IWCLUL 2018

The 4th International Workshop on Computational Linguistics for Uralic Languages

by ACL SIG for Uralic Languages

Proceedings of the Workshop

January 8–9, 2018
Helsinki, Finland

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

ISBN XXX-X-XXXXXX-XX-X

Preface

The 4th International Workshop on Computational Linguistics for the Uralic Languages (IWCLUL) continues the annual meetings ACL SIGUR (Association of computational linguistics' special interest group for Uralic languages) after St. Petersburg (2017), Szeged (2016), and Tromsø (2015). It took place in Helsinki from 8th to 9th January, 2018 and was organized in collaboration with the NLP Research Group at the University of Helsinki.

This year we received a total of 20 submissions of which we accepted 15 (one of which was withdrawn by the authors) giving total of 14 high-quality papers in the final proceedings and an acceptance rate of 75 %. The accepted papers represent a variety of languages and growing resources in the Uralic landscape: Finnish, Komi-Zyrian, Udmurt, Erzya, Northern Sámi, Pite Sámi, Nganasan and Estonian; topics covered treebanks, parsing, code-switching, language generation, automatic speech recognition, morphology, and typological treatment across all Uralic languages, among others.

During this year's annual meeting we also had the first election of the ACL SIGUR board after the establishment of the new SIG in Szeged in 2016. The current board was re-elected by the ACL SIGUR membership for two further years.

We thank the programming committee, local organisers and participants for making annual meetings of ACL SIG for Uralic languages possible.

— Helsinki, 10th of January 2018, The organisers

Organizers:

Tommi A. Pirinen, Universität Hamburg
Michael Rießler, Universität Bielefeld
Jack Rueter, Helsingin yliopisto
Trond Trosterud, UiT Norgga árktalaš universitehta
Francis M. Tyers, Higher School of Economics

Program Committee:

Timofey Arkhangelskiy, national research university, Higher school of Economics (Russia) / Alexander von Humboldt Foundation (Germany)
Csilla Horvath, Research Institute for Linguistics, Hungarian Academy of Sciences (Hungary)
Mans Hulden, University of Colorado Boulder (USA)
Heiki-Jaan Kaalep, University of Tartu (Estonia)
Tommi A. Pirinen, Universität Hamburg (Germany)
Michael Rießler, Universität Bielefeld (Germany)
Miikka Silfverberg, University of Colorado Boulder (USA)
Eszter Simon, Magyar tudományos akadémia (Hungary)
Trond Trosterud, UiT Norgga árktalaš universitehta (Norway)
Francis M. Tyers, Higher School of Economics (Russia)
Veronika Vincze, Szegedi tudományegyetem (Hungary)

Invited Speaker:

Veronika Laippala, University of Turku

Table of Contents

Conference Program

Monday, January 8, 2018

9:00–9:30 ***Registration***

9:30–10:30 *Invited Talk by Veronika Laippala "How we built the Finnish dependency treebank and all the many things we did with it"*

10:30–12:00 Poster presentations

10:30–10:35 *Dependency Parsing of Code-Switching Data with Cross-Lingual Feature Representations*
Niko Partanen, Kyungtae Lim, Michael Rießler and Thierry Poibeau

10:35–10:40 *Building a Finnish SOM-based ontology concept tagger and harvester*
Seppo Nyrkkö

10:40–10:45 *Sound-aligned corpus of Udmurt dialectal texts*
Timofey Arkhangelskiy and Ekaterina Georgieva

10:45–10:50 *Automatic Generation of Wiktionary Entries for Finno-Ugric Minority Languages*
Zsanett Ferenczi, Iván Mittelholcz and Eszter Simon

10:50–10:55 *Development of an Open Source Natural Language Generation Tool for Finnish*
Mika Hämäläinen and Jack Rueter

10:55–11:00 *Guessing lexicon entries using finite-state methods*
Kimmo Koskenniemi

11:00–12:00 ***Posters and coffee***

Monday, January 8, 2018 (continued)

Presentations 1

12:00–12:30 *Tracking Typological Traits of Uralic Languages in Distributed Language Representations*
Johannes Bjerva and Isabelle Augenstein

12:30–13:00 *New Baseline in Automatic Speech Recognition for Northern Sámi*
Juho Leinonen, Peter Smit, Sami Virpioja and Mikko Kurimo

13:00–13:30 *Initial Experiments in Data-Driven Morphological Analysis for Finnish*
Miikka Silfverberg and Mans Hulden

13:30–14:00 *Towards an open-source universal-dependency treebank for Erzya*
Jack Rueter and Francis Tyers

14:00–15:00 ***Lunch***

Presentations 2

15:00–15:30 *Utilization of Nganasan digital resources: a statistical approach to vowel harmony*
László Fejes

15:30–16:00 *Parallel Forms in Estonian Finite State Morphology*
Heiki-Jaan Kaalep

16:00–16:30 *Extracting inflectional class assignment in Pite Saami: Nouns, verbs and those pesky adjectives*
Joshua Wilbur

16:30–17:00 *Analysing Finnish with word lists: the DDI approach to morphology revisited*
Atro Voutilainen and Maria Palolahti

17:00–17:30 ***Posters and coffee***

17:30–18:30 ***SIGUR general meeting***

Tuesday, January 9, 2018

10:00–16:00 Tutorials and discussions