

CG/FST-NLP 2025

**The 9th Workshop on Constraint Grammar and Finite State
NLP – Rule-based and hybrid methods and tools for user
communities**

Proceedings of the Workshop

March 5, 2025

©2025 University of Tartu Library, Estonia

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 978-9908-53-113-7

Introduction

We are delighted to invite you to CG-FST NLP 2025, the ninth NoDaLiDa workshop on Constraint Grammar (CG). Constraint grammars often take text analysed with the help of finite state transducers (FST), which is why we have also invited papers on FST this year. The conference is being held as a physical event only, on March 5th 2025, at the NoDaLiDa conference in Tallinn.

Constraint Grammar is a grammar formalism developed in the Nordic countries. The main open source implementation FSTs, Helsinki finite state transducer, or HFST, was made in Finland. It is thus only to be expected that this workshop series is a pendant to NoDaLiDa/Baltic-HLT. The first CG workshop was arranged at NoDaLiDa in Odense in 2009, and it has been arranged at every NoDaLiDa conference ever since, except for at the 23rd NoDaLiDa in Reykjavik in 2021 (which was held mostly as a virtual event, due to the COVID pandemic).

Topic-wise, the papers in this year's workshop may be divided into three groups: Five of the papers deal mainly with one language, and present a spellchecker (Horváth, Rueter and Trosterud), grammar checker (Bick; Denbæk) or grammatical analysis (Gerstenberger; Trosterud and Vonen). Four of the papers take a more general approach, and deal with the functionalities of Constraint grammar (Swanson; Wiechetek and Unhammer) or FST (Pirinen and Moshagen) as proofing tools. Finally, one paper (Torstensson and Holmström) is a hybrid paper investigating the role of (possibly CG) annotated data may play for LLMs. This year's workshop covers, not only a wide range of applications, but also different languages (Esperanto, Romanian, Mansi, Tokelau, Greenlandic, Lule Sámi, Irish).

We would like to thank the members of the program committee (Eckhard Bick, Tino Didriksen, Kaili Müürisep, Daniel Glen Swanson and Francis Tyers) for taking part in planning and organising the workshop. We would also like to thank the anonymous reviewers for reviewing the incoming papers. Without anonymous reviews there are no peer-reviewed proceedings, and their work is thus highly appreciated.

Trond Trosterud, General Chair

Linda Wiechetek and Flammie Pirinen, Program Co-Chairs

Organizing Committee

General Chair

Trond Trosterud, UiT The Arctic University of Norway

Program Chairs

Linda Wiechetek, UiT The Arctic University of Norway

Flammie Pirinen, UiT The Arctic University of Norway

Program Committee

Program Chairs

Eckhard Bick, University of Southern Denmark - SDU
Tino Didriksen, Oqaasileriffik
Kaili Müürisep, institute of computer science, University of Tartu
Flammie A Pirinen, Norgga árktaš universitehta
Daniel Glen Swanson, Indiana University
Trond Trosterud, University of Tromsø
Francis M. Tyers, Indiana University, Bloomington
Linda Wiechetek, University of Tromsø

Reviewers

Eckhard Bick, University of Southern Denmark - SDU
Tino Didriksen, Oqaasileriffik
Kimmo Koskenniemi, University of Helsinki
Inari Listenmaa, Chalmers University of Technology
Kaili Müürisep, institute of computer science, University of Tartu
Marja-Liisa Olthuis
Flammie A Pirinen, Norgga árktaš universitehta
Daniel Glen Swanson, Indiana University
Trond Trosterud, University of Tromsø
Francis M. Tyers, Indiana University, Bloomington
Kevin Brubeck Unhammer, Trigram
Linda Wiechetek, University of Tromsø

Table of Contents

<i>An Annotated Error Corpus for Esperanto</i> Eckhard Bick	1
<i>Rule-based Surface Realization of Romanian Weak Pronouns</i> Ciprian Gerstenberger	9
<i>Drawing Blue Lines - What can Constraint Grammar do for GEC?</i> Linda Wiechetek and Kevin Brubeck Unhammer	19
<i>Towards Natural Language Explanations of Constraint Grammar Rules</i> Daniel Glen Swanson	28
<i>A Mansi FST and spellchecker</i> Jack Rueter, Csilla Horváth and Trond Trosterud	32
<i>A grammatical analyser for Tokelau</i> Trond Trosterud and Arnfinn Muruvik Vonen	38
<i>A Grammar-Based Method for Instilling Empirical Dependency Structure in LLMs</i> Olle Torstensson and Oskar Holmström	45
<i>Case error corrections for noun phrases containing deverbal attributive nouns in Greenlandic</i> Judithe Denbæk	50
<i>Divvunspell—Finite-State Spell-Checking and Correction on Modern Platforms</i> Flammie A Pirinen and Sjur Nørstebø Moshagen	59

Program

Wednesday, March 5, 2025

08:30 - 08:45 *Opening Remarks*

09:30 - 11:30 *Oral Session 1*

11:00 - 12:00 *Break*

12:00 - 14:00 *Oral Session 2*

14:00 - 15:00 *Break*

15:00 - 16:00 *Oral Session 3*

Wednesday, March 5, 2025 (continued)

16:00 - 17:00 *Panel Discussions and Business Meetings*