Proceedings of the
Australasian Language Technology Workshop 2004

December 8th, 2004

Macquarie University
Sydney, Australia

Editors:

Ash Asudeh, Cécile Paris, Stephen Wan

Sponsors:

Australian Government
Department of Defence
Defence Science and
Technology Organisation

Proceedings of the Australasian Language Technology Workshop 2004

Website: http://www.alta.asn.au/events/altw2004/publication/index.html

Published by:

The Australian Speech Science and Technology Association (ASSTA)
Email: secretary@assta.org
or
G.P.O. Box 143, Canberra City, ACT, 2601.

Contact for additional copies:

ALTA Workshop 2004
Email: workshop@alta.asn.au

ISBN: 0 9581946 1 0

Copyright(c) 2004 ASSTA and individual authors

This work is copyright. Apart from any use as permitted under the Copyright Act 1968, no part may be reproduced by any process without written permission from the copyright holders.

Australasian Language Technology Workshop 2004

Foreword

ALTW 2004 is the second workshop organised by the recently founded Australasian Language Technology Association (ALTA), following previous Australasian Natural Language Processing Workshops. The first Australasian Natural Language Processing Workshop was held at Macquarie University, in 2001 and the second in Canberra in 2002. The first ALTW was held in Melbourne in 2003, in conjunction with a Summer School in Language Technology (ALTSS). ALTW is now ALTA's annual meeting. This year, ALTW and the Summer School in Language Technology are being held in conjunction with the Tenth Australian International Conference on Speech Science & Technology (SST 2004).

Every year, the number of submissions (including submissions from overseas) increases significantly, indicating both the growth of the research area in Australasia and the higher international visibility of the workshop. As a result, every year the standard of the workshop also increases. This year, as in previous years, each submission was a full paper and was reviewed by at least two members of an international programme committee of leading researchers in language technology, listed on the next page. We would like to thank the programme committee for their dedicated and careful review of the papers. The call for papers for ALTW solicited papers presenting substantial, original and unpublished research on all aspects of language technology. We have selected 22 papers for the workshop, which is organised in two parallel sessions.

We would like to thank the local organisers of ALTW/ALTSS and SST for their help in organising this workshop, in particular Diego Molla and Steve Cassidy. We would also like to thank the ALTA executive committee for assistance in various respects. Finally, we would like to thank the sponsors (DSTO, Appen, and VeCommerce) for their generous help in supporting the workshop, and Dominique Estival as the sponsorship coordinator.

As a final word, we hope you will all enjoy the workshop. We are very excited to see `that the language technology community in Australia and the region is vibrant and growing. We hope the workshop will once again foster interactions amongst researchers in language technology research.

Ash Asudeh, Cécile Paris and Stephen Wan
ALTW 2004 programme chairs
November 2004

ISBN: 0 9581946 1 0

Program Committee

Ash Asudeh, University of Canterbury (NZ) (Co-chair)
Cecile Paris, CSIRO (AU) (Co-chair)
Stephen Wan, CSIRO and Macquarie University (AU) (Student Chair)
Steven Bird, University of Melbourne (AU)
Steve Cassidy, Macquarie University (AU)
Nathalie Colineau, CSIRO (AU)
Dominique Estival, DSTO (AU)
James Curran, Sydney University (AU)
Alistair Knott, University of Otago (NZ)
Mirella Lapata, University of Sheffield (UK)
Corrin Lakeland, University of Otago (NZ)
Nadine Ozkan, Scansoft (Canada)
Hiroshi Masuichi, Fujixerox (JP)
Daniel Midgley, University of Western Australia (AU)
Harold Somers, UMIST (UK)

Table of Contents:

Ari Chanen, Jon Patrick	Complex, Corpus-Driven, Syntactic Features for Word Sense Disambiguation	1–8
Diego Molla, Mary Gardiner	Answerfinder: Question Answering by Combining Lexical, Syntactic and Semantic Information	9–16
David Bell, Jon Patrick	Using WordNet Domains In A Supervised Learning Word Sense Disambiguation System	17–24
Luiz Augusto, Sangoi Pizzato	Using a Trie-based Structure for Question Analysis	25–31
Mingfang Wu, Ross Wilkinson, Cecile Paris	An Evaluation on Query-biased Summarisation for the Question Answering Task	32–38
Jon Patrick, Pham Hong Nguyen	Thin Parsing: A Balance between Wide Scale Parsing and Chunking	39–46
Corrin Lakeland, Alistair Knott	Implementing a lexicalised statistical parser	47–54
Rolf Schwitter, Marc Tilbrook	Controlled Natural Language meets the SemanticWeb	55–62
Andrew Lampert, Cecile Paris	Information Assembly for Automatic Content Adaptation	63–70
Bernd Bohnet, Robert Dale	Referring Expression Generation as a Search Problem	71–77
Mark Foreman, Daniel McMichael	A Meta-grammar for CCG	78–84
Cecile Paris, Nathalie Colineau, Dominique Estival	Intelligent Multi Media Presentation of information in a semi-immersive Command and Control environment	85–92

Casey Whitelaw, Jon Patrick	Selecting Systemic Features for Text Classification	93–100
Pont Lurcock, Peter Vlugter, Alistair Knott	A framework for utterance disambiguation in dialogue	101–108
Phil Blunsom	Maximum Entropy Markov Models for Semantic Role Labelling	109–116
Yuanyong Wang, Achim Hoffmann	A New Measure for Extracting Semantically Related Words	117–122
David Penton, Steven Bird	Representing and Rendering Linguistic Paradigms	123–130
Maarten van Schagen, Alistair Knott	Tauira: A tool for acquiring unknown words in a dialogue context	131–138
Catherine Lai, Steven Bird	Querying and Updating Treebanks: A Critical Survey and Requirements Analysis	139–146
Matthew Honnibal	Converting the Penn Treebank to Systemic Functional Grammar	147–154
Patrick Ye	Selection Preference Based Verb Sense Disambiguation Using WordNet	155–162
Jon Patrick, Jeremy Fletcher	Differentiating Types of Verb Particle Constructions	163–170