

State-of-the-art NLP Approaches to Coreference Resolution: Theory and Practical Recipes

Simone Paolo Ponzetto

Seminar für Computerlinguistik

University of Heidelberg

ponzetto@cl.uni-heidelberg.de

Massimo Poesio

DISI

University of Trento

massimo.poesio@unitn.it

1 Introduction

The identification of different nominal phrases in a discourse as used to refer to the same (discourse) entity is essential for achieving robust natural language understanding (NLU). The importance of this task is directly amplified by the field of Natural Language Processing (NLP) currently moving towards high-level linguistic tasks requiring NLU capabilities such as e.g. recognizing textual entailment. This tutorial aims at providing the NLP community with a gentle introduction to the task of coreference resolution from both a theoretical and an application-oriented perspective. Its main purposes are: (1) to introduce a general audience of NLP researchers to the core ideas underlying state-of-the-art computational models of coreference; (2) to provide that same audience with an overview of NLP applications which can benefit from coreference information.

2 Content Overview

1. Introduction to machine learning approaches to coreference resolution. We start by focusing on machine learning based approaches developed in the seminal works from Soon et al. (2001) and Ng & Cardie (2002). We then analyze the main limitations of these approaches, i.e. their clustering of mentions from a local pairwise classification of nominal phrases in text. We finally move on to present more complex models which attempt to model coreference as a global discourse phenomenon (Yang et al., 2003; Luo et al., 2004; Daumé III & Marcu, 2005, inter alia).

2. Lexical and encyclopedic knowledge for coreference resolution. Resolving anaphors to their correct antecedents requires in many cases lexical and encyclopedic knowledge. We accordingly introduce approaches which attempt to include semantic information into the coreference models from a variety of knowledge sources,

e.g. WordNet (Harabagiu et al., 2001), Wikipedia (Ponzetto & Strube, 2006) and automatically harvested patterns (Poesio et al., 2002; Markert & Nissim, 2005; Yang & Su, 2007).

3. Applications and future directions. We present an overview of NLP applications which have been shown to profit from coreference information, e.g. question answering and summarization. We conclude with remarks on future work directions. These include: a) bringing together approaches to coreference using semantic information with global discourse modeling techniques; b) exploring novel application scenarios which could potentially benefit from coreference resolution, e.g. relation extraction and extracting events and event chains from text.

References

- Daumé III, H. & D. Marcu (2005). A large-scale exploration of effective global features for a joint entity detection and tracking model. In *Proc. HLT-EMNLP '05*, pp. 97–104.
- Harabagiu, S. M., R. C. Bunescu & S. J. Maiorano (2001). Text and knowledge mining for coreference resolution. In *Proc. of NAACL-01*, pp. 55–62.
- Luo, X., A. Ittycheriah, H. Jing, N. Kambhatla & S. Roukos (2004). A mention-synchronous coreference resolution algorithm based on the Bell Tree. In *Proc. of ACL-04*, pp. 136–143.
- Markert, K. & M. Nissim (2005). Comparing knowledge sources for nominal anaphora resolution. *Computational Linguistics*, 31(3):367–401.
- Ng, V. & C. Cardie (2002). Improving machine learning approaches to coreference resolution. In *Proc. of ACL-02*, pp. 104–111.
- Poesio, M., T. Ishikawa, S. Schulte im Walde & R. Vieira (2002). Acquiring lexical knowledge for anaphora resolution. In *Proc. of LREC '02*, pp. 1220–1225.
- Ponzetto, S. P. & M. Strube (2006). Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution. In *Proc. of HLT-NAACL-06*, pp. 192–199.
- Soon, W. M., H. T. Ng & D. C. Y. Lim (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544.
- Yang, X. & J. Su (2007). Coreference resolution using semantic relatedness information from automatically discovered patterns. In *Proc. of ACL-07*, pp. 528–535.
- Yang, X., G. Zhou, J. Su & C. L. Tan (2003). Coreference resolution using competition learning approach. In *Proc. of ACL-03*, pp. 176–183.