

# Constraint Grammar Applications

Eckhard Bick

University of Southern Denmark

This paper collection presents the contributions to the 2011 NoDaLiDa Constraint Grammar workshop.

The NoDaLiDa CG workshop has now seen 3 editions, and become the main physical forum for the exchange of ideas and results in the field of Constraint Grammar at the research level. As organizers we were pleased to note a growing number of participants and a consistently high quality of contributions. The last years have seen extensions of CG usage along several lines, both to the formalism as such, and in terms of areas of application. Thus, the CG3 formal language now allows the emulation of very diverse grammatical approaches, covering besides the traditional topological and dependency frameworks also the use of probabilistic, generative and unification techniques. Most strikingly, however, is the potential CG has shown in the applicational arena, where it is now successfully used across a wide range of languagetechnological issues, such as machine translation, grammar checking, dialogue systems and lexicography. Finally, CG continues to be used for the production of linguistic research and teaching resources, such as annotated corpora in general, and treebanks in particular.

The papers in this collection provide an insider's view on some of these developments. Annotation topics cover both treebanks (Voutilainen) and spoken data (Donnelly & Deuchar), and MT features prominently in the applicational area, touching both on methodology such as valency portability (Arriola & Wiechetek) and grand-scale projects such as Wikipedia translation (Bick). Within the field of CG theory, hybrid solutions (Johannessen et al.) and compiler optimization through FST methods (Peltonen) are explored. Parsing efficiency is also the central theme in Yli-Jyrä's paper presenting a full CG compiler implementation with an inward deterministic method.

Finally, it should be noted that CG's rule-based

approach continues to allow researchers to tackle also smaller languages, where the quality or indeed feasibility of machine-learning suffers from the lack of existing training resources. Thus, Basque (Arriola & Wiechetek), Sami (Antonsen & Trosterud) and Esperanto (Bick) are represented in this collection of workshop papers.

On behalf of the organizing team,  
Eckhard Bick

## The papers

Lene Antonsen and Trond Trosterud: *Next to nothing – a cheap South Saami disambiguator*

Eckhard Bick: *WikiTrans: The English Wikipedia in Esperanto*

Kevin Donnelly and Margaret Deuchar: *Using constraint grammar in the Bangor Autoglosser to disambiguate multilingual spoken text*

Janne Bondi Johannessen, Kristin Hagen, André Lynum, Anders Nøklestad: *OBT+Stat: Evaluation of a combined CG and statistical tagger*

Janne Peltonen: *A Finite State Constraint Grammar Parser*

Atro Voutilainen: *FinnTreeBank: Creating a research resource and service for language researchers with Constraint Grammar*

Anssi Yli-Jyrä: *An Efficient Constraint Grammar Parser based on Inward Deterministic Automata*

Linda Wiechetek and Jose Mari Arriola: *An Experiment of Use and Reuse of Verb Valency in Morphosyntactic Disambiguation and Machine Translation for Euskara and North Sámi*