

**Proceedings of the
2004 ACL Workshop on Discourse Annotation**

Held in cooperation with ACL-2004

**25–26 July 2004
Barcelona, Spain**

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
73 Landmark Center
East Stroudsburg, PA 18301
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

PREFACE

This volume contains the papers accepted for presentation at the 2004 Workshop on Discourse Annotation, held in Barcelona, Spain, July 25 and 26, in conjunction with the International and the European Conferences of the ACL. Out of 24 papers submitted for review, 16 were selected for inclusion in the workshop proceedings.

The workshop is being held both in recognition of renewed interest in gold standard corpora for the study and exploration of discourse phenomena and in recognition of growing interest in resources that integrate various types of annotation in order to better characterize such phenomena and thereby help to develop more sensitive and robust tools for dealing with discourse. The workshop is neutral as to whether consensus annotation is possible for every type of discourse phenomenon. Its aims are rather to:

- bring a fuller range of discourse annotation activity to the attention of researchers working on discourse phenomena and their usefulness for language technologies;
- highlight tools used in the annotation process or used to display, query or further analyse the results of annotation;
- discuss obstacles to discourse-level annotation, such as the greater subjectivity in bracketting and labelling judgments, the ambiguity of discourse phenomena, and data sparseness;
- discuss where and how automated and semi-automated annotation can effectively augment or complement manual gold standard annotation;
- discuss opportunities for creating a significantly large, reusable corpus (or set of corpora) annotated for multiple discourse and sentence-level phenomena, as a much richer basis for both assessing theories and building better tools.

The papers selected for the workshop demonstrate a range of discourse phenomena being annotated today, tools being used to create and query annotation, and conclusions being drawn from both manually and automatically annotated corpora. At the workshop itself, blocks of time have been allocated for group discussion of topics related to the above aims.

We are very grateful to our programme committee for their significant efforts in not only reviewing the papers but also shaping the workshop. Finally, we are grateful to the ACL/EACL-2004 conference organizers for enabling this workshop to happen.

Bonnie Webber & Donna Byron (editors)

June 2004

PROGRAM COMMITTEE:

Bonnie Webber, University of Edinburgh (co-chair)
Donna Byron, Ohio State University (co-chair)

Steven Bird, Melbourne University
Liesbeth Degand, University of Louvain
Eva Hajicova, Charles University
Aravind Joshi, University of Pennsylvania
Andrew Kehler, UC San Diego
Daniel Marcu, ISI
Katja Markert, Leeds University
Malvina Nissim, Edinburgh University
Livia Polanyi, FXPAL
Frank Schilder, Thomson Legal and Regulatory
Andrea Setzer, Sheffield University
Wilbert Spooren, Free University of Amsterdam
Manfred Stede, University of Potsdam
Michael Strube, EML Research, Heidelberg
Joel Tetreault, University of Rochester
Martin van den Berg, FXPAL
Annie Zaenen, PARC

WORKSHOP WEBSITE:

<http://www.clit.osu.edu/dbyron/acl04>

Table of Contents

<i>A Framework for Feature based Description of Low level Discourse</i> Laura Alonso Alemany, Ezequiel Andújar Hinojosa and Robert Sola Salvatierra	1
<i>COOPML: Towards Annotating Cooperative Discourse</i> Farah Benamara, Véronique Moriceau and Patrick Saint-Dizier	9
<i>Using a Probabilistic Model of Discourse Relations to Investigate Word Order Variation</i> Cassandra Creswell	17
<i>On the Use of Automatic Tools for Large-scale Semantic Analyses of Causal Connectives</i> Liesbeth Degand, Wilbert Spooren and Yves Bestgen	25
<i>Korean Null Pronouns: Classification and Annotation</i> Na-Rae Han	33
<i>Discourse-level Annotation for Investigating Information Structure</i> Ivana Kruijff-Korbayová and Geert-Jan M. Kruijff	41
<i>Text Type Structure and Logical Document Structure</i> Hagen Langer, Harald Lungen and Petra Saskia Bayerl	49
<i>Temporal Discourse Models for Narrative Structure</i> Inderjeet Mani and James Pustejovsky	57
<i>Exploiting Semantic Information for Manual Anaphoric Annotation in Cast3LB Corpus</i> Borja Navarro, Rubén Izquierdo and Maximiliano Saiz-Noeda	65
<i>Discourse Annotation and Semantic Annotation in the GNOME corpus</i> Massimo Poesio	72
<i>Sentential Structure and Discourse Parsing</i> Livia Polanyi, Chris Culy, Martin van den Berg, Gian Lorenzo Thione and David Ahn	80
<i>Annotation and Data Mining of the Penn Discourse TreeBank</i> Rashmi Prasad, Eleni Miltsakaki, Aravind Joshi and Bonnie Webber	88
<i>The Potsdam Commentary Corpus</i> Manfred Stede	96
<i>Discourse Annotation in the Monroe Corpus</i> Joel Tetreault, Mary Swift, Preethum Prithviraj, Myroslava Dzikovska and James Allen	103
<i>LiveTree: An Integrated Workbench for Discourse Processing</i> Gian Lorenzo Thione, Martin van den Berg, Chris Culy and Livia Polanyi	110
<i>Animacy Encoding in English: Why and How</i> Annie Zaenen, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor and Tom Wasow	118

Preliminary Program Schedule

Sunday, July 25

- 8:45-9:00 Welcome
- 9:00-9:20 *Text Type Structure and Logical Document Structure*
Hagen Langer, Harald Lüngen, Petra Saskia Bayerl
- 9:20-9:40 *Discourse Annotation and Semantic Annotation in the GNOME Corpus*
Massimo Poesio
- 9:40-10:00 *Discourse Annotation in the Monroe Corpus*
Joel Tetreault, Mary Swift, Preethum Prithviraj, Myroslava Dzikovska, James Allen
- 10:00-10:30 Coffee Break
- 10:30-10:50 *Sentential Structure and Discourse Parsing*
Livia Polanyi, Chris Culy, Martin van den Berg, Gian Lorenzo Thione, David Ahn
- 10:50-11:10 *Annotation and Data Mining of the Penn Discourse TreeBank*
Rashmi Prasad, Eleni Miltsakaki, Aravind Joshi, Bonnie Webber
- 11:10-11:30 *The Potsdam Commentary Corpus*
Manfred Stede
- 11:30-12:00 **Moderated Discussion:** Integrating multiple levels of annotation;
Producing annotation for re-use
- 12:00-13:30 Lunch
- 13:30-14:10 Short Descriptions of Demo Software
- 14:10-15:20 Demos of corpus annotation tools
LiveTree: An Integrated Workbench for Discourse Processing
Gian Lorenzo Thione, Martin van den Berg, Chris Culy, Livia Polanyi
Practical Multi-Level Stand-Off Annotation: The MMAX2 Approach
Christoph Müller, Michael Strube
WordFreak: A Tool for Annotating Discourse Connectives and their Argument Structure
Jeremy LaCivita, Thomas Morton, Nikhil Dinesh, Rashmi Prasad, Eleni Miltsakaki,
Aravind Joshi, Bonnie Webber
The ConAno Annotation and Query Tool
Manfred Stede
- 15:20-15:50 Coffee Break
- 15:50-16:20 *Using a Probabilistic Model of Discourse Relations to Investigate Word Order Variation*
Cassandre Creswell
- 16:20-16:40 *On the Use of Automatic Tools for Large-scale Semantic Analyses of Causal Connectives*
Liesbeth Degand, Wilbert Spooren, Yves Bestgen
- 16:40-17:00 *Discourse-level Annotation for Investigating Information Structure*
Ivana Kruijff-Korbayová, Geert-Jan M. Kruijff
- 17:00-17:20 *Exploiting Semantic Information for Manual Anaphoric Annotation in Cast3LB Corpus*
Borja Navarro, Rubén Izquierdo, Maximiliano Saiz-Noeda
- 17:20-18:00 **Moderated Discussion:** Automatable approximations to Gold Standard annotation;
Multi-lingual discourse annotation

Monday, July 26

- 9:00-9:20 *A Framework for Feature-based Description of Low level Discourse*
Laura Alonso Alemany, Ezequiel Andújar Hinojosa, Robert Sola Salvatierra
- 9:20-9:40 *COOPML: Towards Annotating Cooperative Discourse*
Farah Benamara, Véronique Moriceau, Patrick Saint-Dizier
- 9:40-10:00 *Korean Null Pronouns: Classification and Annotation*
Na-Rae Han
- 10:00-10:30 Coffee Break
- 10:30-10:50 *Temporal Discourse Models for Narrative Structure*
Inderjeet Mani, James Pustejovsky
- 10:50-11:10 *Animacy Encoding in English: Why and How*
Annie Zaenen, Jean Carletta, Gregory Garretson, Joan Bresnan,
Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor,
Tom Wasow
- 11:10-11:45 **Moderated Discussion:** Subjectivity in manual discourse annotation;
Data sparsity problems in manual annotation
- 11:45-12:00 *Wrap-up*