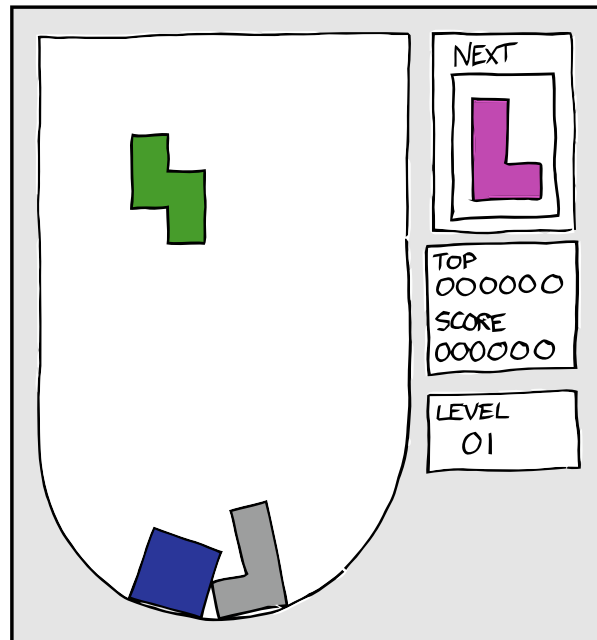


SPMRL-2013

# Fourth Workshop on Statistical Parsing of Morphologically Rich Languages



## Proceedings of the Workshop

18 October 2013  
Grand Hyatt Seattle  
Seattle, Washington, USA

©2013 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 978-1-937284-97-8

*The front-page picture is licensed by [xkcd.com](http://xkcd.com) under the terms of the Creative Commons Attribution-NonCommercial 2.5 License.*

*Original link: <http://xkcd.com/724/> ©[xkcd.com](http://xkcd.com)*

## Introduction

The papers in these proceedings were presented at the fourth Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL 2013), held in Seattle, USA, on October 18th, 2013, in conjunction with the Conference on Empirical Methods in Natural Language Processing (EMNLP 2013). SPMRL is endorsed by the ACL SIGPARSE and SIGLEX interest groups and provides a forum for research in parsing morphologically-rich languages, with the goal of identifying cross-cutting issues in the annotation and parsing methodology for such languages, which typically have more flexible word order and/or higher word-form variation than English.

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. The proceedings include nine contributions to the workshop as well as the system descriptions from the shared task. The workshop included keynote talks by Julia Hockenmaier from the University of Illinois at Urbana-Champaign and by Jinho Choi from IPSOFT (New York). 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 2013 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, University of Lisbon, Uppsala University, University of Stuttgart, University of Szeged and University Paris Diderot (Paris 7). Finally, we would also like to thank the ACL SIGPARSE and SIGLEX interest groups 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 and Yannick Versley (Workshop organisers)

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



**Workshop Organizers:**

Yoav Goldberg (Bar Ilan University, Israel)  
Yuval Marton (IBM Watson Research Center, US)  
Ines Rehbein (Potsdam University, Germany)  
Yannick Versley (Heidelberg University, Germany)

**Shared Task Organizers:**

Sandra Kübler (Indiana University, US)  
Djamé Seddah (Université Paris Sorbonne & INRIAs Alpage Project, France)  
Reut Tsarfaty (Weizmann Institute of Science, Israel)

**Program Committee:**

Mohammed Attia (Dublin City University, Ireland)  
Bernd Bohnet (University of Birmingham, UK)  
Marie Candito (University of Paris 7, France)  
Aoife Cahill (Educational Testing Service Inc., US)  
Özlem Cetinoglu (University of Stuttgart, Germany)  
Jinho D. Choi (IPSoft Inc., US)  
Grzegorz Chrupała (Tilburg University, Netherlands)  
Benoit Crabbé (University of Paris 7, France)  
Gülşen Cebiroglu Eryigit (Istanbul Technical University, Turkey)  
Michael Elhadad (Ben Gurion University, Israel)  
Richard Farkas (University of Szeged, Hungary)  
Jennifer Foster (Dublin City University, Ireland)  
Josef van Genabith (Dublin City University, Ireland)  
Koldo Gojenola (University of the Basque Country, Spain)  
Spence Green (Stanford University, US)  
Samar Husain (Potsdam University, Germany)  
Sandra Kübler (Indiana University, US)  
Jonas Kuhn (University of Stuttgart, Germany)  
Alberto Lavelli (FBK-irst, Italy)  
Joseph Le Roux (Université Paris-Nord, France)  
Wolfgang Maier (University of Düsseldorf, Germany)  
Takuya Matsuzaki (University of Tokyo, Japan)  
Joakim Nivre (Uppsala University, Sweden)  
Kemal Oflazer (Carnegie Mellon University, Qatar)  
Adam Przepiorkowski (ICS PAS, Poland)  
Owen Rambow (Columbia University, US)  
Kenji Sagae (University of Southern California, US)  
Benoit Sagot (Inria Rocquencourt, France)  
Djamé Seddah (Inria Rocquencourt, France)

Reut Tsarfaty (Weizmann Institute of Science, Israel)  
Lamia Tounsi (Dublin City University, Ireland)  
Daniel Zeman (Charles University, Czechia)

**Invited Speakers:**

Julia Hockenmaier, University of Illinois at Urbana-Champaign  
Jinho D. Choi, IPsoft Inc., New York

## Table of Contents

<i>Working with a small dataset - semi-supervised dependency parsing for Irish</i> Teresa Lynn, Jennifer Foster and Mark Dras .....	1
<i>Lithuanian Dependency Parsing with Rich Morphological Features</i> Jurgita Kapociute-Dzikiene, Joakim Nivre and Algis Krupavicius .....	12
<i>Parsing Croatian and Serbian by Using Croatian Dependency Treebanks</i> Željko Agić, Danijela Merkle and Daša Berović.....	22
<i>A Cross-Task Flexible Transition Model for Arabic Tokenization, Affix Detection, Affix Labeling, POS Tagging, and Dependency Parsing</i> Stephen Tratz .....	34
<i>The LIGM-Alpage architecture for the SPMRL 2013 Shared Task: Multiword Expression Analysis and Dependency Parsing</i> Matthieu Constant, Marie Candito and Djamé Seddah.....	46
<i>Exploring beam-based shift-reduce dependency parsing with DyALog: Results from the SPMRL 2013 shared task</i> Eric De La Clergerie.....	53
<i>Effective Morphological Feature Selection with MaltOptimizer at the SPMRL 2013 Shared Task</i> Miguel Ballesteros.....	63
<i>Exploiting the Contribution of Morphological Information to Parsing: the BASQUE TEAM system in the SPMRL'2013 Shared Task</i> Iakes Goenaga, Koldo Gojenola and Nerea Ezeiza .....	71
<i>The AI-KU System at the SPMRL 2013 Shared Task : Unsupervised Features for Dependency Parsing</i> Volkan Cirik and Hüsni Şensoy.....	78
<i>SPMRL'13 Shared Task System: The CADIM Arabic Dependency Parser</i> Yuval Marton, Nizar Habash, Owen Rambow and Sarah Alkhulani .....	86
<i>A Statistical Approach to Prediction of Empty Categories in Hindi Dependency Treebank</i> Puneeth Kukkadapu and Prashanth Mannem .....	91
<i>An Empirical Study on the Effect of Morphological and Lexical Features in Persian Dependency Parsing</i> Mojtaba Khallash, Ali Hadian and Behrouz Minaei-Bidgoli.....	97
<i>Constructing a Practical Constituent Parser from a Japanese Treebank with Function Labels</i> Takaaki Tanaka and Masaaki NAGATA .....	108
<i>Context Based Statistical Morphological Analyzer and its Effect on Hindi Dependency Parsing</i> Deepak Kumar Malladi and Prashanth Mannem .....	119

*Representation of Morphosyntactic Units and Coordination Structures in the Turkish Dependency Treebank*

Umut Sulubacak and Gülşen Eryiğit..... 129

*(Re)ranking Meets Morphosyntax: State-of-the-art Results from the SPMRL 2013 Shared Task*

Anders Björkelund, Ozlem Cetinoglu, Richárd Farkas, Thomas Mueller and Wolfgang Seeker 135

*Overview of the SPMRL 2013 Shared Task: A Cross-Framework Evaluation of Parsing Morphologically Rich Languages*

Djamé Seddah, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho D. Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola Gallettebeitia, Yoav Goldberg, Spence Green, Nizar Habash, Marco Kuhlmann, Wolfgang Maier, Yuval Marton, Joakim Nivre, Adam Przepiórkowski, Ryan Roth, Wolfgang Seeker, Yannick Versley, Veronika Vincze, Marcin Woliński and Alina Wróblewska 146



# Workshop Program

## Friday, October 18, 2013

7:30-9:00 Breakfast and Registration

**(9:00-10:30) SPMRL I**

9:00-9:10 Welcome

9:10-10:05 Invited talk: An HDP Model for Inducing CCGs by Julia Hockenmaier

10:05–10:30 *Working with a small dataset - semi-supervised dependency parsing for Irish*  
Teresa Lynn, Jennifer Foster and Mark Dras

**(10:30-11:00) Coffee**

**(11:00-12:30) SPMRL II**

11:00–11:25 *Lithuanian Dependency Parsing with Rich Morphological Features*  
Jurgita Kapociute-Dzikiene, Joakim Nivre and Algis Krupavicius

11:25–11:50 *Parsing Croatian and Serbian by Using Croatian Dependency Treebanks*  
Željko Agić, Danijela Merkle and Daša Berović

11:50–12:15 *A Cross-Task Flexible Transition Model for Arabic Tokenization, Affix Detection, Affix Labeling, POS Tagging, and Dependency Parsing*  
Stephen Tratz

12:15-12:30 Poster teasers

**Friday, October 18, 2013 (continued)**

**(12:30-14:00) Lunch break**

**(14:00-15:30) Shared task I**

- 14:00-14:10 Shared task intro by Djamé Seddah
- 14:10–14:20 *The LIGM-Alpage architecture for the SPMRL 2013 Shared Task: Multiword Expression Analysis and Dependency Parsing*  
Matthieu Constant, Marie Candito and Djamé Seddah
- 14:20–14:30 *Exploring beam-based shift-reduce dependency parsing with DyALog: Results from the SPMRL 2013 shared task*  
Eric De La Clergerie
- 14:30–14:55 *Effective Morphological Feature Selection with MaltOptimizer at the SPMRL 2013 Shared Task*  
Miguel Ballesteros
- 14:55–15:10 *Exploiting the Contribution of Morphological Information to Parsing: the BASQUE TEAM system in the SPRML'2013 Shared Task*  
Iakes Goenaga, Koldo Gojenola and Nerea Ezeiza
- 15:10–15:20 *The AI-KU System at the SPMRL 2013 Shared Task : Unsupervised Features for Dependency Parsing*  
Volkan Cirik and Hüsnü Şensoy
- 15:20–15:30 *SPMRL'13 Shared Task System: The CADIM Arabic Dependency Parser*  
Yuval Marton, Nizar Habash, Owen Rambow and Sarah Alkhulani
- 15:30-15:35 Q+A

**Friday, October 18, 2013 (continued)**

**(15:35-16:00) Coffee**

**(16:00-16:20) Poster session (+ posters from shared task participants)**

*A Statistical Approach to Prediction of Empty Categories in Hindi Dependency Treebank*  
Puneeth Kukkadapu and Prashanth Mannem

*An Empirical Study on the Effect of Morphological and Lexical Features in Persian Dependency Parsing*

Mojtaba Khallash, Ali Hadian and Behrouz Minaei-Bidgoli

*Constructing a Practical Constituent Parser from a Japanese Treebank with Function Labels*

Takaaki Tanaka and Masaaki NAGATA

*Context Based Statistical Morphological Analyzer and its Effect on Hindi Dependency Parsing*

Deepak Kumar Malladi and Prashanth Mannem

*Representation of Morphosyntactic Units and Coordination Structures in the Turkish Dependency Treebank*

Umut Sulubacak and Gülşen Eryiğit

**(16:20-18:00) Shared task II + panel**

16:20–16:45 *(Re)ranking Meets Morphosyntax: State-of-the-art Results from the SPMRL 2013 Shared Task*

Anders Björkelund, Ozlem Cetinoglu, Richárd Farkas, Thomas Mueller and Wolfgang Seeker

16:45-17:10 Invited talk: Dependency Parsing with Selectional Branching by Jinho D. Choi

17:10-18:00 Panel discussion

18:00-18:15 Closing Remarks

*Overview of the SPMRL 2013 Shared Task: A Cross-Framework Evaluation of Parsing Morphologically Rich Languages*

Djamé Seddah, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho D. Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola Gallettebeitia, Yoav Goldberg, Spence Green, Nizar Habash, Marco Kuhlmann, Wolfgang Maier, Yuval Marton, Joakim Nivre, Adam Przepiórkowski, Ryan Roth, Wolfgang Seeker, Yannick Versley, Veronika Vincze, Marcin Woliński and Alina Wróblewska

