Association for Computational Linguistics

EACL 2003

10th Conference of The European Chapter

Proceedings of 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)

April 13th - 14th 2003

Agro Hotel, Budapest, Hungary

Association for Computational Linguistics

EACL 2003

10th Conference of The European Chapter

Proceedings of 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)

April 13th - 14th 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.

Preface

These proceedings document the fourth event in the LINC workshop series. The first Workshop in this series took place in 1999 in connection with the Meeting of the European Chapter of the Association of Computational Linguistics in Bergen (Uszkoreit, Brants & Krenn, 1999). The second LINC Workshop was held in August of 2000 in Luxembourg in conjunction with COLING (Abeillé, Brants & Uszkoreit, 2000). In 2001 the third workshop was organized by Eva Hajicova in conjunction with the 34th Meeting of the Societas Linguistica Europaea in Leuven.

Whereas the first workshops of the series concentrated more on the basic annotation formats and methodologies for POS-tagging and treebanks, the focus has shifted gradually to reports on available large-scale interpreted corpora, schemes for multi-level annotation and first attempts of semantic annotation. Both the workshop and the EACL conference also document successful cases of corpus exploitation in NLP research. Although no LINC event took place in 2002, the overwhelming response to our call for papers for the 2003 workshop demonstrates the grown interest in the topic and the continued demand for this annual forum of scientific exchange.

Large linguistically interpreted corpora play an increasingly important role for machine learning, evaluation, psycholinguistics as well as theoretical linguistics. Many groups have started to create corpus resources annotated with morphological, syntactic, semantic and discourse information for a variety of languages. Linguistic annotation may consist of morphological analyses, trees, dependencies, grammatical relations, word senses, (co)references, information structure, semantic representations, discourse relations and other types of linguistic information.

This workshop aims at bringing together these activities in order to facilitate advanced and efficient corpus annotations which will provide re-usable resources. The workshop will also provide a forum for reports on the scientific and technological exploitation of interpreted corpora in general, computational or psycholinguistics. Such reports on exploitation results are valuable for the comparison of alternative approaches and will thus serve as feedback to ongoing and new corpus annotation efforts.

We received 34 submissions, 17 of which were accepted for presentation at the two-days workshop. The papers in this volume report on research in the following fields:

- creation of practical annotation schemes
- efficient annotation techniques including automation
- tools supporting corpus conversions
- consistency checking and validation
- tools and methods for searching and browsing
- qualitative and quantitative studies based on linguistically interpreted corpora
- technological advances achieved by the exploitation of interpreted corpora

The papers thus provide the basis for a roadmap session in which current and future directions in corpus annotation and exploitation are discussed.

We hope that you will enjoy your time in Budapest and find this workshop enjoyable and useful for your work.

Anne Abeillé,	Laboratoire de Linguistique Formelle, University Paris 7
Silvia Hansen-Schirra,	Computational Linguistics, Saarland University
Hans Uszkoreit,	Computational Linguistics, Saarland University & German Research Center for AI, Saarbrucken

References

Uszkoreit, H., T. Brants & B. Krenn (1999) (Eds.) Proceedings of the Workshop on Linguistically Interpreted Corpora (LINC-99) at the 9th Conference of the European Chapter of the Association of Computational Linguistics (EACL-99), Bergen.

Abeillé, A., T. Brants & H. Uszkoreit (Eds.) (2000) Proceedings of the Workshop on Linguistically Interpreted Corpora (LINC-2000) at the 18th International Conference on Computational Linguistics (Coling-2000), Luxembourg.

WORKSHOP SPONSORS

The European Chapter of the Association for Computational Linguistics
The European Network of Excellence in Human Language Technologies
Laboratoire de Linguistique Formelle, University Paris 7
Computational Linguistics, Saarland University, Saarbrücken
German Research Center for AI, Saarbrücken

ORGANIZERS

Ann Copestake (University of Cambridge, England)
Jan Hajič (Univerzita Karlova, Prague, Czech Republic)

PROGRAM COMMITTEE

Thorsten Brants (Mountain View),
John Carroll (Sussex),
Tomaz Erjavec (Ljubljana),
Frank Keller (Edinburgh),
Stephan Oepen (Stanford & Trondheim),
Laurent Romary (Nancy),
Geoffrey Sampson (Sussex),
Kiril Simov (Sofia),
Jean Veronis (Aix-en-Provence),
Atro Voutilainen (Helsinki),
Jakub Zavrel (Amsterdam),

ADDITIONAL REVIEWERS

Ulrich Callmeier (Saarbrücken),
Gregor Erbach (Saarbrücken),
Frederik Fouvry (Saarbrücken),
Valia Kordoni (Saarbrücken),
Geert-Jan Kruijff (Saarbrücken),
Stella Neumann (Saarbrücken),

WORKSHOP WEB SITE

http://www.coli.uni-sb.de/linc03

FURTHER INFORMATION

Anne Abeillé
Laboratoire de Linguistic Formelle
University Paris 7
2, place Jussieu
75251 Paris, France

abeille@linguist.jussieu.fr

Silvia Hansen-Schirra, Hans Uszkoreit
Computational Linguistics
Saarland University
P.O.Box 151150
66041 Saarbrücken, Germany

hansen,uszkoreit @coli.uni-sb.de

Table of Contents