

SPONSOR:

CLIF (Computational Linguistics in Flanders)

INVITED SPEAKERS:

Eric Brill, Microsoft Research
John Nerbonne, University of Groningen

PROGRAM COMMITTEE:

Walter Daelemans, University of Antwerp and Tilburg University, Chair
Rémi Zajac, New Mexico State University, Chair

Thorsten Brants, Xerox PARC
Michael Brent, Washington University in Saint Louis
Claire Cardie, Cornell University
James Cussens, University of York
Hervé Déjean, University of Tuebingen, Shared Task coordinator
Gregory Grefenstette, Xerox Grenoble
Raymond Mooney, University of Texas at Austin
John Nerbonne, Groningen University
Kemal Oflazer, Sabanci University
Miles Osborne, University of Edinburgh
David Powers, Flinders University
Ronan Reilly, University College Dublin
Dan Roth, University of Illinois at Urbana-Champaign
Erik Tjong Kim Sang, University of Antwerp, Shared Task coordinator
Antal van den Bosch, Tilburg University
Yorick Wilks, University of Sheffield

CONFERENCE WEBSITE:

<http://lcg-www.uia.ac.be/conll2001/>

PREFACE

This volume contains the papers accepted for presentation at the 2001 workshop on Computational Natural Language Learning (CoNLL-2001), a SIGNLL meeting held in Toulouse, France, July 6 and 7 in conjunction with the International and the European Conferences of the ACL.

CoNLL-2001 is the fifth in a series of meetings organized by SIGNLL¹, the ACL special interest group on natural language learning. The CoNLL meeting is intended to address all aspects of computational natural language learning, including those that are not regularly discussed at computational linguistics meetings, such as computational models of human language acquisition, computational models of the origins and evolution of language, biologically-inspired learning methods etc.

This year, we have a special theme on learning mechanisms for the automated acquisition of language resources (dictionaries, ontologies, grammars) or the automated adaptation of natural language resources/processors to new domains or languages. The dimensions of learning that are of interest include the integration of humans/linguists in the learning process, the structure of the training data (especially, minimization of the training set), and the kind of knowledge that is learned.

This year's CoNLL also features two invited presentations, one by Eric Brill on whether the web changes empirical NLP, and one by John Nerbonne on the results of a successful European project on *Learning Computational Grammars*. As in the previous two editions of CoNLL, we have a shared task session, coordinated by Erik Tjong Kim Sang and Hervé Déjean. For this session, training and test data were made available, and researchers were invited to apply their learning system to this task. This year the task is clause identification. Six systems have participated, five from Europe and one from Australia. You will find descriptions of these systems and their results in this proceedings.

Of the 61 papers submitted, the programme committee selected 20 papers, representative of the state of the art in computational language learning today. We are very grateful to our programme committee for the effort they put in reviewing the papers. We are also grateful to the ACL/EACL-2001 conference organizers on whom we could rely for the local organization. Finally, we most gratefully acknowledge the support of our sponsor, the Computational Linguistics in Flanders research community (CLIF) for sponsoring one of the invited speakers.

Walter Daelemans & Rémi Zajac (editors)
May 2001

¹<http://www.aclweb.org/signll/>