

**CG-MTA 2023**

**Constraint Grammar – Methods, Tools, and Applications**

**Proceedings of the Workshop**

May 22, 2023

©2023 Association of Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 979-8-89176-059-2

Editors: Eckhard Bick, Trond Trosterud, Tanel Alumäe

## Preface

The Constraint Grammar Workshop at NoDaLiDa goes back to 2007 and is the main regular venue of the Constraint Grammar community, bringing together researchers from not only the Nordic region, but also eliciting interest from other countries. With the exception of the Covid year 2021, the workshop has been arranged at every NoDaLiDa since its inception, making it the eighth one in a row. The workshops have focused both on theoretical issues linked to the Constraint Grammar framework and on its uses in the detailed analysis of natural languages at various linguistic levels. In addition, contributions about Constraint Grammar-based applications have played an increasing role, with a growing number of papers dedicated to this area.

This tendency towards applications and real-life integration has been especially clear in the current edition of the workshop, that had an overweight on CG-based practical programs targeting proofing tools. Thus, 4 papers presented work on grammar checkers, covering *Faroese*, *Inari Saami*, *Lule Saami* and *South Saami*. All papers focused on a restricted set of error types, but in each case the error types represented high-frequency problems in the language in question. The Faroese contribution was about errors related to the letter *ð* in morphological suffixes, for Inari Saami the focus was on a specific set of interference errors from Finnish, while the Lule Saami article presented the broadest set of error types, with agreement phenomena being the common denominator. The last grammar checker paper, about South Saami, looked at two problematic parts of the grammar, adjective agreement and negative constructions. All four grammar checkers represented finished and tested work for the error types in question and were released during the workshop.

A further two papers presented applicative, CG-based programs, where a down-stream, higher level NLP task was solved using a dedicated Constraint Grammar rule set. Thus, the paper *Attribution of Quoted Speech in Portuguese Text* exploits new Constraint Grammar techniques dealing with long-distance relations, such as co-reference links spanning several sentences and the dynamic use of stream variables, to automatically annotate news and literary text for quoting and attribution constructions, using syntactic, semantic and pragmatic tags as clues to identify and classify these constructions and to link them to specific speaker IDs. Another applicative contribution was an ICALL paper that used Constraint Grammar for the automatic scoring of learner essays written in *Basque*, with scores expressed in terms of the European framework CEFR for language level assignment.

Finally, a more theoretical paper, *WITH Context: Adding Rule-Grouping to VISL CG-3*, addressed the CG rule formalism itself, to which it added a new operator, *WITH*, hereby opening up for a new rule type that would allow a more efficient grammar writing by fusing entire rule blocks with shared contexts into one, integrated rule.

On behalf of the workshop organizers,

Trond Trosterud and Eckhard Bick

# Organizing Committee

## Workshop organizers

Eckhard Bick, Research lector, Institute of Language and Communication, University of Southern Denmark

Tino Didriksen, Developer, GrammarSoft ApS; University of Southern Denmark

Kristin Hagen, Senior engineer, Tekstlaboratoriet, University of Oslo

Kaili Müürisep, Senior research fellow, Institute of Computer Science, University of Tartu

Trond Trosterud, Professor in Sámi computational linguistics, University of Tromsø

Linda Wiechetek, Senior engineer, University of Tromsø

## Program Committee / Reviewers

Eckhard Bick (Chair)

Edit Bugge

Lina Lejdebros Enwald

Kristin Hagen

Katri Hiovain-Asikainen

Kaili Müürisep

Anssi Yli-Jyrä

Flammie A. Pirinen

Jack Rueter

Daniel Glen Swanson

Trond Trosterud

Elaine Uí Dhonnchadha

Kevin Brubeck Unhammer

Linda Wiechetek

## Table of Contents

<i>Attribution of Quoted Speech in Portuguese Text</i> Eckhard Bick .....	1
<i>WITH Context: Adding Rule-Grouping to VISL CG-3</i> Daniel Swanson, Tino Didriksen and Francis M. Tyers .....	10
<i>To ð or not to ð - A Faroese CG-based grammar checker targeting ð errors</i> Trond Trosterud .....	15
<i>Towards automatic essay scoring of Basque language texts from a rule-based approach based on curriculum-aware systems</i> Jose Maria Arriola, Mikel Iruskieta, Ekain Arrieta and Jon Alkorta .....	20
<i>Correcting well-known interference errors – Towards a L2 grammar checker for Inari Saami</i> Trond Trosterud, Marja-Liisa Olthuis and Linda Wiechetek .....	29
<i>Supporting Language Users - Releasing a Full-fledged Lule Sámi Grammar Checker</i> Inga Lill Sigga Mikkelsen and Linda Wiechetek .....	37
<i>A South Sámi Grammar Checker For Stopping Language Change</i> Linda Wiechetek and Maja Lisa Kappfjell .....	46