New Methods in Language Processing

and

Computational Natural Language Learning

NeMLaP3/CoNLL98

Editor: David M. W. Powers

Organized by: Macquarie University

Sydney University

Flinders University

Held at Macquarie University, January 11-17, 1998

Proceedings of the

Joint Conference on

New Methods in Language Processing

and

Computational Natural Language Learning

NeMLaP3/CoNLL98

Editor: David M. W. Powers

Organized by: Macquarie University
Sydney University
Flinders University

Held at Macquarie University, January 11-17, 1998

CoNLL is the annual conference of the ACL Special Interest Group on Natural Language Learning (SIGNLL)

Order additional copies from

ACL
P.O. Box 6090
Somerset, NJ, 08875 USA
+1-908-873-3898
acl@bellcore.com

ISBN: 0-7258-0634-6

PREFACE

Natural Language Processing is increasingly moving away from the idea of handcoding grammars and systems, and the NeMLaP series of conferences has encouraged the exploration of new approaches to all aspects of the field. The techniques NeMLaP portrays tend to move away from the symbolic toward the connectionist and the statistical, and often have a learning flavour about them.

The ACL Special Interest Group on Natural Language Learning last year initiated the CoNLL series which focus specifically on Learning, so it is highly appropriate that the two conferences, originally scheduled for two different cities in Australia, should be held jointly to maximize the interaction between the two groups.

The timing of the conference is a bit unusual for those used to conferences in the northern summer, and a bit difficult for those who only have short breaks. But attending conferences in our winter semester is something Australians are used to, and the European members of the organizing committee were keen on the idea of escaping their winter for a while. The distance is also a big factor, and combining a joint conference with a series of other events has enabled us to maintain both a good number of papers and a high level of refereeing - at the main conferences we have 32 papers being presented out of 40 odd submissions.

This volume is a bit unusual in one respect, in that we are including the papers from the associated workshops: the Workshop on Human Computer Conversation which is being held in association with the Loebner Prize in Sydney before the joint conference, and the Workshop on Paradigms and Grounding in Language Learning which is being held in Adelaide following the conferences. We felt that registrants and purchasers of the proceedings would appreciate having the additional papers, given that the cost of adding them was less than the cost of producing separate proceedings for the workshops.

We trust that you will enjoy your time in Australia and will find the events worthwhile. We look forward to seeing you.

David Powers
Flinders University

ORGANIZERS

David Powers, Flinders Uni, Australia
Sandra Williams, Macquarie Uni/Microsoft Research Inst, Australia
Harold Somers, UMIST, UK
Chris Manning, Sydney Uni, Australia
Kemal Oflazer, Bilkent Uni, Turkey
Robert Dale, Macquarie Uni/Microsoft Research Inst, Australia

PROGRAM COMMITTEE

Michael Brent, Johns Hopkins Uni, USA
Claire Cardie, Cornell Uni, USA
Walter Daelemans, Tilburg Uni, NL
Robert Dale, Macquarie Uni
Mark Ellison, Edinburgh Uni, UK
Dominique Estival, Melbourne Uni, Australia
Chris Manning, Sydney Uni, Australia
Kemal Oflazer, Bilkent Uni, Turkey
Raymond Mooney, University of Texas at Austin, USA
Kim Plunkett, Oxford University, UK
David Powers, Flinders Uni, Australia
Christer Samuelsson, Bell Labs, USA
Jeffrey M. Siskind, NEC Research Institute, Princeton, USA
Harold Somers, UMIST, UK
Junichi Tsujii, Tokyo Uni, Japan
Antal van den Bosch, Tilburg University, The Netherlands
Atro Voutilainen, Helsinki Uni, Finland
Peter Wallis, DSTO, Australia
Dekai Wu, HKUST, Hong Kong
David Yarowsky, Johns Hopkins Uni, USA
Ingrid Zukerman, Monash Uni, Australia

NeMLaP3/CoNLL98

Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning

January 11-17, 1998
Macquarie University
SYDNEY, AUSTRALIA

TABLE OF CONTENTS

Main Conference

HCC Workshop

PaGiLL Workshop

Tutorial TIMETABLE

MONDAY, 12 JANUARY (BLUE MOUNTAINS)

10.00	
	Christer Samuelsson *Introduction to Statistical Methods in Natural Language Processing*
13.00	**Lunch**
14.00	**Guided Walk - Grand Canyon**
17.00	
	Robert Dale *Introduction to Natural Language Generation*
20.00	**Dinner and relaxation**

TUESDAY, 13 JANUARY (BLUE MOUNTAINS)

7.00	**Early morning stroll / Bird Watching**
8.00	**Breakfast**
9.00	
	Walter Daelemans *Introduction to Memory-Based Language Technology*
12.00	**Lunch**
14.00	**Guided Walk - Grand Canyon**
17.00	
	Dominique Estival *Introduction to Grammatical Formalisms for Natural Language Processing*
20.00	**Dinner and relaxation**

TUESDAY, 20 JANUARY (MELBOURNE)

14.00	
	David Dowe *Introduction to Snob, MML and Mixture Modelling*

HCC TIMETABLE

WEDNESDAY, 14 JANUARY

11.30	
	Jason L. Hutchens and Michael D. Alder *Introducing MegaHaL* *271*
	Véronique Bastin and Denis Cordier *Methods and Tricks Used in an Attempt to Pass the Turing Test* *275*
12.30	**Lunch**
14.00	
	David M. W. Powers *The Total Turing Test and the Loebner Prize* *279*
	Zenshiro Kawasaki, Keiji Takida and Masato Tajima *Language Model and Sentence Structure Manipulations for Natural Language Applications Systems* *281*
	Discussion
15.30	**Break**
16.30	**BrainStorming session**
17.30	

NeMLaP3/CoNLL98 TIMETABLE

THURSDAY, 15 JANUARY

FRIDAY, 16 JANUARY

SATURDAY, 17 JANUARY

PaGiLL TIMETABLE

WEDNESDAY, 21 JANUARY

11.30

Tony C. Smith
Learning Feature-Value Grammars from Plain Text *291*

Hervé Déjean
Morphemes as Necessary Concept for Structures Discovery from Untagged Corpora *295*

12.30 **Lunch**

14.00

Christopher D. Manning
The Segmentation Problem in Morphology Learning *299*

David M. W. Powers
Reconciliation of Unsupervised Clustering, Segmentation and Cohesion *307*

Isabelle Tellier
Syntactico-Semantic Learning of Categorical Grammars *311*

15.30 **Break**

16.30 **BrainStorming session**

17.30

AUTHOR INDEX