

5th Workshop on Recent Advances in Slavonic Natural Language Processing

(RASLAN 2011)

**Karlova Studanka, Czech Republic
2 – 4 December 2011**

Editors:

**A. Horak
P. Rychly**

ISBN: 978-1-62993-831-8

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© (2011) by the Masaryk University Faculty of Informatics
All rights reserved.

Printed by Curran Associates, Inc. (2014)

For permission requests, please contact the Masaryk University Faculty of Informatics
at the address below.

Masaryk University Faculty of Informatics
c/o Dana Hlaváčková
Centre for Natural Language Processing
B206 Botanická 554/68a, Ponava, Brno
Czech Republic

Phone: +420 549 49 1864, 3329

ydana@aurora.fi.muni.cz

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

I Logic and Language

- Analyzing Time-Related Clauses in Transparent Intensional Logic 3
Aleš Horák, Miloš Jakubiček, Vojtěch Kovář
- Mind modeling using Transparent Intensional Logic 11
Andrej Gardoň
- Verbs as Predicates: Towards Inference in a Discourse 19
Zuzana Nevěřilová, Marek Grác

II Syntax, Morphology and Lexicon

- Czech Morphological Tagset Revisited 29
Miloš Jakubiček, Vojtěch Kovář, Pavel Šmerk
- Korean Parsing in an Extended Categorical Grammar 43
Juyeon Kang
- Extracting Phrases from PDT 2.0 51
Vašek Němčík
- Extended VerbaLex Editor Interface Based on the DEB Platform 59
Dana Hlaváčková, Adam Rambousek

III Text Corpora

- Building Corpora of Technical Texts: Approaches and Tools 69
Petr Sojka, Martin Liška, Michal Růžička
- Corpus-based Disambiguation for Machine Translation 81
Vít Baisa
- Building a 50M Corpus of Tajik Language 89
Gulshan Dovudov, Jan Pomikálek, Vít Suchomel, Pavel Šmerk
- Practical Web Crawling for Text Corpora 97
Vít Suchomel, Jan Pomikálek

IV Language Modelling

A Bayesian Approach to Query Language Identification	111
<i>Jiří Materna and Juraj Hreško</i>	
A Framework for Authorship Identification in the Internet Environment .	117
<i>Jan Rygl, Aleš Horák</i>	
chared: Character Encoding Detection with a Known Language	125
<i>Jan Pomikálek, Vít Suchomel</i>	
Words' Burstiness in Language Models	131
<i>Pavel Rychlý</i>	
Author Index	139