WILS 2012

The NAACL-HLT Workshop on the Induction of Linguistic Structure

Proceedings of the Workshop

June 7, 2012
Montréal, Canada

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53707
USA

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-20-6 / 1-937284-20-4

Introduction

Welcome to the proceedings of the NAACL Workshop on the Induction of Linguistic Structure (WILS). This workshop solicited papers addressing the challenges of learning in an unsupervised or minimally supervised context with questions of linguistic structure. Inducing structured linguistic representations from text has long been a fundamental problem in Computational Linguistics and Natural Language Processing, drawing from theoretical Computer Science and Machine Learning. The popularity of the area is driven by two different motivations. Firstly, it can help us to better understand the cognitive process of language acquisition in humans. Secondly, it can help with portability of NLP applications into new domains and new languages. Most NLP algorithms rely on syntactic parse structure created by supervised methods, however in many cases there is no available training data, thus limiting the portability of these algorithms. Consequently work on unsupervised induction of the linguistic structure of language holds considerable promise, although current approaches are a long way from solving the general problems. This workshop aimed to foster continuing research in structure induction, and bring together different communities working on these problems, be it from a cognitive or a text processing perspective.

The workshop also hosted the PASCAL Unsupervised Grammar Induction Challenge, which aimed to foster continuing research in grammar induction and part-of-speech induction, while also opening up the problem to more ambitious settings, including a wider variety of languages, removing the reliance on gold standard parts-of-speech and, critically, providing a thorough evaluation.

Trevor Cohn, Phil Blunsom and João Graça, Workshop Chairs

Organizers:

Trevor Cohn, University of Sheffield
Phil Blunsom, University of Oxford
João Graça, Spoken Language Systems Lab, INESC-ID Lisboa

Program Committee:

Ben Taskar - University of Pennsylvania
Percy Liang - Stanford University
Andreas Vlachos - University of Cambridge
Chris Dyer - CMU
Mark Drezde - John Hopkins
Shai Cohen - Columbia University
Kuzman Ganchev - Google Inc.
André Martins - CMU/IST Portugal
Greg Druck - Yahoo
Ryan McDonald - Google Inc.
Nathan Schneider - CMU
Partha Talukdar - CMU
Dipanjan Das - CMU
Mark Steedman - University of Edinburgh
Luke Zettlemoyer - University of Washington
Roi Reichart - MIT
David Smith - University of Massachusetts
Ivan Titov - Saarland University
Alex Clarke - Royal Holloway University
Khalil Sima'an - University of Amsterdam
Stella Frank - University of Edinburgh

Invited Speakers:

Alex Clark - Royal Holloway University
Regina Barzilay - MIT
Noah Smith - CMU

Table of Contents

Conference Program

Thursday, June 7, 2012

9:00–10:00 Invited talk by Alex Clark

Session 1: Spotlight talks

Transferring Frames: Utilization of Linked Lexical Resources
Lars Borin, Markus Forsberg, Richard Johansson, Kristiina Muhonen, Tanja Purtonen and Kaarlo Voionmaa

Unsupervised Induction of Frame-Semantic Representations
Ashutosh Modi, Ivan Titov and Alexandre Klementiev

Capitalization Cues Improve Dependency Grammar Induction
Valentin I. Spitkovsky, Hiyan Alshawi and Daniel Jurafsky

10:30–11:00 Break

11:00–12:00 Invited talk by Regina Barzilay

Session 2: Spotlight talks

Toward Tree Substitution Grammars with Latent Annotations
Francis Ferraro, Benjamin Van Durme and Matt Post

Exploiting Partial Annotations with EM Training
Dirk Hovy and Eduard Hovy

Using Senses in HMM Word Alignment
Douwe Gelling and Trevor Cohn

Unsupervised Part of Speech Inference with Particle Filters
Gregory Dubbin and Phil Blunsom

Nudging the Envelope of Direct Transfer Methods for Multilingual Named Entity Recognition
Oscar Täckström

Thursday, June 7, 2012 (continued)

1:00–2:15 Lunch

2:15–3:15 Invited talk by Noah Smith

Session 3: PASCAL challenge and poster session

3:15–3:30 *The PASCAL Challenge on Grammar Induction*
Douwe Gelling, Trevor Cohn, Phil Blunsom and Joao Graca

3:30–4:00 Break and poster session

4:00–5:30 Poster session continues. Posters include the above spotlight papers and the following system descriptions

Two baselines for unsupervised dependency parsing
Anders Søgaard

Unsupervised Dependency Parsing using Reducibility and Fertility features
David Mareček and Zdeněk Žabokrtský

Induction of Linguistic Structure with Combinatory Categorial Grammars
Yonatan Bisk and Julia Hockenmaier

Turning the pipeline into a loop: Iterated unsupervised dependency parsing and PoS induction
Christos Christodoulopoulos, Sharon Goldwater and Mark Steedman

Hierarchical clustering of word class distributions
Grzegorz Chrupała

Combining the Sparsity and Unambiguity Biases for Grammar Induction
Kewei Tu