

HLT/NAACL-06

# **Scalable Natural Language Understanding**

**(ScaNaLU 2006)**

**Proceedings of the Workshop**

8 June 2006  
New York City, USA

Production and Manufacturing by  
*Omnipress Inc.*  
2600 Anderson Street  
Madison, WI 53704

©2006 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

## Introduction

This volume contains the papers presented at the HLT/NAACL 2006 workshop entitled: Scalable Natural Language Understanding. The workshop was held on June 8th, 2006 and is the third in a series that started in Heidelberg, Germany on May 23rd and 24th 2002 and continued on May 6th, 2004 at HLT/NAACL in Boston. The papers were refereed by an international panel of experts in the field. The workshop is the second held under the auspices of the HLT/NAACL to be directed at issues concerning the scalability of natural language understanding and generation systems.

There is a growing need for systems that can understand and generate natural language in applications that require substantial amounts of knowledge as well as reasoning capabilities. Most current implemented systems for natural language understanding (NLU) are decoupled from any reasoning processes, which makes them narrow and brittle. Furthermore, they do not appear to be scalable in the sense that the techniques used in such systems do not appear to generalize to more complex applications. While significant work has been done in developing theoretical underpinnings of systems that use knowledge and reasoning (e.g., development of models of linguistic interpretation using abductive reasoning, intention recognition, formal models of dialogue, formal models of lexical and utterance meaning, and utterance planning), it has often proved difficult to utilize such theories in robust working systems.

Another major barrier has been the vast amount of linguistic and world knowledge needed. However, there is now significant progress in compiling the required knowledge, using manual and increasingly automated techniques for ontology and grammar learning. But even as these resources become available, we still lack some key conceptual and computational frameworks that will form the foundation for effective scalable natural language systems, e.g., in terms of incremental processing, dialogical alignment or pragmatics. The collection of researchers who face the challenges involved in scaling human language technology is growing in conjunction with greater efforts to develop systems that robustly interact with users in intuitive and conversational ways.

We wish to thank the organizers of HLT/NAACL 2006 for their professional support and the members of the Program Committee for reviewing the submissions on a very tight schedule.

James Allen, University of Rochester  
Jan Alexandersson, German Research Center for Artificial Intelligence, Saarbücken  
Jerome Feldman, International Computer Science Institute and University of Berkeley  
Robert Porzel, European Media Laboratory, Heidelberg



**Organizers:**

James Allen, University of Rochester  
Jan Alexandersson, German Research Center for Artificial Intelligence, Saarbrücken  
Jerome Feldman, International Computer Science Institute and University of Berkeley  
Robert Porzel, European Media Laboratory, Heidelberg

**Program Committee:**

Tilman Becker, German Research Center for Artificial Intelligence, Saarbrücken  
Paul Buitelaar, German Research Center for Artificial Intelligence, Saarbrücken  
Aldo Gangemi, Laboratory for Applied Ontology, Rome  
Dan Gildea, University of Rochester  
Gnter Grz, University of Erlangen  
Joachim de Beule, University of Brussels  
Dan Flickinger, Oslo, Saarland and Stanford Universities  
Berenike Loos, European Media Laboratory, Heidelberg  
Rainer Malaka, University of Bremen  
Tomasz Marciniak, Language Computer Corporation, Richardson, Texas  
Vanessa Micelli, European Media Laboratory, Heidelberg  
Stanley Peters, Stanford University  
Massimo Poesio, University of Essex  
Len Schubert, University of Rochester  
Hans Uszkoreit, Saarland University  
Hans-Peter Zorn, European Media Laboratory, Heidelberg

**Invited Speaker:**

Luc Steels, University of Brussels and Sony Computer Science Laboratory, Paris



## Table of Contents

<i>Pragmatic information extraction from subject ellipsis in informal English</i> Shigeko Nariyama .....	1
<i>Backbone Extraction and Pruning for Speeding Up a Deep Parser for Dialogue Systems</i> Myroslava O. Dzikovska and Carolyn P. Ros .....	9
<i>Understanding Complex Natural Language Explanations in Tutorial Applications</i> Pamela W. Jordan, Maxim Makatchev and Umarani Pappuswamy .....	17
<i>Increasing the coverage of a domain independent dialogue lexicon with VERBNET</i> Benoit Crabb, Myroslava O. Dzikovska, William de Beaumont and Mary Swift .....	25
<i>Scaling Natural Language Understanding via User-driven Ontology Learning</i> Berenike Loos .....	33
<i>Catching Metaphors</i> Matt Gedigian, John Bryant, Srini Narayanan and Branimir Ciric .....	41
<i>Scaling Construction Grammar up to Production Systems: the Situated Constructional Interpretation Model</i> Guillaume Pitel .....	49
<i>Searching for Grammar Right</i> Vanessa Micelli .....	57
<i>Embodied construction grammar as layered modal languages</i> Anders Sogaard .....	65
<i>A (very) Brief Introduction to Fluid Construction Grammar</i> Luc Steels and Joachim de Beule .....	73





# Conference Program

**Thursday, June 8, 2006**

8:45–9:00 Welcome: Robert Porzel

## **Session 1: Elision, Deep Parsing, and Tutoring**

9:00–9:30 Pragmatic Information Extraction from Subject Ellipsis in Informal English: Shigeko Nariyama

9:30–10:00 Backbone Extraction and Pruning for Speeding Up a Deep Parser for Dialogue Systems: Myroslava O. Dzikovska and Carolyn P. Rosé

10:00–10:30 Understanding Complex Natural Language Explanations in Tutorial Applications: Pamela W. Jordan, Maxim Makatchev and Umarani Pappuswamy

10:30–11:00 Coffee Break

## **Session 2: Lexica, Ontologies and Metaphors**

11:00–11:30 Increasing the Coverage of a Domain-independent Dialogue Lexicon with VERB-NET: Benoit Crabbé, Myroslava O. Dzikovska, William de Beaumont and Mary Swift

11:30–12:00 Scaling Natural Language Understanding via User-driven Ontology Learning: Berenike Loos

12:00–12:30 Catching Metaphors: Matt Gedigian, John Bryant, Srinu Narayanan, and Branimir Ciric

12:30–14:00 Lunch Break

## **Session 3: Formal and Computational Construction Grammar**

14:00–14:30 Scaling Construction Grammar up to Production Systems: the Situated Constructional Interpretation Model: Guillaume Pitel

14:30–15:00 Searching for Grammar Right: Vanessa Micelli

15:00–15:30 Embodied Construction Grammar as Layered Modal Languages: Anders Søgaard

15:30–16:00 Coffee Break

## **Session 4: Invited Talk, Demo Session and Closure**

16:00–16:45 A (very) Brief Introduction to Fluid Construction Grammar: Luc Steels and Joachim de Beule (Invited Talk)

16:45–17:15 Demo Session

17:15–17:30 Discussion (Next Workshop)

