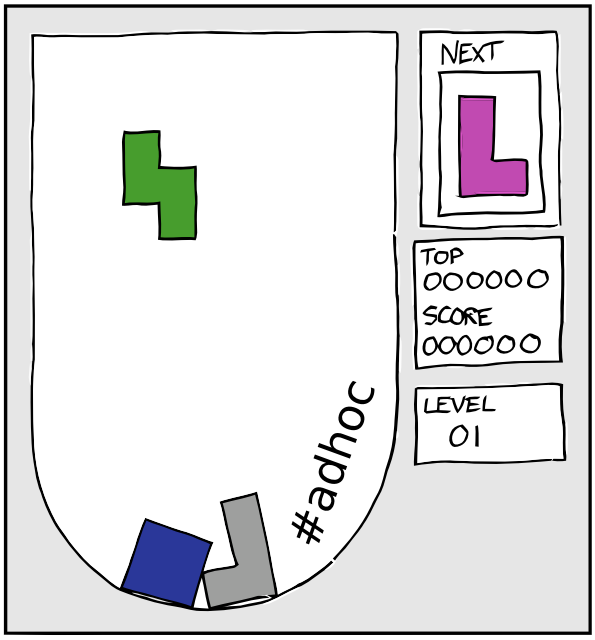
SPMRL-SANCL 2014

First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages (SPMRL-SANCL 2014)



Proceedings of the Workshop

©2014 The Authors

The papers in this volume are licensed by the authors under a *Creative Commons Attribution 4.0 International* License. Feel free to print your own copy.

The graphic on a titlepage is based on a work by Randall Munroe (XKCD 724: Hell, `xkcd.com/724/`) and is licensed under a *Creative Commons Attribution NonCommercial 2.5* License.

ISBN 978-1-941643-30-3

Proceedings of the First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-canonical Language (SPMRL-SANCL)

Yoav Goldberg, Yuval Marton, Yannick Versley, Özlem Çetinoğlu, Ines Rehbein, Joel Tetreault, Sandra Kübler, Djamé Seddah and Reut Tsarfaty (eds.)

Introduction

The papers in these proceedings were presented at the First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages (SPMRL-SANCL 2014), held in Seattle, USA, on October 18th, 2013, in conjunction with the 25th international Conference on Computational Linguistics (Coling 2014).

SPMRL-SANCL is endorsed by the ACL SIGPARSE interest group and provides a forum for research in parsing morphologically-rich languages and non-canonical language, with the goal of identifying cross-cutting issues in the annotation and parsing methodology, in the face of more flexible word order and/or higher word-form variation, or lexical sparseness and ad-hoc structures than English newspaper text.

SPMRL has also been host to discussions on realistic and appropriate evaluation methods that can be applied in the face of morphological and/or segmentation ambiguities; these discussions have culminated in the first shared task for parsing morphologically-rich languages, co-located with SPMRL 2013, and the second shared task for semi-supervised parsing of morphologically-rich languages, co-located with SPMRL 2014. The proceedings include nine contributions to the workshop as well as one system description from the shared task. The workshop included a keynote talk by Joakim Nivre (Uppsala).

We would like to thank all submitting authors for their contributions, the program committee for their fine work on reviewing the submissions, the participants of the shared task for their contributions and of course our invited speaker. For their precious help preparing the SPMRL 2014 Shared Task and for allowing their data to be part of it, we warmly thank and the Linguistic Data Consortium, the Knowledge Center for Processing Hebrew (MILA), the Ben Gurion University, Columbia University, Institute of Computer Science (Polish Academy of Sciences), Korea Advanced Institute of Science and Technology, University of the Basque Country, Uppsala University, University of Stuttgart, University of Szeged, University Paris Diderot (Paris 7), University of Marne La Vallée, and University of Tübingen. We gratefully acknowledge the contribution of Språkbanken and the University of Gothenburg for providing the PAROLE corpus and the help of Dr. Jungyeul Park and Prof. Key-Sun Cho for the KAIST annotated news corpus. Finally, we would also like to thank the ACL SIGPARSE interest group for their endorsement, for the support of INRIA's Alpage project, and everybody who participated in the workshop and contributed to the discussions.

Yoav Goldberg, Yuval Marton, Ines Rehbein, Yannick Versley, Özlem Çetinoğlu, Joel Tetrault
(Workshop organisers)

Sandra Kübler, Djamé Seddah and Reut Tsarfaty
(Shared Task organisers)

Workshop Organizers

Yoav Goldberg (Bar Ilan University, Israel)

Yuval Marton (Microsoft Corp., US)

Ines Rehbein (Potsdam University, Germany)

Yannick Versley (Heidelberg University, Germany)

Özlem Çetinoğlu (University of Stuttgart, Germany)

Joel Tetreault (Yahoo! Labs, US)

SANCL Special Track

Ines Rehbein (Potsdam University, Germany)

Djamé Seddah (Université Paris Sorbonne & INRIA's Alpage Project, France)

Özlem Çetinoğlu (University of Stuttgart, Germany)

Joel Tetreault (Yahoo! Labs, US)

SPMRL Shared Task

Sandra Kübler (Indiana University, US)

Djamé Seddah (Université Paris Sorbonne & INRIA's Alpage Project, France)

Reut Tsarfaty (Weizmann Institute of Science, Israel)

Invited Speaker:

Joakim Nivre (Uppsala University)

Program Committee:

Bernd Bohnet (University of Birmingham, UK)
Marie Candito (University of Paris 7, France)
Aoife Cahill (Educational Testing Service, US)
Jinho D. Choi (University of Massachusetts Amherst, US)
Grzegorz Chrupała (Tilburg University, Netherlands)
Markus Dickinson (Indiana University, US)
Stefanie Dipper (Ruhr-Universität Bochum, Germany)
Jacob Eisenstein (Georgia Institute of Technology, US)
Richárd Farkas (University of Szeged, Hungary)
Jennifer Foster (Dublin City University, Ireland)
Josef van Genabith (DFKI, Germany)
Koldo Gojenola (University of the Basque Country, Spain)
Spence Green (Stanford University, US)
Samar Husain (Potsdam University, Germany)
Sandra Kübler (Indiana University, US)
Joseph Le Roux (Université Paris-Nord, France)
John Lee (City University of Hong Kong, China)
Wolfgang Maier (University of Düsseldorf, Germany)
Takuya Matsuzaki (University of Tokyo, Japan)
David McClosky (IBM Research, US)
Detmar Meurers (University of Tübingen, Germany)
Joakim Nivre (Uppsala University, Sweden)
Kemal Oflazer (Carnegie Mellon University, Qatar)
Adam Przepiórkowski (ICS PAS, Poland)
Owen Rambow (Columbia University, US)
Kenji Sagae (University of Southern California, US)
Benoît Sagot (Inria, France)
Djamé Seddah (Univ. Paris Sorbonne, France)
Wolfgang Seeker (University of Stuttgart, Germany)
Anders Søgaard (University of Copenhagen, Denmark)
Reut Tsarfaty (Weizmann Institute of Science, Israel)
Lamia Tounsi (Dublin City University, Ireland)
Daniel Zeman (Charles University, Czech Republic)

Table of Contents

Parsing German: How Much Morphology Do We Need?
Wolfgang Maier, Sandra Kübler, Daniel Dakota and Daniel Whyatt 1

Joint Ensemble Model for POS Tagging and Dependency Parsing
Iliana Simova, Dimitar Vasilev, Alexander Popov, Kiril Simov and Petya Osenova 15

Improving the parsing of French coordination through annotation standards and targeted features
Assaf Urieli 26

Experiments with Easy-first nonprojective constituent parsing
Yannick Versley 39

Exploring Options for Fast Domain Adaptation of Dependency Parsers
Viktor Pekar, Juntao Yu, Mohab El-karef and Bernd Bohnet 54

Self-Training for Parsing Learner Text
Aoife Cahill, Binod Gyawali and James Bruno 66

The effect of disfluencies and learner errors on the parsing of spoken learner language
Andrew Caines and Paula Buttery 74

Initial Explorations in Two-phase Turkish Dependency Parsing by Incorporating Constituents
İlknur Durgar El-Kahlout, Ahmet Afşın Akın and Ertugrul Yılmaz 82

Experiments for Dependency Parsing of Greek
Prokopis Prokopidis and Haris Papageorgiou 90

Introducing the IMS-Wrocław-Szeged-CIS entry at the SPMRL 2014 Shared Task: Reranking and Morpho-syntax meet Unlabeled Data
Anders Björkelund, Özlem Çetinoğlu, Agnieszka Faleńska, Richárd Farkas, Thomas Mueller, Wolfgang Seeker and Zsolt Szántó 97

Introducing the SPMRL 2014 Shared Task on Parsing Morphologically-rich Languages
Djamé Seddah, Sandra Kübler and Reut Tsarfaty 103

Conference Program

Sunday August 24, 2014

9:00 Opening

9:05 Invited Talk: Universal Dependency Parsing (Joakim Nivre)

SPMRL

10:00 *Parsing German: How Much Morphology Do We Need?*
Wolfgang Maier, Sandra Kübler, Daniel Dakota and Daniel Whyatt

10:30 (coffee break)

11:00 *Joint Ensemble Model for POS Tagging and Dependency Parsing*
Iliana Simova, Dimitar Vasilev, Alexander Popov, Kiril Simov and Petya Osenova

11:30 *Improving the parsing of French coordination through annotation standards and targeted features*
Assaf Urieli

12:00 *Experiments with Easy-first nonprojective constituent parsing*
Yannick Versley

12:25 (lunch)

SANCL

14:00 *Exploring Options for Fast Domain Adaptation of Dependency Parsers*
Viktor Pekar, Juntao Yu, Mohab El-karef and Bernd Bohnet

14:30 *Self-Training for Parsing Learner Text*
Aoife Cahill, Binod Gyawali and James Bruno

14:50 *The effect of disfluencies and learner errors on the parsing of spoken learner language*
Andrew Caines and Paula Buttery

15:30 Poster session (SPMRL short papers and shared task)

Sunday August 24, 2014 (continued)

SPMRL short papers

Initial Explorations in Two-phase Turkish Dependency Parsing by Incorporating Constituents
İlknur Durgar El-Kahlout, Ahmet Afşın Akın and Ertugrul Yılmaz

Experiments for Dependency Parsing of Greek
Prokopis Prokopidis and Haris Papageorgiou

SANCL special session

I lack words and I don't know why: Solving elliptical structures in the syntactic annotation of private letters (Clara Pinto and Catarina Carvalheiro)

SPMRL shared task

Introducing the IMS-Wrocław-Szeged-CIS entry at the SPMRL 2014 Shared Task: Reranking and Morpho-syntax meet Unlabeled Data
Anders Björkelund, Özlem Çetinoğlu, Agnieszka Faleńska, Richárd Farkas, Thomas Mueller, Wolfgang Seeker and Zsolt Szántó

Introducing the SPMRL 2014 Shared Task on Parsing Morphologically-rich Languages
Djamé Seddah, Sandra Kübler and Reut Tsarfaty