Frontiers in Linguistically Annotated Corpora 2006

A Merged Workshop with 7th International Workshop on Linguistically Interpreted Corpora (LINC-2006) and Frontiers in Corpus Annotation III

Proceedings of the Workshop

Chairs:
Adam Meyers, Shigeko Nariyama, Timothy Baldwin and Francis Bond

22 July 2006
Sydney, Australia
# Table of Contents

Preface ................................................................. v  
Organizers .............................................................. vii  
Workshop Program ................................................... ix  

Challenges for Annotating Images for Sense Disambiguation  
   Cecilia Ovesdotter Alm, Nicolas Loeff and David A. Forsyth ....................... 1  

A Semi-Automatic Method for Annotating a Biomedical Proposition Bank  
   Wen-Chi Chou, Richard Tzong-Han Tsai, Ying-Shan Su, Wei Ku, Ting-Yi Sung and  
   Wen-Lian Hsu .......................................................... 5  

How and Where do People Fail with Time: Temporal Reference Mapping Annotation by Chinese  
   and English Bilinguals  
   Yang Ye and Steven Abney .............................................. 13  

Probing the Space of Grammatical Variation: Induction of Cross-Lingual Grammatical Constraints  
   from Treebanks  
   Felice Dell’Orletta, Alessandro Lenci, Simonetta Montemagni and Vito Pirrelli ........ 21  

Frontiers in Linguistic Annotation for Lower-Density Languages  
   Mike Maxwell and Baden Hughes ........................................ 29  

Annotation Compatibility Working Group Report  
   Adam Meyers ............................................................. 38  

Manual Annotation of Opinion Categories in Meetings  
   Swapna Somasundaran, Janyce Wiebe, Paul Hoffmann and Diane Litman ........... 54  

The Hinoki Sensebank — A Large-Scale Word Sense Tagged Corpus of Japanese —  
   Takaaki Tanaka, Francis Bond and Sanae Fujita ..................................... 62  

Issues in Synchronizing the English Treebank and PropBank  
   Olga Babko-Malaya, Ann Bies, Ann Taylor, Szuting Yi, Martha Palmer, Mitch Marcus,  
   Seth Kulick and Libin Shen .................................................. 70  

On Distance between Deep Syntax and Semantic Representation  
   Václav Novák .................................................................. 78  

Corpus Annotation by Generation  
   Elke Teich, John A. Bateman and Richard Eckart ........................................ 86  

Constructing an English Valency Lexicon  
   Jiří Semecký and Silvie Cinková ................................................. 94  

Author Index .................................................................. 99
Preface

Large linguistically interpreted corpora play an increasingly important role for machine learning, evaluation, psycholinguistics as well as theoretical linguistics. Many research groups are engaged in the creation of corpus resources annotated with morphological, syntactic, semantic, discourse and other linguistic information for a variety of languages. This workshop brings together researchers interested in the annotation of linguistically interpreted corpora by combining two workshops: Frontiers in Linguistic Annotation III and the 7th International Workshop on Linguistically Interpreted Corpora (LINC-2006). The goal of the workshop is to identify and disseminate best practice in the development and utilization of linguistically interpreted corpora.

We would like to thank all the authors who submitted papers. There were 19 submissions, of which 10 appear in the proceedings. We would like to particularly thank all the members of the program committee for their time and effort in ensuring that the papers were fairly assessed, and for the useful comments they provided.

In addition to regular papers, the workshop features presentations by working groups on two topics:

**Annotation Compatibility:** A roadmap of the compatibility of current annotation schemes with each other: *Annotation Compatibility Working Group Report.*

**Low-density Languages:** A discussion of low density languages and the problems associated with them: *Frontiers in Linguistic Annotation for Lower-Density Languages.*

The Innovative Student Annotation Award, for best paper by a student, went to Václav Novák, for his paper *On Distance between Deep Syntax and Semantic Representation.*
Organizers

Chairs:
Timothy Baldwin, University of Melbourne
Francis Bond, NTT Communication Science Laboratories
Adam Meyers, New York University
Shigeko Nariyama, University of Melbourne

Program Committee:
Lars Ahrenberg, Linköpings Universitet
Kathy Baker, U.S. Dept. of Defense
Steven Bird, University of Melbourne
Alex Chengyu Fang, City University Hong Kong
David Farwell, Computing Research Laboratory, New Mexico State University
Chuck Fillmore, International Computer Science Institute, Berkeley
Anette Frank, DFKI
John Fry, SRI International
Eva Hajicova, Center for Computational Linguistics, Charles University, Prague
Erhard W. Hinrichs, University of Tübingen
Ed Hovy, University of Southern California
Baden Hughes, University of Melbourne
Emi Izumi, NICT
Aravind Joshi, University of Pennsylvania, Philadelphia
Sergei Nirenburg, University of Maryland, Baltimore County
Stephan Oepen, University of Oslo
Boyan A. Onyshkevych, U.S. Dept. of Defense
Kyounghee Paik, KLI
Martha Palmer, University of Colorado
Manfred Pinkal, DFKI
Massimo Poesio, University of Essex
Owen Rambow, Columbia University
Peter Rossen Skadhauge, Copenhagen Business School
Beth Sundheim, SPAWAR Systems Center
Jia-Lin Tsai, Tung Nan Institute of Technology
Janyce Wiebe, University of Pittsburgh
Nianwen Xue, University of Pennsylvania

Working Group Organizers:
Adam Meyers, New York University
Mike Maxwell, University of Maryland
Workshop Program

Saturday, 22 July 2006

09:00–09:10 Opening Remarks

09:10–09:30 Challenges for Annotating Images for Sense Disambiguation
Cecilia Ovesdotter Alm, Nicolas Loeff and David A. Forsyth

09:30–10:00 A Semi-Automatic Method for Annotating a Biomedical Proposition Bank
Wen-Chi Chou, Richard Tzong-Han Tsai, Ying-Shan Su, Wei Ku, Ting-Yi Sung and
Wen-Lian Hsu

10:00–10:30 How and Where do People Fail with Time: Temporal Reference Mapping Annotation
by Chinese and English Bilinguals
Yang Ye and Steven Abney

10.30–11.00 Coffee Break

11:00–11:30 Probing the Space of Grammatical Variation: Induction of Cross-Lingual Grammatical Constraints from Treebanks
Felice Dell’Orletta, Alessandro Lenci, Simonetta Montemagni and Vito Pirrelli

11:30–12:00 Frontiers in Linguistic Annotation for Lower-Density Languages
Mike Maxwell and Baden Hughes

12:00–12:30 Annotation Compatibility Working Group Report
Adam Meyers

12:30–14.00 Lunch

14:00–14:30 Manual Annotation of Opinion Categories in Meetings
Swapna Somasundaran, Janyce Wiebe, Paul Hoffmann and Diane Litman

14:30–15:00 The Hinoki Sensebank — A Large-Scale Word Sense Tagged Corpus of Japanese —
Takaaki Tanaka, Francis Bond and Sanae Fujita

15:00–15:30 Issues in Synchronizing the English Treebank and PropBank
Olga Babko-Malaya, Ann Bies, Ann Taylor, Szuting Yi, Martha Palmer, Mitch Marcus, Seth Kulick and Libin Shen

15.30–16.00 Coffee Break

16:00–16:30 On Distance between Deep Syntax and Semantic Representation
Václav Novák

16.30–17.30 Discussion

17.30–17.40 Closing Remarks

Alternate papers

Corpus Annotation by Generation
Elke Teich, John A. Bateman and Richard Eckart

Constructing an English Valency Lexicon
Jiří Semecký and Silvie Cinková