ACL-04

42nd Annual Meeting of the Association for Computational Linguistics

Proceedings of the Conference

21–26 July , 2004
Barcelona, Spain

Preface to the Student Research Workshop Proceedings

We are pleased to introduce the papers accepted for presentation at the Student Research Workshop of the *42nd Annual Meeting of the Association for Computational Linguistics* in Barcelona, Spain. We received 43 papers – a near-record number – of which we were able to accept 12. We thank all students who submitted and we hope the reviews provided useful feedback and directions for future research both for those presenting and for those whose papers we were not able to accept.

Many people contributed to the organization of the ACL-04 Student Research Workshop. First of all, we thank the many reviewers for the time they spent reading and reviewing the papers. A special thank you to those reviewers who helped us out with extra reviews, often on short notice, in the face of a number of submissions that surpassed our expectations.

The members of the ACL-04 Student Research Workshop Program Committee were:

Laura Alonso, *Universidad de Barcelona, Spain*
Jordi Atserias i Batalla, *Universidad Politecnica de Catalunya, Spain*
Amit Bagga, *Avaya Labs Research, USA*
John Beavers, *Stanford University, USA*
Francis Bond, *NTT, Japan*
Christina Bosco, *Turin University, Italy*
Gosse Bouma, *Rijksuniversiteit Groningen, Netherlands*
Thorsten Brants, *Google Inc, USA*
Eric de la Clergerie, *INRIA, France*
Eric Fosler-Lussier, *Ohio State University, USA*
Heidi Fox, *Brown University, USA*
Oren Glickman, *Bar Ilan University, Israel*
Michelle Gregory, *University of Buffalo, USA*
Jon Herring, *University of Brighton, UK*
Chu-Ren Huang, *Academia Sinica, Taiwan*
Alistair Knott, *University of Otago, New Zealand*
Milen Kouylekov, *ITC-irst, Trento, Italy*
Ashwani Kumar, *MIT, USA*
Alberto Lavelli, *ITC-irst, Trento, Italy*

Rob Malouf, *San Diego State University, USA*
Erwin Marsi, *University of Tilburg, Netherlands*
Fernando Martinez Santiago, *Universidad de Jaén, Spain*
Joakim Nivre, *Växjö University, Sweden*
Jahna Otterbacher, *University of Michigan, USA*
Sebastian Padó *Saarland University, Germany*
Bill Raymond, *Ohio State University, USA*
Hannah Rohde, *University of California–San Diego, USA*
Anoop Sarkar, *Simon Fraser University, Canada*
Jennifer Spenader, *Stockholm University, Sweden*
Paul Tepper, *Northwestern University, USA*
Kristina Toutanova, *Stanford University, USA*
Olga Uryupina, *Saarland University, Germany*
Gertjan van Noord, *Rijksuniversiteit Groningen, Netherlands*
Menno van Zaanen, *University of Tilburg, Netherlands*
Begoña Villada, *Rijksuniversiteit Groningen, Netherlands*
Stephen Wan, *Macquarie University, Australia*
Mirjam Wester, *University of Edinburgh, UK*
Maria Wolters, *Rhetorical Systems, UK*
Milena Yankova, *Bulgaria Academy of Sciences*

Twenty researchers have agreed to serve as respondents to the papers presented during the Student Research Workshop. We are thankful for their thoughtful comments which are of prime importance to this next generation of computational linguists.

Special thanks to Donia Scott for her input, to Claire Gardent for her helpful suggestions, and to our tireless faculty advisor Justine Cassell for her advice, support, and obtaining funding. Above all, we wish to thank the American National Science Foundation for providing financial support that has allowed the student presenters to travel to Barcelona.

ACL-04 Student Research Workshop Co-chairs:
Leonoor van der Beek, *University of Groningen, Netherlands*
Dmitriy Genzel, *Brown University, USA*
Daniel Midgley, *University of Western Australia, Australia*

Table of Contents

Programme for Poster/Demo session

Thursday, July 22

10:30-12:10 Short Presentations

13:40-17:30 Session 1

TransType2 - An Innovative Computer-Assisted Translation System
José Esteban, José Lorenzo, Antonio S. Valderrábanos and Guy Lapalme

Improving Domain-Specific Word Alignment for Computer Assisted Translation
Wu Hua and Wang Haifeng

Constructing Transliteration Lexicons from Web Corpora
Jin-Shea Kuo and Ying-Kuei Yang

Subsentential Translation Memory for Computer Assisted Writing and Translation
Jian-Cheng Wu, Thomas C. Chuang, Wen-Chi Shei and Jason S. Chang

Customizing Parallel Corpora at the Document Level
Monica Rogati and Yiming Yang

An Automatic Filter for Non-Parallel Texts
Chris Pike and I. Dan Melamed

Exploiting Aggregate Properties of Bilingual Dictionaries For Distinguishing Senses of English Words and Inducing English Sense Clusters
Charles Schafer and David Yarowsky

Interactive grammar development with WCDG
Kilian A. Foth, Michael Daum and Wolfgang Menzel

Wide Coverage Symbolic Surface Realization
Charles Callaway

Part-of-Speech Tagging Considering Surface Form for an Agglutinative Language
Do-Gil Lee and Hae-Chang Rim

Is Conceptual Combination Influenced by Word Order?
Phil Maguire and Arthur Cater

Corpus representativeness for syntactic information acquisition
Núria Bel

Exploiting Unannotated Corpora for Tagging and Chunking
Rie Kubota Ando

Improving Bitext Word Alignments via Syntax-based Reordering of English
Elliott Franco Drabek and David Yarowsky

Friday, July 23

8:45-10:00 Short Presentations

13:40-16:40 Session 2

Hierarchy Extraction based on Inclusion of Appearance
Eiko Yamamoto, Kyoko Kanzaki and Hitoshi Isahara

Knowledge intensive e-mail summarization in CARPANTA
Laura Alonso, Irene Castellón, Bernardino Casas and Lluís Padró

Finding Anchor Verbs for Biomedical IE Using Predicate-Argument Structures
Akane Yakushiji, Yuka Tateisi, Yusuke Miyao and Jun'ichi Tsujii

Resource Analysis for Question Answering
Lucian Vlad Lita, Warren A. Hunt and Eric Nyberg

TANGO: Bilingual Collocational Concordancer
Jia-Yan Jian, Yu-Chia Chang and Jason S. Chang

Graph-based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization
Rada Mihalcea

Compiling Boostexter Rules into a Finite-state Transducer
Srinivas Bangalore

Combining Lexical, Syntactic, and Semantic Features with Maximum Entropy Models for Information Extraction
Nanda Kambhatla

On the Equivalence of Weighted Finite-state Transducers
Julien Quint

A New Feature Selection Score for Multinomial Naive Bayes Text Classification Based on KL-Divergence
Karl-Michael Schneider

Saturday, July 24

9:00-10:20 Short Presentations

14:30-17:05 Session 3

Incorporating topic information into semantic analysis models
Tony Mullen and Nigel Collier

A Practical Solution to the Problem of Automatic Word Sense Induction
Reinhard Rapp

Automatic clustering of collocation for detecting practical sense boundary
Saim Shin and Key-Sun Choi

Co-training for Predicting Emotions with Spoken Dialogue Data
Beatriz Maeireizo, Diane Litman and Rebecca Hwa

Multimodal Database Access on Handheld Devices
Elsa Pecourt and Norbert Reithinger

Wysiwym with wider coverage
Richard Power and Roger Evans

NLTK: The Natural Language Toolkit
Steven Bird and Edward Loper

Dyna: A Language for Weighted Dynamic Programming
Jason Eisner, Eric Goldlust and Noah A. Smith

MATCHkiosk: A Multimodal Interactive City Guide
Michael Johnston and Srinivas Bangalore

Fragments and Text Categorization
Jan Blaták, Eva Mráková and Lubos Popelínsky