

Proceedings of the 2006

Australasian Language Technology Workshop

Sancta Sophia College, Sydney
Nov 30-Dec 1, 2006

Sponsored by

HCSNet

ALTA



Editors:

Lawrence Cavedon

Ingrid Zukerman

<http://www.altas.asn.au/events/altw2006/>

ISBN: 1-74108-146-7

Preface

This volume contains the papers accepted for presentation at the Australasian Language Technology Workshop (ALTW) 2006, held at the University of Sydney, Sydney, Australia, on November 30 – December 1, 2006. This is the fourth annual installment of the workshop in its most-recent incarnation, and the continuation of an annual workshop series that has existed under various guises since the early 90s.

The goals of the workshop are:

- to bring together the growing Language Technology (LT) community in Australia and New Zealand and encourage interactions;
- to foster interaction between academic and industrial researchers;
- to encourage dissemination of research results for new and ongoing projects; and
- to increase the visibility of LT research in Australia and New Zealand.

This year's Australasian Language Technology Workshop includes regular talks as well as poster presentations and student posters. Of the 36 papers submitted, 19 papers were selected by the program committee for publication and appear in these proceedings. Of these, 13 are oral presentations papers and 6 are poster presentations. Additionally, we have included 6 student posters to encourage feedback on early results. Each full-length submission was independently peer reviewed by at least two members of the international program committee, in accordance with the DEST requirements for E1 conference publications.

We would like to thank all the authors who submitted papers, as well as the members of the program committee for the time and effort they contributed in reviewing the papers. Our thanks also go to local organizer Rolf Schwitter, to Jiawen Rong for installing and configuring the submission management system, and to members of the ALTA executive for their support in organizing the workshop.

This year, ALTW is held as part of the HCSNet (ARC Network in Human Communication Science) SummerFest. We thank HCSNet for its invaluable financial and organisational support, including organising the venue and refreshments, printing the proceedings, handling registration, and particularly for providing travel funds to all presenters, including the invited international speaker: Michael Johnston (AT&T Labs Research). A special thanks goes to Kym Buckley of HCSNet for her immense effort in managing the organisation of the HCSNet events, including ALTW.

Lawrence Cavedon and Ingrid Zukerman
Program Chairs

Committees

Program Chairs:

Lawrence Cavedon (National ICT Australia and RMIT University, Australia)
Ingrid Zukerman (Monash University)

Local Chair:

Rolf Schwitter (Macquarie University)

Program Committee:

Our thanks to the following Program Committee members for their efforts in the reviewing process.

Jan Alexandersson (DFKI)	Timothy Baldwin (University of Melbourne)
Steven Bird (University of Melbourne)	Frances Bond (NTT Communication Science Labs)
Stephen Clark (Oxford University)	Nathalie Colineau (CSIRO ICT Centre)
James Curran (University of Sydney)	Robert Dale (Macquarie University)
Mark Dras (Macquarie University)	Mark Ellison (Analith)
Dominique Estival (Appen)	Tanja Gaustad (Appen)
Graeme Hirst (University of Toronto)	Baden Hughes (University of Melbourne)
Michael Johnston (ATT Labs Research)	Arne Jonsson (Linkoping University)
Dan Jurafsky (Stanford University)	Min-Yen Kan (National University of Singapore)
Alistair Knott (University of Otago)	Kazunori Komatani (Kyoto University)
Kathleen McCoy (University of Delaware)	Chris Manning (Stanford University)
Yuval Marom (Monash University)	Diego Molla (Macquarie University)
Ng Hwee Tou (National University of Singapore)	Cecile Paris (CSIRO ICT Centre)
Matthew Purver (Stanford University)	Tony Smith (Waikato University)
Nicola Stokes (NICTA Victoria)	Seyed M.M. Tahaghoghi (RMIT University)
Takaaki Tanaka (NTT Communication Science Labs)	Stephen Wan (Macquarie University and CSIRO)
Menno van Zaanen (Macquarie University)	

Table of Contents

Invited Presentations

Robust Multimodal Understanding for Interactive Systems 1
Michael Johnston

User Modelling for Language Technologists 2
Judy Kay

Full Papers

Efficient Combinatory Categorical Grammar Parsing 3
Bojan Djordjevic, James R. Curran

Improved Default Sense Selection for Word Sense Disambiguation 11
Tobias Hawker, Matthew Honnibal

Experiments with Sentence Classification 18
Anthony Khoo, Yuval Marom, David Albrecht

Computational Semantics in the *Natural Language Toolkit* 26
Ewan Klein

Classifying Speech Acts using Verbal Response Modes 34
Andrew Lampert, Robert Dale, Cécile Paris

Word Relatives in Context for Word Sense Disambiguation 42
David Martinez, Eneko Agirre, Xinglong Wang

Named Entity Recognition for Question Answering 51
Diego Mollá, Menno van Zaanen, Daniel Smith

Named Entity Recognition for Astronomy Literature 59
Tara Murphy, Tara McIntosh, James R. Curran

Die Morphologie (f): Targeted Lexical Acquisition for Languages other than English 67
Jeremy Nicholson, Timothy Baldwin, Phil Blunsom

Automatic Mapping Clinical Notes to Medical Terminologies 75
Jon Patrick, Yefeng Wang, Peter Budd

Pseudo Relevance Feedback Using Named Entities for Question Answering 83
Luiz Augusto Pizzato, Diego Mollá, Cécile Paris

Evidence for Gradient Salience: What Happens with Competing Non-salient Referents during
Pronoun Resolution? 91
Ralph L. Rose

Web Readability and Computer-Assisted Language Learning	99
<i>Alexandra L. Uitdenbogerd</i>	
Questions Require an Answer: A Deductive Perspective on Questions and Answers	107
<i>Willemijn Vermaat</i>	
Towards the Evaluation of Referring Expression Generation	115
<i>Jette Viethen, Robert Dale</i>	
Error Correction Using Utterance Disambiguation Techniques	123
<i>Peter Vlugter, Edwin van der Ham, Alistair Knott</i>	
Using Dependency-Based Features to Take the “Para-farce” out of Paraphrase	131
<i>Stephen Wan, Mark Dras, Robert Dale, Cécile Paris</i>	
Verb Sense Disambiguation Using Selectional Preferences Extracted with a State-of-the-art Semantic Role Labeler	139
<i>Patrick Ye, Timothy Baldwin</i>	
This Phrase-Based SMT System is Out of Order: Generalised Word Reordering in Machine Translation	149
<i>Simon Zwarts, Mark Dras</i>	
Student Posters	
Analysis and Prediction of User Behaviour in a Museum Environment	157
<i>Karl Grieser, Timothy Baldwin, Steven Bird</i>	
Using Dialogue Acts to Suggest Responses in Support Services via Instant Messaging	159
<i>Edward Ivanovic</i>	
Probabilities Improve Stress-Prediction in a CFG of Hawaiian Phonology	161
<i>‘Ōiwi Parker Jones</i>	
Towards Cognitive Optimisation of a Search Engine Interface	163
<i>Kenneth Treharne, Darius Pfitzner, David M. W. Powers</i>	
Natural Language Processing and XML Retrieval	165
<i>Alan Woodley, Xavier Tannier, Marcus Hassler, Shlomo Geva</i>	
Extracting Patient Clinical Profiles from Case Reports	167
<i>Yitao Zhang, Jon Patrick</i>	

Thursday November 30 2006

- **9.00am – 10.45am: Session 1**
 - **Welcome and Opening**
 - Evidence for gradient salience: What happens with competing non-salient referents during pronoun resolution?, *Ralph L. Rose*
 - Error correction using utterance disambiguation techniques, *Peter Vlugter, Edwin van der Ham, and Alistair Knott*
 - Classifying speech acts using verbal response modes, *Andrew Lampert, Robert Dale, and Cecile Paris*
- **10.45am – 11.15am: Coffee break**
- **11.15am – 12.45pm: Session 2**
 - Verb sense disambiguation using selectional preferences extracted with a state-of-the-art semantic role labeler, *Patrick Ye and Timothy Baldwin*
 - Word relatives in context for word sense disambiguation, *David Martinez, Eneko Agirre, and Xinglong Wang*
 - Die morphologie (f): targeted lexical acquisition for languages other than English, *Jeremy Nicholson, Timothy Baldwin, and Phil Blunsom*
- **12.45pm – 2.00pm: Lunch (provided by HCSNet)**
- **2.00pm – 3.30pm: Poster Session**

Accepted papers

- Improved default sense selection for word sense disambiguation, *Tobias Hawker and Matthew Honnibal*
- Automatic mapping clinical notes to medical terminologies, *Jon Patrick, Yefeng Wang and Peter Budd*
- Pseudo relevance feedback using named entities for Question Answering, *Luiz Augusto Pizzato, Diego Molla, and Cecile Paris*
- Web readability and Computer-Assisted Language Learning, *Alexandra L. Uitdenbogerd*
- Questions require an answer: A deductive perspective on questions and answers, *Willemijn Vermaat*
- Towards the evaluation of referring expression generation, *Jette Viethen and Robert Dale*

Student posters

- Analysis and prediction of user behaviour in a museum environment, *Karl Grieser, Tim Baldwin and Steven Bird*
- Using dialogue acts to suggest responses in support services via Instant Messaging, *Edward Ivanovic*
- Probabilities improve stress-prediction in a CFG of Hawaiian phonology, *'Oiwī Parker Jones*
- Towards cognitive optimisation of a search engine interface, *Kenneth Treharne, Darius Pfitzner and David M. W. Powers*
- Natural language processing and XML retrieval, *Alan Woodley, Xavier Tannier, Marcus Hassler, and Shlomo Geva*
- Extracting patient clinical profiles from case reports, *Yitao Zhang and Jon Patrick*

- **3.30pm – 4.00pm: Coffee break**
- **4.00pm – 5.30pm: Session 4**
 - Efficient combinatory categorial grammar parsing, *Bojan Djordjevic and James R. Curran*
 - **HCSNet Keynote Presentation:**
Robust multimodal understanding for interactive systems
Michael Johnston (AT&T Labs Research)
- **5.30pm—6.00pm: Drinks (courtesy of HCSNet)**
- **6.00pm: Musical Recital (courtesy of HCSNet)**

Friday December 1 2006

- **9.00am—10.00am:**
 - **HCSNet Keynote Presentation:**
Human-computer interaction based on acoustic signals, muscle movements, and brainwaves, *Tanja Schultz (Carnegie Mellon University)*
- **10.15am—10.45am: Session 5**
 - Computational semantics in the *Natural Language Toolkit*, *Ewan Klein*
- **10.45am—11.15am: Coffee break**
- **11.45am – 12.45pm: Session 6**
 - Experiments with sentence classification, *Anthony Khoo, Yuval Marom and David Albrecht*
 - **Invited Presentation:** User modeling for language technologists
Judy Kay (University of Sydney)
- **12.45pm – 2.00pm: Lunch (provided by HCSNet)**
- **2.00pm – 3.30pm: Session 7**
 - Named entity recognition for Question Answering, *Diego Molla, Menno van Zaanen, and Daniel Smith*
 - Named entity recognition for astronomy literature, *Tara Murphy, Tara McIntosh, and James R. Curran*
 - This phrase-based SMT system is out of order: Generalised word reordering in machine translation, *Simon Zwarts and Mark Dras*
- **3.30pm – 4.00pm: Coffee break**
- **4.00pm – 4.45pm: Session 8**
 - Using dependency-based features to take the "para-farce" out of paraphrase, *Stephen Wan, Mark Dras, Robert Dale, and Cecile Paris*
 - **Closing Remarks**
- **4.45pm—5.45pm: ALTA Annual General Meeting**