

## Preface

The fourteenth meeting of the computational linguistics community in the Netherlands (CLIN 2003) took place at the University of Antwerp on December 19 of 2003. Over one hundred registered participants came to Antwerp to see more than 50 researchers present their work on a wide variety of subjects. The number of participants and submissions seems to increase with every CLIN meeting and this year was no exception, showing that interest in computational linguistics is still on the rise in our region. In fact, the number of submissions was so high this year, CLIN 2003 had to host a poster session to accommodate all of the presentations.

We were very happy to welcome this year's keynote speaker, Thorsten Brants from Google Inc., who provided a very interesting outline of the role that Natural Language Processing can play in information retrieval systems. Coming from an industrial background allowed him to critically assess the state of the art of NLP with respect to the information retrieval technology Google is working on. Thorsten was kind enough to take the time to prepare a written version of his insights, which kicks off these proceedings.

Thorsten's paper complements a varied selection of the numerous submissions received after the CLIN 2003 call for papers. We were very fortunate to have a list of skilled reviewers at our disposal to handle the refereeing of these papers. Gosse Bouma, Walter Daelemans, Maarten de Rijke, Koenraad de Smedt, Iris Hendrickx, Jean-Pierre Martens, Mark-Jan Nederhof, John Nerbonne, Jan Odijk, Marie-Laure Reinberger, Khalil Sima'an, Erik Tjong Kim Sang, Antal van den Bosch, Frank Van Eynde, Hans van Halteren, Gertjan van Noord, Menno van Zaanen and Patrick Wambacq all helped us to select the 11 most suited papers from the 23 submissions. We would like to thank them as well as the authors for all the hard work and for safeguarding the quality of the CLIN proceedings.

It is interesting to note that almost all papers in these proceedings deal with some form of corpus-based research. Some work uses corpus data or corpus-induced resources to find empirical proof for a linguistic status questionis (Bouma;van der Beek and Baldwin;Xanthos). Others use corpora to extract resources, such as a list of multi-word lexemes (Kis, Villada Moirón, Bíró, Bouma, Pohl, Ugray and Nerbonne). Not surprisingly the CLIN 2003 proceedings contain a large number of papers on parsing. You will find some work on the adjustment and extension of existing parsing algorithms (Prescher, Scha, Sima'an and Zollmann; Schelfhout and Coppen), as well as work on the application of parsing technology to adapt treebanks to a specific task (Tylman Ule and Jorn Veenstra). Another research effort uses recently made available data to construct a shallow parser for spoken Dutch (Canisius and van den Bosch). You will also find work on NLP applications in these proceedings, such as automatic detection of plagiarism in student essays (van Halteren) and sentence reduction in automatic subtitling (Tjong Kim Sang, Daelemans and Höthker). A text-to-speech application is used as a test-case for a new efficient algorithm for the incremental construction of a

minimal acyclic sequential transducer (Skut).

We would like to thank Walter Daelemans, Erik Tjong Kim Sang, Kim Luyckx and Barbara Buls for lending us a helping hand during the organization of CLIN 2003. We want to thank SIKS (School for Information and Knowledge Systems) and CLIF (Computational Linguistics in Flanders) for sponsoring the event. Publishing these proceedings would also not have been possible without the further financial aid of CLIF and the logistic framework of APIL (Antwerp Papers in Linguistics). Finally, we would like to extend our gratitude to all of the participants, presenters, submitters and reviewers of CLIN 2003.

Antwerpen, August 2004

*Bart Decadt, Veronique Hoste, Guy De Pauw*  
*CNTS - Language Technology Group*  
*University of Antwerp*

---

For more information on CLIN 2003, see:  
<http://cnts.uia.ac.be/clin2003/>

For more information about the CLIN workshop series, see:  
<http://www.let.rug.nl/vannoord/Clin/clin.html>