

**EACL'99**

**CoNLL-99**

**Computational Natural Language  
Learning**

**Proceedings of a Workshop Sponsored by  
The Association for Computational Linguistics**

**Editors: Miles Osborne & Erik Tjong Kim Sang**

12 June 1999  
Universty of Bergen  
Bergen, Norway



CoNLL99

Computational Natural Language  
Learning

Bergen, Norway 12th June, 1999

Editors: Miles Osborne & Erik Tjong Kim Sang

In conjunction with the *Special Interest Group in Natural Language Learning* and the TMR Project *Learning Computational Grammars*

Collection copyright ACL. Authors retain individual copyright. Order additional copies from:

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

## Preface

Welcome to CoNLL99 (<http://lcg-www.uia.ac.be/conll99>), the third meeting of the *Special Interest Group in Natural Language Learning*. Regular papers accepted this year deal with orthography, morphology, syntax and grammatical relations, using technologies based around information theory, transformation-based learning, instance-based learning and semi-automated knowledge acquisition. In addition to these regular papers, CoNLL99 has as a special theme the task of recovering Noun Phrases from free text. On this topic, we are pleased to have Lance Ramshaw as our guest speaker.

In addition to the papers in the proceedings dealing with Noun Phrase identification, the meeting will include short reports by other groups. The CoNLL99 web site will contain updated information.

We would like to thank all the people who helped contribute to this workshop: our guest speaker, the programme committee, SiGNLL, LCG (funded by the TMR programme of the European Union), our EACL hosts at Bergen, and finally, and most importantly, all the authors for their contributions.

Miles Osborne and Erik Tjong Kim Sang.

## Chairs

Miles Osborne (U. Groningen)  
Erik Tjong Kim Sang (U. Antwerp)

## Programme Committee

Antal van den Bosch (U. Tilburg)  
Ted Briscoe (U. Cambridge)  
Walter Daelemans (U. Antwerp)  
John Nerbonne (U. Groningen)  
David Powers (U. Flinders)  
Christer Samuelsson (Xerox)  
Jeffrey Mark Siskind (NEC)

## Invited Speaker

Lance Ramshaw (BBN).

## Table of Contents

<b>Chunyu Kit and Yorick Wilks</b> Unsupervised Learning of Word Boundary with Description Length Gain	1-6
<b>André Kempe</b> Experiments in Unsupervised Entropy-Based Corpus Segmentation	7-13
<b>Kemal Oflazer and Sergei Nirenburg</b> Practical Bootstrapping of Morphological Analyzers	14-23
<b>Stephan Raaijmakers</b> Finding Representations for Memory-Based Language Learning	24-32
<b>Torbjörn Lager</b> The $\mu$ -TBL System: Logic Programming Tools for Transformation-Based Learning	33-42
<b>Lisa Ferro, Marc Vilain and Alexander Yeh</b> Learning Transformation Rules to Find Grammatical Relations	43-52
<b>Walter Daelemans, Sabine Buchholz and Jorn Veenstra</b> Memory-Based Shallow Parsing	53-60
<b>Miles Osborne</b> MDL-based DCG Induction for NP Identification	61-68

## Timetable

- 09.00 - 09.30 Unsupervised Learning of Word Boundary  
with Description Length Gain  
**Chunyu Kit and Yorick Wilks**
- 09.30 - 10.00 Experiments in Unsupervised Entropy-Based  
Corpus Segmentation  
**André Kempe**
- 10.00 - 10.30 Practical Bootstrapping of Morphological Analyzers  
**Kemal Oflazer and Sergei Nirenburg**
- 10.30 - 11.00 break
- 11.00 - 12.00 Invited talk (title to be announced)  
**Lance Ramshaw**
- 12.00 - 12.30 Finding Representations for Memory-based  
Language Learning  
**Stephan Raaijmakers**
- 12.30 - 13.30 Lunch
- 13.30 - 14.00 The  $\mu$ -TBL System: Logic Programming Tools for  
Transformation-Based Learning  
**Torbjörn Lager**
- 14.00 - 14.30 Learning Transformation Rules to Find  
Grammatical Relations  
**Lisa Ferro, Marc Vilain and Alexander Yeh**
- 14.30 - 15.00 Memory-Based Shallow Parsing  
**Walter Daelemans, Sabine Buchholz  
and Jorn Veenstra**
- 15.00 - 15.30 break
- 15.30 - 16.00 MDL-based DCG Induction for NP Identification  
**Miles Osborne**
- 16.00 - 16.30 Results CoNLL task  
(speakers to be announced)
- 16.30 - 17.00 Closing