IWPT 2020

The 16th International Conference on Parsing Technologies and IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies

Proceedings of the Conference

July 9, 2020

Organized by SIGPARSE
the ACL Special Interest Group on Natural Language Parsing

©2020 The Association for Computational Linguistics

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 978-1-952148-11-8

Preface

Welcome to the 16th International Conference on Parsing Technologies (IWPT 2020), which this year (for the first time since 2007) is co-located with the Annual Meeting of the Association for Computational Linguistics (ACL). The IWPT meeting series, hosted by the ACL Special Interest Group in Natural Language Parsing (SIGPARSE), has been held biennualy since its inaugual meeting in 1989 in Pittsburgh, PA (USA).

For 2020, the SIGPARSE steering group decided to try out something new, co-location with the main ACL meeting in the form of a reduced one-day IWPT programme. The main motivation for this move was to reduce fragmentation (and travel) and to increase IWPT visibility in the 'mainstream' ACL community (we already know that at least one of these goals was attained). At the same time, IWPT launches its own series of parsing shared tasks this year, which strengthens the experimental and applied perspective on parsing technologies in the conference programme.

The IWPT 2020 shared task focuses on the parsing of Enhanced Universal Dependencies (EUD) over 17 languages. This is the first time that graph-based representations of syntactic structures are evaluated on such a large scale, and we believe it will pave way for research on richer models and representations. The task attracted system submissions from ten teams from around the world and, thus, establishes a highly relevant point of comparison for this line of syntactic analysis. We are very grateful to everyone who contributed to this shared task, starting with the data providers who worked hard to meet our deadline. Thanks to the participant teams who worked tirelessly in a short time period to provide such a set of great and interesting systems!

Owing to the COVID-19 pandemic this year, the meeting will regrettably be held entirely virtual, where for IWPT we have adopted a mostly-asynchronous format: Accepted papers (of three different types: long, short, and shared task) will be presented through pre-recorded talks, which become available on-line for individual viewing before the actual conference day. On the original date of the conference, July 9, there will be a three-hour live session, scheduled so that the timing should be convenient (all things considered) for participants around the world: 14:00–17:00 UTC, which translates, for example, into a starting time at 7:00 in the morning at the US West Coast and wrapping up at 1:00 in the morning in Melbourne, Australia. The live sessions will be devoted exclusively to questions and answers, organized into five thematic sessions. Authors of papers associated with each session will be available to answer questions and disucss their work (possibly also among themselves).

There has been (and to some degree still is) much uncertainty about the format of ACL and IWPT this year, and in a sense we were positively surprised to receive a number of submissions comparable to recent IWPT instances. Out of 24 regular paper submissions, the programme committee accepted 14 for presentation at the conference. The IWPT 2020 programme is complemented by one invited talk, by Paola Merlo of the University of Geneva (to whom we are immensely grateful for honoring her commitment despite the mostly-asynchronous, virtual format) and by an overview paper and ten system descriptions from the IWPT 2020 shared task. We further gratefully acknowledge the work of authors and reviewers, as well as of the ACL workshop chairs, who had to try and shepherd our community through a difficult logistics process.

Copenhagen, Davis, Groningen, Kyoto, Oslo, Paris, Peking, Prague, and Tel Aviv

Gosse Bouma, Yuji Matsumoto, Stephan Oepen, Kenji Sagae, Djamé Seddah,
Weiwei Sun, Anders Søgaard, Reut Tsarfaty, and Dan Zeman

Organizers:

Kenji Sagae, University of California at Davis (General Chair)
Anders Søgaard, University of Copenhagen (Programme Co-Chair)
Weiwei Sun, Peking University (Programme Co-Chair)
Gosse Bouma, University of Groningen (Shared Task Co-Chair)
Djamé Seddah, University Paris-Sorbonne (Shared Task Co-Chair)
Dan Zeman, Charles University in Prague (Shared Task Co-Chair)
Stephan Oepen, University of Oslo (Publicity Chair)

Program Committee:

Željko Agić
Mark Anderson
Miguel Ballesteros
James Barry
Steven Bethard
Anders Björkelund
Gosse Bouma
Marie Candito
Xavier Carreras
John Carroll
Özlem Çetinoğlu
Grzegorz Chrupała
Ryan Cotterell
Miryam de Lhoneux
Mathieu Dehouck
Chris Dyer
Adam Ek
Jennifer Foster
Annemarie Friedrich
Yoav Goldberg
Carlos Gómez-Rodríguez
Han He
Johannes Heinecke
James Henderson
Daniel Hershcovich
Jenna Kanerva
Sandra Kübler
Marco Kuhlmann
Jonathan K. Kummerfeld
Xuezhe Ma
Gabriel Marzinotto
Yusuke Miyao
Mark-Jan Nederhof
Joakim Nivre
Stephan Oepen
Lilja Øvrelid
Barbara Plank
Ines Rehbein

Roi Reichart
Kenji Sagae
Giorgio Satta
Natalie Schluter
Djamé Seddah
Anders Søgaard
Weiwei Sun
Ivan Titov
Gertjan van Noord
Joachim Wagner
Rui Yan
Daniel Zeman
Yi Zhang
Yue Zhang

Invited Speaker:

Paola Merlo, University of Geneva

Table of Contents

Syntactic Parsing in Humans and Machines
Paola Merlo 1

Distilling Neural Networks for Greener and Faster Dependency Parsing
Mark Anderson and Carlos Gómez-Rodríguez 2

End-to-End Negation Resolution as Graph Parsing
Robin Kurtz, Stephan Oepen and Marco Kuhlmann 14

Integrating Graph-Based and Transition-Based Dependency Parsers in the Deep Contextualized Era
Agnieszka Falenska, Anders Björkelund and Jonas Kuhn 25

Semi-supervised Parsing with a Variational Autoencoding Parser
Xiao Zhang and Dan Goldwasser 40

Memory-bounded Neural Incremental Parsing for Psycholinguistic Prediction
Lifeng Jin and William Schuler 48

Obfuscation for Privacy-preserving Syntactic Parsing
Zhifeng Hu, Serhii Havrylov, Ivan Titov and Shay B. Cohen 62

Tensors over Semirings for Latent-Variable Weighted Logic Programs
Esma Balkir, Daniel Gildea and Shay B. Cohen 73

Advances in Using Grammars with Latent Annotations for Discontinuous Parsing
Kilian Gebhardt 91

Lexicalization of Probabilistic Linear Context-free Rewriting Systems
Richard Mörbitz and Thomas Ruprecht 98

Self-Training for Unsupervised Parsing with PRPN
Anhad Mohananey, Katharina Kann and Samuel R. Bowman 105

Span-Based LCFRS-2 Parsing
Miloš Stanojević and Mark Steedman 111

Analysis of the Penn Korean Universal Dependency Treebank (PKT-UD): Manual Revision to Build Robust Parsing Model in Korean
Tae Hwan Oh, Ji Yoon Han, Hyonsu Choe, Seokwon Park, Han He, Jinho D. Choi, Na-Rae Han, Jena D. Hwang and Hansaem Kim 122

Statistical Deep Parsing for Spanish Using Neural Networks
Luis Chiruzzo and Dina Wonsever 132

The Importance of Category Labels in Grammar Induction with Child-directed Utterances
Lifeng Jin and William Schuler 145

Overview of the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies
Gosse Bouma, Djamé Seddah and Daniel Zeman 151

Turku Enhanced Parser Pipeline: From Raw Text to Enhanced Graphs in the IWPT 2020 Shared Task
Jenna Kanerva, Filip Ginter and Sampo Pyysalo 162

Hybrid Enhanced Universal Dependencies Parsing
Johannes Heinecke 174

Adaptation of Multilingual Transformer Encoder for Robust Enhanced Universal Dependency Parsing
Han He and Jinho D. Choi 181

Efficient EUD Parsing
Mathieu Dehouck, Mark Anderson and Carlos Gómez-Rodríguez 192

Linear Neural Parsing and Hybrid Enhancement for Enhanced Universal Dependencies
Giuseppe Attardi, Daniele Sartiano and Maria Simi 206

Enhanced Universal Dependency Parsing with Second-Order Inference and Mixture of Training Data
Xinyu Wang, Yong Jiang and Kewei Tu 215

How Much of Enhanced UD Is Contained in UD?
Adam Ek and Jean-Philippe Bernardy 221

The ADAPT Enhanced Dependency Parser at the IWPT 2020 Shared Task
James Barry, Joachim Wagner and Jennifer Foster 227

Køpsala: Transition-Based Graph Parsing via Efficient Training and Effective Encoding
Daniel Hershcovich, Miryam de Lhoneux, Artur Kulmizev, Elham Pejhan and Joakim Nivre . . 236

RobertNLP at the IWPT 2020 Shared Task: Surprisingly Simple Enhanced UD Parsing for English
Stefan Grünewald and Annemarie Friedrich 245

Conference Program

July 9, 2020

14:00 UTC–14:15 UTC Session 1: Invited Talk Q&A

Syntactic Parsing in Humans and Machines
Paola Merlo

14:15 UTC–14:40 UTC Session 2: Regular Papers Q&A

Distilling Neural Networks for Greener and Faster Dependency Parsing
Mark Anderson and Carlos Gómez-Rodríguez

End-to-End Negation Resolution as Graph Parsing
Robin Kurtz, Stephan Oepen and Marco Kuhlmann

Integrating Graph-Based and Transition-Based Dependency Parsers in the Deep Contextualized Era
Agnieszka Falenska, Anders Björkelund and Jonas Kuhn

Semi-supervised Parsing with a Variational Autoencoding Parser
Xiao Zhang and Dan Goldwasser

14:40 UTC–15:00 UTC Session 3: Regular Papers Q&A

Memory-bounded Neural Incremental Parsing for Psycholinguistic Prediction
Lifeng Jin and William Schuler

Obfuscation for Privacy-preserving Syntactic Parsing
Zhifeng Hu, Serhii Havrylov, Ivan Titov and Shay B. Cohen

Tensors over Semirings for Latent-Variable Weighted Logic Programs
Esma Balkir, Daniel Gildea and Shay B. Cohen

July 9, 2020 (continued)

15:10 UTC–15:35 UTC Session 4: Regular Papers Q&A

Advances in Using Grammars with Latent Annotations for Discontinuous Parsing
Kilian Gebhardt

Lexicalization of Probabilistic Linear Context-free Rewriting Systems
Richard Mörbitz and Thomas Ruprecht

Self-Training for Unsupervised Parsing with PRPN
Anhad Mohananey, Katharina Kann and Samuel R. Bowman

Span-Based LCFRS-2 Parsing
Miloš Stanojević and Mark Steedman

15:35 UTC–16:00 UTC Session 5: Regular Papers Q&A

Analysis of the Penn Korean Universal Dependency Treebank (PKT-UD): Manual Revision to Build Robust Parsing Model in Korean
Tae Hwan Oh, Ji Yoon Han, Hyonsu Choe, Seokwon Park, Han He, Jinho D. Choi, Na-Rae Han, Jena D. Hwang and Hansaem Kim

Statistical Deep Parsing for Spanish Using Neural Networks
Luis Chiruzzo and Dina Wonsever

The Importance of Category Labels in Grammar Induction with Child-directed Utterances
Lifeng Jin and William Schuler

July 9, 2020 (continued)

16:10 UTC–17:00 UTC Session 6: Shared Task Q&A

Overview of the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies
Gosse Bouma, Djamé Seddah and Daniel Zeman

Turku Enhanced Parser Pipeline: From Raw Text to Enhanced Graphs in the IWPT 2020 Shared Task
Jenna Kanerva, Filip Ginter and Sampo Pyysalo

Hybrid Enhanced Universal Dependencies Parsing
Johannes Heinecke

Adaptation of Multilingual Transformer Encoder for Robust Enhanced Universal Dependency Parsing
Han He and Jinho D. Choi

Efficient EUD Parsing
Mathieu Dehouck, Mark Anderson and Carlos Gómez-Rodríguez

Linear Neural Parsing and Hybrid Enhancement for Enhanced Universal Dependencies
Giuseppe Attardi, Daniele Sartiano and Maria Simi

Enhanced Universal Dependency Parsing with Second-Order Inference and Mixture of Training Data
Xinyu Wang, Yong Jiang and Kewei Tu

How Much of Enhanced UD Is Contained in UD?
Adam Ek and Jean-Philippe Bernardy

The ADAPT Enhanced Dependency Parser at the IWPT 2020 Shared Task
James Barry, Joachim Wagner and Jennifer Foster

Køpsala: Transition-Based Graph Parsing via Efficient Training and Effective Encoding
Daniel Hershcovich, Miryam de Lhoneux, Artur Kulmizev, Elham Pejhan and Joakim Nivre

RobertNLP at the IWPT 2020 Shared Task: Surprisingly Simple Enhanced UD Parsing for English
Stefan Grünewald and Annemarie Friedrich