IWPT-09

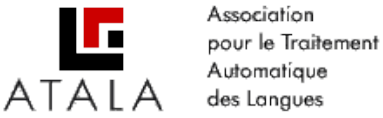
Proceedings of the 11th International Conference on Parsing Technologies

7-9 October 2009
Paris, France

Production and Manufacturing by
RAPHIDOC
1 rue Isabey
92500 Rueil-Malmaison
France

Sponsored by

ACL/SIGParse



ANR Passage

Photo credit: Fotolia

Preface

Welcome to the Eleventh International Conference on Parsing Technologies, IWPT'09, in the splendid city of Paris.

IWPT'09 continues the tradition of biennial conferences on parsing technology organized by SIGPARSE, the Special Interest Group on Parsing of the Association for Computational Linguistics (ACL). The first conference, in 1989, took place in Pittsburgh and Hidden Valley, Pennsylvania. Subsequently, IWPT conferences were held in Cancun (Mexico) in 1991; Tilburg (Netherlands) and Durbuy (Belgium) in 1993; Prague and Karlovy Vary (Czech Republic) in 1995; Boston/Cambridge (Massachusetts) in 1997; Trento (Italy) in 2000; Beijing (China) in 2001; Nancy (France) in 2003; Vancouver (Canada) in 2005; and Prague (Czech Republic) in 2007.

Over the years the IWPT Workshops have become the major forum for researchers in natural language parsing. They have lead to the publication of four books on parsing technologies; a fifth one about to be published.

Where the IWPT conferences from 1989 through 2003 were standalone conferences, the last two IWPTs were organised as co-satellite event of large conferences: IWPT 2005 was co-loated with the HLT-EMNLP conference in Vancouver, and IWPT 2007 with the main ACL conference in Prague. This worked well from a logistic point of view, thanks to the support from ACL, but it was felt to lead to somewhat less interesting events than in the past, sitting in the shadow of the larger conference and competing with other satellite events. It was therefore decided to return to the standalone format in 2009, with INRIA Rocquencourt and the University of Paris 7 volunteering to take charge of the organisation. We would like to thank Eric de la Clergerie, Laurence Danlos, Benoit Sagot and the support staff at INRIA and University of Paris 7 for their efforts to realize IWPT'09.

IWPT'09 is fortunate to have three very distinguished invited speakers: John Carroll from the university of Sussex, Mark Johnson from Brown University, and Joakim Nivre from the University of Uppsala.

I would like to thank all the programme committee members for their careful and timely work, especially those that took up extra rewiewing obligations at very short notice and those who participated in discussions on diverging reviews. Special thanks go to Eric de la Clergerie, the programme chair, for organising the reviewing, designing the workshop programme and producing the proceedings. The scientific programme includes 14 accepted full papers and 27 accepted short papers (the latter being an all-time high for IWPT), covering virtually all currently hot topics in parsing technology. Together with the three invited talks by top experts in parsing, these papers provide a fascinating picture of the state of the art in parsing natural language, that I hope you will enjoy and will find inspiring.

Harry Bunt
IWPT'09 General Chair

Organizers

General Chair:

Harry Bunt (Tilburg University, Netherlands)

Programme Chair:

Éric Villemonte de la Clergerie (INRIA, France)

Logistic Arrangements Chair:

Laurence Danlos (University Paris Diderot, France)

Programme Committee:

Philippe Blache (CNRS/Provence University, Aix-en-Provence, France)
Harry Bunt (TiCC, Tilburg University, Netherlands)
David Chiang(USC/ISI, Marina del Rey, USA)
John Carroll (University of Sussex, Brighton, UK)
Stephen Clark (University of Cambridge, UK)
Éric Villemonte de la Clergerie (INRIA, Rocquencourt, France) (chair)
Jason Eisner (Johns Hopkins University, Baltimore, USA)
James Henderson (University of Edinburgh,UK)
Julia Hockenmaier (University of Pennsylvania, Philadelphia, USA)
Aravind Joshi (University of Pennsylvania, Philadelphia, USA)
Ronald Kaplan (Xerox Palo Alto Research Center, USA)
Martin Kay (Xerox Palo Alto Research Center, USA)
Sadao Kurohashi (University of Kyoto, Japan)
Alon Lavie (Carnegie-Mellon University, Pittsburgh, USA)
Rob Malouf (San Diego State University, USA)
Yuji Matsumoto (Nara Institute of Science and Technology, Japan)
Paola Merlo (University of Geneva, Switzerland)
Bob Moore (Microsoft, Redmond, USA)
Mark-Jan Nederhof (University of St. Andrews, Scotland)
Joakim Nivre (University of Uppsala, Sweden)
Gertjan van Noord (University of Groningen, Netherlands)
Stephan Oepen (University of Oslo, Norway)
Stefan Riezler (Xerox Palo Alto Research Center, USA)
Giorgio Satta (University of Padua, Italy)
Kenji Sagae (Institute for Creative Technologies, Marina del Rey, USA)
Khalil Sima'an (University of Amsterdam, Netherlands)
Hozumi Tanaka (Chukyo University, Japan)
K. Vijay-Shanker (University of Delaware, USA)
Éric Wehrli (LATL, University of Geneva, Switzerland)
David Weir (University of Sussex, Brighton, UK)
Shuly Wintner (University of Haifa, Israel)
Dekai Wu (Hong Kong University of Science and Technology, China)

Additional Reviewers:

Tejaswini Deoskar (ILLC, University of Amsterdam, Netherlands)
Sylvain Schmitz (ENS Cachan, France)

Invited Speakers:

John Carroll (University of Sussex, Brighton, UK)
Mark Johnson (Brown University, USA)
Joakim Nivre (University of Uppsala, Sweden)

Panel Chair:

Josef van Genabith (DCU, Dublin, Ireland)

Table of Contents

Conference Program

Wednesday, October 7, 2009

9:00–9:15 Opening Remarks

9:15–10:15 Invited Talk by John Carroll

Coffee Break and Poster Display

10:45–11:15 *Parsing Algorithms based on Tree Automata*
Andreas Maletti and Giorgio Satta

11:15–11:45 *Weighted parsing of trees*
Mark-Jan Nederhof

11:45–12:20 **Short Paper Session I**

Automatic Adaptation of Annotation Standards for Dependency Parsing ? Using Projected Treebank as Source Corpus
Wenbin Jiang and Qun Liu

Learning Stochastic Bracketing Inversion Transduction Grammars with a Cubic Time Biparsing Algorithm
Markus Saers, Joakim Nivre and Dekai Wu

Empirical lower bounds on translation unit error rate for the full class of inversion transduction grammars
Anders Søgaard and Dekai Wu

Lunch

14:00–14:30 *Predictive Text Entry using Syntax and Semantics*
Sebastian Ganslandt, Jakob Jörwall and Pierre Nugues

14:30–15:00 *Parsing Formal Languages using Natural Language Parsing Techniques*
Jens Nilsson, Welf Löwe, Johan Hall and Joakim Nivre

15:00–16:00 **Short Paper Session II**

An Incremental Earley Parser for Simple Range Concatenation Grammar
Laura Kallmeyer and Wolfgang Maier

Deductive Parsing in Interaction Grammars
Joseph Le Roux

Synchronous Rewriting in Treebanks
Laura Kallmeyer, Wolfgang Maier and Giorgio Satta

Wednesday, October 7, 2009 (continued)

An Improved Oracle for Dependency Parsing with Online Reordering
Joakim Nivre, Marco Kuhlmann and Johan Hall

Two stage constraint based hybrid approach to free word order language dependency parsing
Akshar Bharati, Samar Husain, Dipti Misra and Rajeev Sangal

Coffee Break and Poster Display

16:35–17:00 **Short Paper Session III**

Analysis of Discourse Structure with Syntactic Dependencies and Data-Driven Shift-Reduce Parsing
Kenji Sagae

Evaluating Contribution of Deep Syntactic Information to Shallow Semantic Analysis
Sumire Uematsu and Jun'ichi Tsujii

17:00–17:30 *Weight Pushing and Binarization for Fixed-Grammar Parsing*
Matt Post and Daniel Gildea

17:30–18:00 *Co-Parsing with Competitive Models*
Lidia Khmylko, Kilian A. Foth and Wolfgang Menzel

Thursday, October 8, 2009

9:00–10:00 Invited Talk by Mark Johnson

Coffee Break and Poster Display

10:30–11:00 *Capturing Consistency between Intra-clause and Inter-clause Relations in Knowledge-rich Dependency and Case Structure Analysis*
Daisuke Kawahara and Sadao Kurohashi

11:00–11:30 *Constructing parse forests that include exactly the n-best PCFG trees*
Pierre Boullier, Alexis Nasr and Benoît Sagot

11:30–12:30 **Short Paper Session IV**

Hebrew Dependency Parsing: Initial Results
Yoav Goldberg and Michael Elhadad

Scalable Discriminative Parsing for German
Yannick Versley and Ines Rehbein

Improving generative statistical parsing with semi-supervised word clustering
Marie Candito and Benoît Crabbé

Thursday, October 8, 2009 (continued)

Application of feature propagation to dependency parsing
Kepa Bengoetxea and Koldo Gojenola

Guessing the Grammatical Function of a Non-Root F-Structure in LFG
Anton Bryl, Josef Van Genabith and Yvette Graham

Lunch

14:00–14:30 *Cross parser evaluation : a French Treebanks study*
Djamé Seddah, Marie Candito and Benoît Crabbé

14:30–15:00 *Transition-Based Parsing of the Chinese Treebank using a Global Discriminative Model*
Yue Zhang and Stephen Clark

15:00–15:25 **Short Paper Session V**

Grammar Error Detection with Best Approximated Parse
Jean-Philippe Prost

The effect of correcting grammatical errors on parse probabilities
Joachim Wagner and Jennifer Foster

Coffee Break and Poster Display

16:00–18:15 Panel: Statistical Parsing for Morphologically-rich Languages

Friday, October 9, 2009

9:00-10:00 Invited Talk by Joakim Nivre

10:00–10:30 *Effective Analysis of Causes and Inter-dependencies of Parsing Errors*
Tadayoshi Hara, Yusuke Miyao and Jun'ichi Tsujii

10:30–11:00 *Clustering Words by Syntactic Similarity improves Dependency Parsing of Predicate-argument Structures*
Kenji Sagae and Andrew S. Gordon

Coffee Break and Poster Display

11:30–12:30 **Short Paper Session VI**

The chunk as the period of the functions length and frequency of words on the syntagmatic axis
Jacques Vergne

Using a maximum entropy-based tagger to improve a very fast vine parser
Anders Søgaard and Jonas Kuhn

Friday, October 9, 2009 (continued)

HPSG Supertagging: A Sequence Labeling View
Yao-zhong Zhang, Takuya Matsuzaki and Jun'ichi Tsujii

Smoothing fine-grained PCFG lexicons
Tejaswini Deoskar, Mats Rooth and Khalil Sima'an

Wide-coverage parsing of speech transcripts
Jeroen Geertzen

Lunch

13:45–14:15 ACL/SIGParse Business Meeting

14:15–15:15 **Short Paper Session VII**

Interactive Predictive Parsing
Ricardo Sánchez-Sáez, Joan-Andreu Sánchez and José-Miguel Benedí

Using Treebanking Discriminants as Parse Disambiguation Features
Md. Faisal Mahbub Chowdhury, Yi Zhang and Valia Kordoni

Heuristic search in a cognitive model of human parsing
John Hale

Dependency Parsing with Energy-based Reinforcement Learning
Lidan Zhang and Kwok Ping Chan

A generative re-ranking model for dependency parsing
Federico Sangati, Willem Zuidema and Rens Bod

Coffee Break and Poster Display

15:45–16:15 *Dependency Constraints for Lexical Disambiguation*
Guillaume Bonfante, Bruno Guillaume and Mathieu Morey

16:15–16:45 *Parsing Directed Acyclic Graphs with Range Concatenation Grammars*
Pierre Boullier and Benoît Sagot

16:45–17:00 Closing Remarks

Invited Talks

Moving Parsing into the Real World: Noisy Text, Grammatical Representations and Applications

John Carroll
University of Sussex, Brighton, UK
J.A.Carroll@sussex.ac.uk

Much recent research in natural language parsing takes as input carefully crafted, edited text, often from newspapers. However, many real-world applications involve processing text which is not written carefully by a native speaker, is produced for an eventual audience of only one, and is in essence ephemeral. In this talk I will present a number of research and commercial applications of this type which I and collaborators are developing, in which we parse text as diverse as mobile phone text messages, non-native language learner essays, internet chat, and primary care medical notes. I will discuss the problems these types of text pose for a parser, and outline how we integrate information from parsing into applications.

Learning Rules with Adaptor Grammars

Mark Johnson
Brown University, USA
Mark_Johnson@Brown.edu

Nonparametric Bayesian methods are interesting because they may provide a way of learning the appropriate units of generalization (i.e., the "rules" of a grammar) as well as the generalization's probability or weight (i.e., the rule's probability). Adaptor Grammars are a framework for stating a variety of hierarchical nonparametric Bayesian models, where the units of generalization can be viewed as kinds of PCFG rules. This talk describes the mathematical and computational properties of Adaptor Grammars and linguistic applications such as word segmentation, syllabification and named entity recognition. The later part of the talk reviews MCMC inference and describes the MCMC algorithms we use to sample adaptor grammars.

Joint work with Sharon Goldwater and Tom Griffiths.

Discontinuous Dependency Parsing

Joakim Nivre
University of Uppsala, Sweden
joakim.nivre@lingfil.uu.se

There is a strong tendency in natural language syntax such that elements that have a direct syntactic relation are also adjacent in the surface realization of a sentence. Nevertheless, notable exceptions to this generalization exist in practically all languages and are especially common in languages with free or flexible word order. Syntactic theorists, on the one hand, have developed a variety of representational devices for dealing with these exceptions, including phonetically null elements, gap threading, and non-projective dependency trees. Syntactic parsers, on the other hand, use these devices very restrictively since they add to the complexity of an already daunting task. This is especially true of data-driven parsers, where discontinuity is often simply ignored. In this talk, I will review techniques for dealing with discontinuous structures in the framework of dependency parsing, focusing on parsing algorithms that build structures from non-adjacent elements and in particular transition-based algorithms that use online reordering.