Association for Computational Linguistics

EACL 2003

10th Conference of The European Chapter

Proceedings of the Workshop on Morphological Processing of Slavic Languages

April 13th 2003

Agro Hotel, Budapest, Hungary

Association for Computational Linguistics

EACL 2003

10th Conference of The European Chapter

Proceedings of the Workshop on Morphological Processing of Slavic Languages

April 13th 2003

Agro Hotel, Budapest, Hungary

The conference, the workshops and the tutorials are sponsored by:

Chief Patron of the Conference:
Dr. Ferenc Baja
Political State Secretary
Office of Government Information Technology and Civil Relations
Prime Minister's Office

Linguistic Systems BV
Leo Konst (Managing director)
Postbus 1186, 6501 BD Nijmegen, Nederland
tel: +31 24 322 63 02
fax: +31 24 324 21 16
e-mail: info@euroglot.nl, leokonst@telebyte.nl,
http://www.euroglot.nl

Xerox Research Centre Europe
Irene Maxwell
6 chemin de Maupertuis
38240 Meylan, France
Tel: +33 (0)4.76.61.50.83
Fax: +33 (0)4.76.61.50.99
email: info@xrce.xerox.com
website: www.xrce.xerox.com

ATALA
Jean Veronis
Jean.Veronis@up.univ-mrs.fr
45 rue d'Ulm
75230 Paris Cedex 5, France
http://www.atala.org

ELRA/ELDA
Khalid Choukri
choukri@elda.fr
55-57 rue Brillat Savarin
75013 Paris, France
Tel: (+33 1) 43 13 33 33,
Fax: (+33 1) 43 13 33 30
http://www.elda.fr

Order copies of ACL proceedings from: Priscilla Rasmussen, Association for Computational Linguistics, 3 Landmark Center, East Stroudsburg, PA 18301 USA, Phone +1-570-476-8006, Fax +1-570-476-0860, URL http://www.acl-web.org.

INTRODUCTION

This volume contains the papers accepted for the EACL-03 workshop on Morphological Processing of Slavic Languages held in Budapest on April 13, 2003, just preceding the 11th Conference of the European Chapter of the Association for Computational Linguistics.

The aim of this workshop was to present, in one place, different aspects of the morphological processing of Slavic languages and to establish the current relation between linguistic knowledge and the possibilities of fulfilling the computational needs for Slavic languages. Different approaches to modelling morphological structure, to lexical and corpus annotation and to processing morphological information have been developed, and some of them for more than one language. Yet annotation schemes, morphological analysers, part-of-speech taggers or language resources that encompass all - or even a larger number of - Slavic languages are rare. At the same time, a systematic review of existing approaches to the morphological processing of Slavic languages and their relations does not yet exist.

The topic of the workshop was the morphological computational analysis and annotation of Slavic languages, encountered on both the inflective and the derivational levels. The workshop discussed the lexical structures necessary for morphological analysis and presented standardisation efforts in the field that can, for instance, enable the transfer of applied methods from one language to another, or aid in the annotation of morphological information in corpora. In addition to resources, these papers also discuss methods for word-level syntactic tagging, lexicon acquisition, collocation and term extraction.

From the fifteen papers submitted for this workshop, the reviewers selected the twelve papers included in these proceedings. The papers cover seven Slavic languages - Bulgarian, Czech, Croatian, Polish, Russian, Serbian, and Slovenian - as well as the majority of suggested workshop topics. The organisers and editors would like to thank the reviewers as well as the authors for their work in making the Workshop on Morphological Processing of Slavic Languages a success.

Tomaž Erjavec & Duško Vitas

April 2003

ORGANISERS:

Tomaž Erjavec, Jožef Stefan Institute, Ljubljana
Duško Vitas, University of Belgrade, Belgrade

PROGRAMME COMMITTEE:

František Čermak, Charles University, Czech Republic
Greville G. Corbett, University of Surrey, UK
Tomaž Erjavec, Jožef Stefan Institute, Slovenia
Roger Evans, University of Brighton, UK
Karel Pala, Masaryk University, Czech Republic
Vladimír Petkevíč, Charles University, Czech Republic
Vladimir Plungian, Russian Academy of Sciences, Russia
Max Silberztein, Université de Franche-Comté, France
Kiril Simov, Bulgarian Academy of Sciences, Bulgaria
Marko Tadić, University of Zagreb, Croatia
Duško Vitas, University of Belgrade, Serbia & Montenegro

FURTHER INFORMATION:

Tomaž Erjavec
Jožef Stefan Institute
Jamova 39
SI-1000 Ljubljana, Slovenia

WORKSHOP WEBSITE:

http://nl.ijs.si/mpsl03/

WORKSHOP PROGRAMME

Sunday, April 13

8:45-9:00	Welcome
09:00-09:30	*Relations between Inflectional and Derivation Patterns* Karel Pala, Radek Sedláček and Marek Veber
09:30-10:00	*A Large-scale Inheritance-based Morphological Lexicon for Russian* Roger Evans, Carole Tiberius, Dunstan Brown and Greville C. Corbett
10:00-10:30	*Automatic Lexical Acquisition from Raw Corpora: An Application to Russian* Antoni Oliver, Irene Castellón and Lluís Màrquez
10:30-11:00	Morning Break
11:00-11:30	*The MULTEXT-East Morphosyntactic Specification for Slavic Languages* Tomaž Erjavec, Cvetana Krstev, Vladimir Petkevíč, Kiril Simov, Marko Tadić and Duško Vitas
11:30-12:00	*A Flexemic Tagset for Polish* Adam Przepiórkowski and Marcin Woliński
12:00-12:30	*Building the Croatian Morphological Lexicon* Marko Tadić and Sanja Fulgosi
12:30-14:00	Lunch
14:00-14:30	*Unsupervised Learning of Bulgarian POS Tags* Derrick Higgins
14:30-15:00	*Composite Tense Recognition and Tagging in Serbian* Duško Vitas and Cvetana Krstev
15:00-15:30	*A Reconfigurable Stochastic Tagger for Languages with Complex Tag Structure* Łukasz Dębowski
15:30-16:00	Afternoon Break
16:00-16:30	*Some Aspects of the Morphological Processing of Bulgarian* Milena Slavcheva
16:30-17:00	*Morpho-syntactic Clues for Terminological Processing in Serbian* Goran Nenadić, Irena Spasić and Sophia Ananiadou
17:00-17:30	*Russian Morphology: Ressources and Java Software Application* Serge Yablonsky
17:30-18:30	Round Table *Slavic Languages: Between Linguistic Description and Computational Needs*

Table of Contents