

CORBON 2017

**Second Workshop on
Coreference Resolution beyond OntoNotes**

Proceedings of the Workshop

EACL 2017 Workshop
April 4, 2017
Valencia, Spain

©2017 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 978-1-945626-46-3

Introduction

Many NLP researchers, especially those not working in the area of discourse processing, tend to equate coreference resolution with the sort of coreference that people did in MUC, ACE, and OntoNotes, having the impression that coreference is a well-worn task owing in part to the large number of papers reporting results on the MUC/ACE/OntoNotes corpora. This is an unfortunate misconception: the previous SemEval 2010 and CoNLL 2012 shared tasks on coreference resolution have largely focused on entity coreference, which constitutes only one of the many kinds of coreference relations that were discussed in theoretical and computational linguistics in the past few decades. In fact, by focusing on entity coreference resolution, NLP researchers have only scratched the surface of the wealth of interesting problems in coreference resolution.

The decision to focus on entity coreference resolution was initially made by information extraction (IE) researchers when coreference was selected as one of the tasks in the MUC-6 coreference in 1995. Many interesting kinds of coreference relations, such as bridging and reference to abstract entities, were left out not because they were not important, but because “it was felt that the menu was simply too ambitious”. It turned out that this decision had an important consequence: the progress made in coreference research in the past two decades was largely driven by the availability of coreference-annotated corpora such as MUC, ACE, and OntoNotes, where entity coreference was the focus.

Being aware of other fora gathering coreference-related papers (such as LAW, DiscoMT or EVENTS), in 2016 we started a new workshop on the single topic of knowledge-oriented coreference resolution under the name of *Coreference Resolution Beyond OntoNotes* (CORBON) that would bring together researchers who were interested in under-investigated coreference phenomena, willing to contribute both theoretical and applied computational work on coreference resolution, especially for languages other than English, less-researched forms of coreference and new applications of coreference resolution.

The success of the first edition of the workshop (held in conjunction with NAACL HLT 2016 in San Diego, USA) and our intention to verify the role of the Europe-based researchers in the field encouraged us to organise the second edition of the workshop in conjunction with EACL 2017 in Valencia, Spain. Our call attracted 12 submissions (nine from European research institutions and three from India). We are pleased to see that the submissions covered not only a variety of less-studied languages in the coreference community (e.g., Basque, French, German, Polish, Portuguese, Russian, and Tamil) but also many under-investigated topics in coreference resolution (e.g., feature representation, the use of semantics and deep syntax for coreference resolution, difficult cases of anaphora, and the use of coreference chains in high-level natural language applications). Each submission was rigorously reviewed by three to five programme committee members. We would like to thank the 29 programme committee members for their hard work. Based on their recommendations, we accepted six papers.

We are grateful to Massimo Poesio for accepting our invitation to be this year’s invited speaker. Massimo will give us an overview of his new project on developing better games and techniques to collect and analyse data about anaphora and using them to train probabilistic resolvers.

To further enrich the workshop participants’ experience, we included in this year’s programme a panel discussion on the interplay of referential and discourse relations in text. We thank Ruslan Mitkov, Anna Nedoluzhko, Massimo Poesio, and Arndt Riester for agreeing to serve as panelists. We are excited about this new addition to the programme.

To promote work on coreference resolution in low-resource languages, we included in our call for papers a shared task on projection-based coreference resolution. The goal was to perform German and Russian coreference resolution by projecting automatically generated coreference chains from English to these languages via a parallel corpus. In particular, the participants were not allowed to employ any knowledge of these languages or use any German and Russian coreference-annotated data to train resolvers in these languages. To our knowledge, this is the first shared task on projection-based coreference resolution. We are indebted to our shared task coordinator, Yulia Grishina, who capably handled all aspects of the shared task, ranging from data preparation to the scoring of the participating systems. Papers related to the shared task, including Yulia’s overview paper and the participating team’s system description paper, are included in the proceedings and will be presented during the workshop.

Finally, we would like to thank the workshop participants for joining in. We look forward to an exciting workshop in Valencia.

— Maciej Ogrodniczuk and Vincent Ng

Organizing Committee and Proceedings Editors:

Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences
Vincent Ng, University of Texas at Dallas

Shared Task Coordinator:

Yulia Grishina, University of Potsdam

Programme Committee:

Anders Björkelund, University of Stuttgart
Antonio Branco, University of Lisbon
Chen Chen, Google
Dan Cristea, A. I. Cuza University of Iași
Pascal Denis, MAGNET, INRIA Lille Nord-Europe
Sobha Lalitha Devi, AU-KBC Research Center, Anna University, Chennai
Yulia Grishina, University of Potsdam
Lars Hellan, Norwegian University of Science and Technology
Veronique Hoste, Ghent University
Yufang Hou, IBM
Ryu Iida, National Institute of Information and Communications Technology (NICT)
Ekaterina Lapshinova-Koltunski, Saarland University
Emmanuel Lassalle, Citadel, UK
Chen Li, Microsoft
Sebastian Martschat, Heidelberg University
Ruslan Mitkov, University of Wolverhampton
Costanza Navarretta, University of Copenhagen
Anna Nedoluzhko, Charles University in Prague
Michal Novak, Charles University in Prague
Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences
Constantin Orasan, University of Wolverhampton
Massimo Poesio, University of Essex
Manfred Stede, University of Potsdam
Veselin Stoyanov, Facebook
Yannick Versley, Heidelberg University
Rob Voigt, Stanford University
Sam Wiseman, Harvard University
Amir Zeldes, Georgetown University
Heike Zinsmeister, University of Hamburg

Invited Speaker:

Massimo Poesio, University of Essex

Panelists:

Ruslan Mitkov, University of Wolverhampton
Anna Nedoluzhko, Charles University in Prague
Massimo Poesio, University of Essex
Arndt Riestler, University of Stuttgart

Table of Contents

<i>Use Generalized Representations, But Do Not Forget Surface Features</i> Nafise Sadat Moosavi and Michael Strube	1
<i>Enriching Basque Coreference Resolution System using Semantic Knowledge sources</i> Ander Soraluze, Olatz Arregi, Xabier Arregi and Arantza Díaz de Ilarraza	8
<i>Improving Polish Mention Detection with Valency Dictionary</i> Maciej Ogrodniczuk and Bartłomiej Nitoń	17
<i>A Google-Proof Collection of French Winograd Schemas</i> Pascal Amsili and Olga Seminck	24
<i>Using Coreference Links to Improve Spanish-to-English Machine Translation</i> Lesly Miculicich Werlen and Andrei Popescu-Belis	30
<i>Multi-source annotation projection of coreference chains: assessing strategies and testing opportunities</i> Yulia Grishina and Manfred Stede	41
<i>CORBON 2017 Shared Task: Projection-Based Coreference Resolution</i> Yulia Grishina	51
<i>Projection-based Coreference Resolution Using Deep Syntax</i> Michal Novák, Anna Nedoluzhko and Zdeněk Žabokrtský	56

Workshop Program: April 4, 2017

09:30–11:00 Session 1: Invited Talk, Feature Representations

09:30–09:40 *Introduction*

Maciej Ogrodniczuk and Vincent Ng

09:40–10:40 Invited talk: *Exploring Anaphoric Ambiguity Using Games-With-a-Purpose: The Dali Project*

Massimo Poesio

10:40–11:00 *Use Generalized Representations, But Do Not Forget Surface Features*

Nafise Sadat Moosavi, Michael Strube

11:00–11:30 Coffee Break

11:30–13:00 Session 2: Coreference in Under-Investigated Languages

11:30–12:00 *Enriching Basque Coreference Resolution System using Semantic Knowledge Sources*

Ander Soraluze, Olatz Arregi, Xabier Arregi, Arantza Díaz de Ilarraza

12:00–12:30 *Improving Polish Mention Detection with Valency Dictionary*

Maciej Ogrodniczuk, Bartłomiej Nitoń

12:30–13:00 *A Google-Proof Collection of French Winograd Schemas*

Pascal Amsili, Olga Seminck

13:00–14:30 Lunch Break

14:30–16:00 Session 3: Panel Discussion, Coreference for NLP Applications

14:30–15:30 Panel discussion: *Referential vs. discourse relations*

Ruslan Mitkov, Anna Nedoluzhko, Massimo Poesio, Arndt Riester

15:30–16:00 *Using Coreference Links to Improve Spanish-to-English Machine Translation*

Lesly Miculicich Werlen, Andrei Popescu-Belis

16:00–16:30 Coffee Break

16:30–18:00 Session 4: Projection-Based Coreference Resolution

16:30–17:00 *Multi-source Annotation Projection of Coreference Chains: Assessing Strategies and Testing Opportunities*

Yulia Grishina, Manfred Stede

17:00–17:30 *CORBON 2017 Shared Task: Projection-Based Coreference Resolution*

Yulia Grishina

17:30–18:00 *Projection-based Coreference Resolution Using Deep Syntax*

Michal Novák, Anna Nedoluzhko, Zdeněk Žabokrtský

