

## Preface

This volume has its origins in the first Learning Language in Logic (LLL) workshop which took place on 30 June 1999 in Bled, Slovenia immediately after the Ninth International Workshop on Inductive Logic Programming (ILP'99) and the Sixteenth International Conference on Machine Learning (ICML'99). LLL is a research area lying at the intersection of computational linguistics, machine learning and computational logic. As such it is of interest to all those working in these three fields. I am pleased to say that the workshop attracted submissions from both the Natural Language Processing (NLP) community and the ILP community, reflecting the essentially multi-disciplinary nature of LLL. Eric Brill and Ray Mooney were invited speakers at the workshop and their contributions to this volume reflect the topics of their stimulating invited talks.

After the workshop authors were given the opportunity to improve their papers, the results of which are contained here. However, this volume also includes a substantial amount of two sorts of additional material. Firstly, since our central aim is to introduce LLL work to the widest possible audience, two introductory chapters have been written. Džeroski, Cussens and Manandhar provide an introduction to ILP and LLL and Thompson provides an introduction to NLP. In both cases no previous knowledge is assumed. Secondly, to give a full account of research in LLL we invited a number of researchers from the NLP and machine learning communities to contribute. We are pleased to say that all those invited contributed—resulting in the contributions from Tjong Kim Sang & Nerbonne; Kazakov; Eineborg & Lindberg; Riezler and Thompson & Califf.

A LLL home page has been created at <http://www.cs.york.ac.uk/aig/lll> which provides pointers to papers, research group, datasets and algorithms. A page for this volume is at <http://www.cs.york.ac.uk/aig/lll/book>.

We have many people to thank for bringing this volume about. Thanks firstly to all those authors who contributed to this volume. Thanks also to the LLL'99 programme committee who provided the initial reviews for many of the papers here. We greatly appreciate the generous financial support towards the LLL'99 workshop from the ILP2 project and the MLNet-II, ILPNet2 and CompulogNet Networks of Excellence. Sašo Džeroski is supported by the Slovenian Ministry of Science and Technology. Thanks also to Chris Kilgour and Heather Maclaren at York for help with Word. Alfred Hofmann and colleagues at Springer deserve thanks for efficiently producing this book. Finally, we would like to thank Stephen Muggleton. Stephen not only pioneered many of the ILP techniques used by contributors to this volume, he also initiated much of the interaction between the NLP and ILP community which is central to LLL—indeed the slogan “Learning Language in Logic” is due to Stephen. We are very glad that he agreed to provide the foreword to this volume.

July 2000

James Cussens & Sašo Džeroski

# Table of Contents

---

## I Introductions & Overviews

---

- An Introduction to Inductive Logic Programming and Learning Language  
in Logic ..... 3  
*Sašo Džeroski, James Cussens, Suresh Manandhar*
- A Brief Introduction to Natural Language Processing for Non-Linguists ... 36  
*Cynthia A. Thompson*
- A Closer Look at the Automatic Induction of Linguistic Knowledge ..... 49  
*Eric Brill*
- Learning for Semantic Interpretation: Scaling Up Without Dumbing Down 57  
*Raymond J. Mooney*

---

## II Morphology & Phonology

---

- Learning to lemmatise Slovene words ..... 69  
*Sašo Džeroski, Tomaž Erjavec*
- Achievements and Prospects of Learning Word Morphology with Inductive  
Logic Programming ..... 89  
*Dimitar Kazakov*
- Learning the Logic of Simple Phonotactics ..... 110  
*Erik F. Tjong Kim Sang, John Nerbonne*

---

## III Syntax

---

- Grammar Induction as Substructural Inductive Logic Programming ..... 127  
*Pieter Adriaans, Erik de Haas*
- Experiments in Inductive Chart Parsing ..... 144  
*James Cussens, Stephen Pulman*
- ILP in Part-of-Speech Tagging — An Overview ..... 159  
*Martin Eineborg, Nikolaj Lindberg*
- Iterative Part-of-Speech Tagging ..... 172  
*Alípio Jorge, Alneu de Andrade Lopes*
- DCG Induction using MDL and Parsed Corpora ..... 186  
*Miles Osborne*

Learning Log-Linear Models on Constraint-Based Grammars for Disambiguation .....	201
<i>Stefan Riezler</i>	

Unsupervised Lexical Learning with Categorical Grammars Using the LLL Corpus .....	220
<i>Stephen Watkinson, Suresh Manandhar</i>	

---

## **IV Semantics**

---

Induction of Recursive Transfer Rules .....	239
<i>Henrik Boström</i>	

Learning for Text Categorization and Information Extraction with ILP ...	249
<i>Markus Junker, Michael Sintek, Matthias Rinck</i>	

Corpus-based Learning of Semantic Relations by the ILP System, Asium ..	261
<i>Claire Nedellec</i>	

Improving Learning by Choosing Examples Intelligently in Two Natural Language Tasks .....	281
<i>Cynthia A. Thompson, Mary Elaine Califf</i>	

<b>Author Index</b> .....	303
---------------------------	-----