Preface

Linguistically interpreted corpora cover a wide variety of topics in Computational Linguistics. They are used for developing processing methods and for linguistic investigations, at the word level and the structural level, for speech and language applications, for syntactic and semantic information, for translation and information retrieval, and for many other topics. More and more types of linguistic information are annotated in corpora.

What was started mainly for English is now performed for many other languages. This allows the transfer of methods previously used for English to other languages as well as the development of new techniques covering needs of other languages. Having corpora in many languages allows to develop and identify techniques applicable to a wide variety of languages. And it allows to adapt to a particular language if necessary.

This workshops aims at bringing together work on linguistic annotation schemes, annotation tools, algorithms aiming at automation of annotation, procedures for efficiently combining human and automatic processes, algorithms aiming at detecting errors and inconsistencies in corpora, search in copora, and the use of corpora for linguistic investigations.

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

Anne Abeille, Department of Linguistics, University Paris 7
Thorsten Brants, Computational Linguistics, Saarland University
Hans Uszkoreit, Computational Linguistics, Saarland University
and German Research Center for AI, Saarbrücken

Workshop on Linguistically Interpreted Corpora

LINC-2000

Luxembourg, August 6, 2000

Organizers:

Anne Abeille (Paris)
Thorsten Brants (Saarbrücken)
Hans Uszkoreit (Saarbrücken)

Programme Committee:

John Carroll (Brighton)
Lionel Clement (Paris)
Tomaz Erjavec (Ljubljana)
Frank Keller (Edinburgh)
Laurent Romary (Nancy)
Geoffrey Sampson (Brighton)
Jean Veronis (Aix-en-Provence)
Atro Voutilainen (Helsinki)
Jakub Zavrel (Antwerp)

Further Information:

Anne Abeille
UFR de Linguistique
Université Paris 7
2, place Jussieu
F-75251 Paris, France
abeille@linguist.jussieu.fr

Thorsten Brants, Hans Uszkoreit
FR 8.7 Computational Linguistics
Saarland University
P.O.Box 171150
D-66041 Saarbrücken, Germany
{brants,uszkoreit}@coli.uni-sb.de

http://www.coli.uni-sb.de/linc2000

Table of Contents

Comparing Linguistic Interpretation Schemes for English Corpora
Eric Atwell, George Demetriou, John Hughes, Amanda Schiffrin, Clive Souter, Sean Wilcock 1

Dependency-based Syntactic Annotation of a Chinese Corpus
Tom B. Y. Lai, Huang Changning 11

The Italian Syntactic-Semantic Treebank: Architecture, Annotation, Tools and Evaluation
S. Montemagni, F. Barsotti, M. Battista, N. Calzolari, O. Corazzari, A. Zampolli, F. Fanciulli, M. Massetani, R. Raffaelli, R. Basili, M. T. Pazienza, D. Saracino, F. Zanzotto, N. Mana, F. Pianesi, R. Delmonte 18

Where Should Annotation Stop?
Geoffrey Sampson 28

A Statistical Account on Word Order Variation in German
Daniela Kurz 35

Bottom-Up Tagset Design from Maximally Reduced Tagset
Péter Dienes, Csaba Oravecz 42

The Detection of Inconsistency in Manually Tagged Text
Hans van Halteren 48

Grammar-Based Corpus Annotation
Stefanie Dipper 56

Automatic Procedures in Tectogrammatical Tagging
Alena Böhmová, Petr Sgall 65

Considering Automatic Aids to Corpus Annotation
David Day, Benjamin Wellner 71

Author Index

Eric Atwell 1
F. Barsotti 18
R. Basili 18
M. Battista 18
Alena Böhmová 65
N. Calzolari 18
Huang Changning 11
O. Corazzari 18
David Day 71
R. Delmonte 18
George Demetriou 1
Péter Dienes 42
Stefanie Dipper 56
F. Fanciulli 18
John Hughes 1
Daniela Kurz 35
Tom Lai 11
N. Mana 18
M. Massetani 18
S. Montemagni 18
Csaba Oravecz 42
M. Pazienza 18
F. Pianesi 18
R. Raffaelli 18
Geoffrey Sampson 28
D. Saracino 18
Amanda Schiffrin 1
Petr Sgall 65
Clive Souter 1
Hans van Halteren 48
Benjamin Wellner 71
Sean Wilcock 1
A. Zampolli 18
F. Zanzotto 18