IWCLUL 2017

3rd International Workshop for Computational Linguistics of Uralic Languages

Proceedings of the Workshop

23–24 January 2017
Norwegian University Centre of Oslo, St. Petersburg
St. Petersburg, Russia

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

Introduction

Uralic is an interesting group of languages from the computational-linguistic perspective. The Uralic languages share large parts of morphological and morphophonological complexity that is not present in the Indo-European language family, which has traditionally dominated computational-linguistic research. This can be seen for example in number of morphologically complex forms belonging to one word, which in Indo-European languages is in range of ones or tens whereas for Uralic languages, it is in the range of hundreds and thousands. Furthermore, Uralic language situations share a lot of geo-political aspects: the three national languages—Finnish, Estonian and Hungarian—are comparably small languages and only moderately resourced in terms of computational-linguistics while being stable and not in threat of extinction. The recognised minority languages of western-European states, on the other hand—such as North Smi, Kven and Vro—do clearly fall in the category of lesser resourced and more threatened languages, whereas the majority of Uralic languages in the east of Europe and Siberia are close to extinction. Common to all rapid development of more advanced computational-linguistic methods is required for continued vitality of the languages in everyday life, to enable archiving and use of the languages with computers and other devices such as mobile applications.

Computational linguistic Research inside Uralistics is being carried out only in a handful of universities, research institutes and other sites and only by relatively few researchers. Our intention with organising this conference is to gather these researchers from scattered institutions together in order to share ideas and resources, and avoid duplicating efforts in gathering and enriching these scarce resources. We want to initiate more concentrated effort in collecting and improving language resources and technologies for the survival of the Uralic languages and hope that our effort today will become an ongoing tradition in the future.

For the current proceedings of The Third International Workshop on Computational Linguistics for Uralic Languages, we accepted 10 high-quality submissions about topics including computational lexicography, language documentation, optical character recognition, dependency parsing, web-as-corpus as well as automatic and rule-based morphological analysis methods. The covered languages are very broad and reach from different Smi languages, over Kven, Finnish, Komi, Udmurt, Mari, Khanty, Mansi, and Tundra Nenets. Whereas some papers describe language-specific research, others compare different languages or work on small Uralic languages in general. These contributions are all very important for the preservation and development of Uralic languages as well as for future linguistic investigations on them.

The conference was organized in collaboration with The University of Oslo St. Petersburg Representative Office and held in St. Petersburg, Russia, on January 23rd and 24th 2017. The program consisted of an invited speech by Heiki-Jaan Kaalep, a poster session, and four talks during the first day and an open discussion and individual project workshops during the second day. The current proceedings include the written versions all oral and poster presentations.

—Tommi A Pirinen, Trond Trosterud, Francis M. Tyers, Michael Rieler
Conference organisers,
January 22, 2017, St. Petersburg

Organizers:

Francis M. Tyers, UiT Norgga árktalaš universitehta
Michael Rießler, Albert-Ludwigs-Universität Freiburg
Tommi A. Pirinen, Universität Hamburg
Trond Trosterud, UiT Norgga árktalaš universitehta

Program Committee:

Eszter Simon, Magyar tudományos akadémia (Hungary)
Francis M. Tyers, UiT Norgga árktalaš universitehta (Norway)
Jack Rueter, Helsingin yliopisto (Finland)
Mans Hulden, University of Colorado at Boulder (USA)
Michael Rieler, Albert-Ludwigs-Universität Freiburg (Germany)
Miikka Silfverberg, University of Helsinki (Finland)
Tommi A. Pirinen, Universität Hanmburg (Germany)
Trond Trosterud, UiT Norgga árktalaš universitehta (Norway)
Veronika Vincze, Szegedi tudományegyetem (Hungary)
a , - Ł (Russia)
, ” ” (Russia)

Invited Speaker:

Heiki-Jaan Kaalep, Tartu ülikool

Table of Contents

Conference Program

Monday, January 23, 2017

10:00–10:15 Opening Remarks

10:15–11:30 Invited Talk by Heiki-Jaan Kaalep

Session 1: Poster boasters

11:30–11:33 *Synchronized Mediawiki based analyzer dictionary development*
Jack Rueter and Mika Hämäläinen

11:33–11:36 *DEMO: Giellatekno Open-source click-in-text dictionaries for bringing closely related languages into contact.*
Jack Rueter

11:36–11:39 *Languages under the influence: Building a database of Uralic languages*
Eszter Simon and Nikolett Mus

11:39–11:42 *Instant Annotations –Applying NLP Methods to the Annotation of Spoken Language Documentation Corpora*
Ciprian Gerstenberger, Niko Partanen, Michael Rießler and Joshua Wilbur

11:45–12:30 Posters and Coffee

12:30–14:00 Lunch

Session 2: Oral presentations

14:00–14:30 *Preliminary Experiments concerning Verbal Predicative Structure Extraction from a Large Finnish Corpus*
Guersande Chaminade and Thierry Poibeau

14:30–15:00 *Language technology resources and tools for Mansi: an overview*
Csilla Horváth, Norbert Szilágyi, Veronika Vincze and Ágoston Nagy

15:00–15:30 *Annotation schemes in North Sámi dependency parsing*
Francis M. Tyers and Mariya Sheyanova

15:30–16:00 *A morphological analyser for Kven*
Sindre Reino Trosterud, Trond Trosterud, Anna-Kaisa Räisänen, Leena Niiranen, Mervi Haavisto and Kaisa Maliniemi

Monday, January 23, 2017 (continued)

16:00–16:30	LRE Map, LR Matrices and LR Impact Factor
16:30–17:15	Posters, demos and coffee
17:15–18:30	ACL SIGUR business meeting
19:00–late	Conference Dinner

Tuesday, January 24, 2017

10:00–18:00	Round tables, discussions and workshopping