i> Sylwia Ozdowska	53