

NAACL HLT 2016

**Workshop on Discontinuous Structures  
in Natural Language Processing (DiscoNLP)**

**Proceedings of the Workshop**

June 17, 2016  
San Diego, California, USA

©2016 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](mailto:acl@aclweb.org)

ISBN 978-1-941643-85-3

## Introduction

This volume presents the papers presented at the Workshop on Discontinuous Structures in Natural Language Processing, held in San Diego, California on June 17, 2016 during the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

The modeling of certain structures in natural language requires a mechanism for discontinuity, in the sense that we must account for two or more parts of the structure that are not adjacent. This is true across many languages and on different description levels. For instance, on the lexical level, this concerns discontinuous morphological phenomena such as transfixation (templatic morphology), as well as phrasal verbs, and non-contiguous multiword expressions. On the syntactic level, discontinuity is caused by phenomena such as extraposition and topicalization, or argument scrambling. Morphologically rich languages (MRLs) are particularly likely to exhibit such phenomena. Other examples include disfluency and anaphora/coreference resolution with discontinuous antecedents; modeling in both of the latter areas requires an extended domain of locality. On a higher level, discontinuity is a relevant factor in machine translation, as well as in complex question answering and in topic structure modeling. Discontinuity has been studied intensively in a range of different areas, including but not limited to grammar development, syntactic and semantic parsing, morphological analysis, machine translation, anaphora resolution, discourse modeling, automatic summarization and complex question answering.

Nevertheless, the treatment of discontinuous structures remains a challenge, because on the one hand, recovering of non-local information is generally associated with a high computational cost, and on the other hand, discontinuities are inherently a low-frequency phenomenon, which means that statistical approaches have a tendency to analyze them incorrectly as more frequent local phenomena. Additionally, it is not always clear if and how NLP tasks can benefit from knowing about discontinuity, that is, why one should care, particularly considering the given computational cost. The goal of this workshop is to bring together researchers from the different areas to give them a forum to exchange ideas and problem solutions, to create synergy effects, and to enable more powerful solutions. This encompasses not only linguistic analyses and work on analyzing or recovering the corresponding structures, such as, e.g., in non-projective dependency parsing, but also studies on "use cases", which show how information about discontinuity can be used to enhance NLP tasks. We think that given the broad program we have put together, this goal has been more than fulfilled.

Thanks to all authors who have contributed their work! Out of ten submissions, seven were selected for presentation. We would also like to extend our gratitude the program committee, who have dedicated their time and effort in order to make this workshop a high-quality event.

See you in San Diego!

Wolfgang Maier, Sandra Kübler, and Constantin Orăsan



**Organizers:**

Wolfgang Maier, University of Düsseldorf (Germany)  
Sandra Kübler, Indiana University (USA)  
Constantin Orăsan, University of Wolverhampton (UK)

**Program Committee:**

Anne Abeillé, University Paris 7 (France)  
Krasimir Angelov, University of Gothenburg (Sweden)  
Marianna Apidianaki, LIMSIS (France)  
Eric de la Clergerie, INRIA (France)  
Andreas van Cranenburgh, Royal Netherlands Academy for Arts and Sciences (The Netherlands)  
Joachim Daiber, University of Amsterdam (The Netherlands)  
Carlos Gómez Rodríguez, University of A Coruña (Spain)  
Eva Hasler, University of Cambridge (UK)  
Mijaíl Kabadjov, University of Essex (UK)  
Sylvain Kahane, University Paris 10 (France)  
Laura Kallmeyer, University of Düsseldorf (Germany)  
Philipp Koehn, University of Edinburgh (UK)  
Johannes Leveling, Elsevier (The Netherlands)  
Timm Lichte, University of Düsseldorf (Germany)  
Peter Ljunglöf, University of Gothenburg (Sweden)  
Georgiana Marsic, University of Wolverhampton (UK)  
Detmar Meurers, University of Tübingen (Germany)  
Jean-Luc Minel, Université Paris Ouest Nanterre La Défense (France)  
Sara Moze, University of Wolverhampton (UK)  
Philippe Muller, University of Toulouse/IRIT (France)  
Preslav Nakov, Qatar Computing Research Institute (Qatar)  
Mark-Jan Nederhof, University of St. Andrews (UK)  
Yannick Parmentier, University of Orléans (France)  
Ted Pedersen, University of Minnesota (USA)  
Irene Renau, Pontificia Universidad Católica de Valparaíso (Chile)  
Lonneke van der Plas, University of Malta (Malta)  
Natalie Schluter, University of Copenhagen (Denmark)  
Djamé Seddah, University Paris 4 (France)  
Khalil Sima'an, University of Amsterdam (The Netherlands)  
Yannick Versley, University of Heidelberg (Germany)  
Suzan Veberne, University of Nijmegen (The Netherlands)  
Andy Way, Dublin City University (Ireland)

**Invited Speaker:**

David Chiang, University of Notre Dame (USA)



## Table of Contents

<i>An LFG Account of Discontinuous Nominal Expressions</i>	
Liselotte Snijders .....	1
<i>Non-projectivity and valency</i>	
Zdenka Uresova, Eva Fucikova and Jan Hajic .....	12
<i>Machine Translation of Non-Contiguous Multiword Units</i>	
Anabela Barreiro and Fernando Batista .....	22
<i>Discontinuous VP in Bulgarian</i>	
Elisaveta Balabanova .....	31
<i>Discontinuous Genitives in Hindi/Urdu</i>	
Sebastian Sulger .....	37
<i>Discontinuous parsing with continuous trees</i>	
Wolfgang Maier and Timm Lichte .....	47
<i>Discontinuity Revisited: A Minimalist Approach to Pseudoprojective Constituent Parsing</i>	
Yannick Versley .....	58





# Workshop Program

**Friday, June 17, 2016**

- 9:30–10:00    *An LFG Account of Discontinuous Nominal Expressions*  
Liselotte Snijders
- 10:00–10:30    *Non-projectivity and valency*  
Zdenka Uresova, Eva Fucikova and Jan Hajic
- 10:30–11:00    *Coffee break***
- 11:00–12:15    *Invited Talk: Finite automata for free word order languages*  
David Chiang
- 12:15–12:45    *Machine Translation of Non-Contiguous Multiword Units*  
Anabela Barreiro and Fernando Batista
- 12:45–14:30    *Lunch break***
- 2:30–3:00    *Discontinuous VP in Bulgarian*  
Elisaveta Balabanova
- 3:00–3:30    *Discontinuous Genitives in Hindi/Urdu*  
Sebastian Sulger
- 4:00–4:30    *Discontinuous parsing with continuous trees*  
Wolfgang Maier and Timm Lichte
- 4:30–5:00    *Discontinuity Revisited: A Minimalist Approach to Pseudoprojective Constituent Parsing*  
Yannick Versley
- 5:00–5:45    *Panel discussion*

