

ComAComA 2014

**Proceedings of the First Workshop on
Computational Approaches to Compound Analysis**

Held at the 25th International Conference on Computational Linguistics (COLING 2014)

Editors

Ben Verhoeven
Walter Daelemans
Menno van Zaanen
Gerhard van Huyssteen

ISBN: 978-1-873769-43-0

August 24, 2014
Dublin, Ireland

All works in this volume are licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

Publishers

Dublin City University (DCU)
Glasnevin, Dublin 9, Ireland

Association for Computational Linguistics (ACL)
Stroudsburg, PA, USA
<http://aclweb.org/anthology/>

ISBN: 978-1-873769-43-0

Introduction

The ComAComA workshop is an interdisciplinary platform for researchers working on compound processing in different languages, to present recent and ongoing work.

The workshop has several related aims. Firstly, it brings together researchers from different backgrounds (e.g., computational linguistics, linguistics, neurolinguistics, psycholinguistics, language technology) to discuss and evaluate compound processing each from their own point of view. Secondly, based on the interaction between the participants, the workshop provides an overview of existing and desired resources for future research in this area. Finally, we expect that the interdisciplinary approach of the workshop will result in better methodologies to evaluate compound processing systems from different perspectives.

Given the high productivity of compounding in a wide range of languages, compound processing is an interesting subject in linguistics, computational linguistics, and other applied disciplines. For example, for many language technology applications, compound processing remains a challenge (both morphologically and semantically), since novel compounds are created and interpreted on the fly. In order to deal with this productivity, systems that can analyse new compound forms and their meanings need to be developed. From an interdisciplinary perspective, we also need to better understand the process of compounding (as a cognitive process), in order to model its complexity.

Workshop Organizers

Ben Verhoeven, University of Antwerp, Belgium

Walter Daelemans, University of Antwerp, Belgium

Menno van Zaanen, Tilburg University, The Netherlands

Gerhard van Huyssteen, North-West University, South Africa

Program Committee

Preslav Nakov, Qatar Computing Research Institute

Iris Hendrickx, Radboud University Nijmegen

Lonneke Van der Plas, University of Stuttgart

Helmut Schmid, Ludwig Maximilian University Munich

Roald Eiselen, North-West University

Pavol Štekauer, P.J. Safarik University

Diarmuid Ó Séaghdha, University of Cambridge

Rochelle Lieber, University of New Hampshire

Tony Veale, University College Dublin

Pius ten Hacken, University of Innsbruck

Anneke Neijt, Radboud University Nijmegen

Andrea Krott, University of Birmingham

Emmanuel Keuleers, Ghent University

Stan Szpakowicz, University of Ottawa

Invited Speakers

Andrea Krott, University of Birmingham

Diarmuid Ó Séaghdha, University of Cambridge

Table of Contents

| | |
|---|----|
| <i>Modelling Regular Subcategorization Changes in German Particle Verbs</i> Stefan Bott and Sabine Schulte im Walde | 1 |
| <i>Splitting of Compound Terms in non-Prototypical Compounding Languages</i> Elizaveta Clouet and Béatrice Daille | 11 |
| <i>Automatic Compound Processing: Compound Splitting and Semantic Analysis for Afrikaans and Dutch</i> Ben Verhoeven, Menno van Zaanen, Walter Daelemans and Gerhard Van Huyssteen | 20 |
| <i>A Taxonomy for Afrikaans and Dutch Compounds</i> Gerhard Van Huyssteen and Ben Verhoeven | 31 |
| <i>Electrophysiological correlates of noun-noun compound processing by non-native speakers of English</i> Cecile DeCat, Harald Baayen and Ekaterini Klepousniotou | 41 |
| <i>A Comparative Study of Different Classification Methods for the Identification of Brazilian Portuguese Multiword Expressions</i> Alexsandro Fonseca and Fatiha Sadat | 53 |
| <i>Wordsyou dont know: Evaluation of lexicon-based decompounding with unknown handling</i> Karolina Owczarzak, Ferdinand de Haan, George Krupka and Don Hindle | 63 |
| <i>Multiword noun compound bracketing using Wikipedia</i> Caroline Barriere and Pierre André Ménard | 72 |
| <i>Distinguishing Degrees of Compositionality in Compound Splitting for Statistical Machine Translation</i> Marion Weller, Fabienne Cap, Stefan Müller, Sabine Schulte im Walde and Alexander Fraser | 81 |

