

5th International Workshop for Computational Linguistics of Uralic Languages (IWCLUL 2019)

Tartu, Estonia
7 - 8 January 2019

ISBN: 978-1-5108-8132-7

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2019) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2019)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Data-Driven Morphological Analysis for Uralic Languages</i> Miikka Silfverberg and Francis Tyers	1
<i>North Sámi morphological segmentation with low-resource semi-supervised sequence labeling</i> Stig-Arne Grönroos, Sami Virpioja and Mikko Kurimo	15
<i>What does the Nom say? An algorithm for case disambiguation in Hungarian</i> Noémi Ligeti-Nagy, Andrea Dömötör and Noémi Vadász	27
<i>A Contrastive Evaluation of Word Sense Disambiguation Systems for Finnish</i> Frankie Robertson	42
<i>Elliptical Constructions in Estonian UD Treebank</i> Kadri Muischnek and Liisi Torga	55
<i>FiST –towards a free Semantic Tagger of modern standard Finnish</i> Kimmo Kettunen	66
<i>An OCR system for the Unified Northern Alphabet</i> Niko Partanen and Michael Rießler	77
<i>ELAN as a search engine for hierarchically structured, tagged corpora</i> Joshua Wilbur	90
<i>Neural and rule-based Finnish NLP models—expectations, experiments and experiences</i> Tommi A Pirinen	104
<i>Uralic multimedia corpora: ISO/TEI corpus data in the project INEL</i> Timofey Arkhangelskiy, Anne Ferger and Hanna Hedeland	115
<i>Corpora of social media in minority Uralic languages</i> Timofey Arkhangelskiy	125
<i>Is this the end? Two-step tokenization of sentence boundaries</i> Linda Wiecheteck, Sjur Nørstebø Moshagen and Thomas Omma	141
<i>Learning multilingual topics through aspect extraction from monolingual texts</i> Johannes Huber and Myra Spiliopoulou	154
<i>Electronical resources for Livonian</i> Valts Ernštreits	184
<i>The use of Extract Morphology for Automatic Generation of Language Technology for Votic</i> Kristian Kankainen	19H