

**Proceedings of the
Workshop on
Reference Resolution and its Applications**

Held in cooperation with ACL-2004

**25–26 July 2004
Barcelona, Spain**

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
73 Landmark Center
East Stroudsburg, PA 18301
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

PROGRAM COMMITTEE:

Sanda Harabagiu, University of Texas at Dallas, Chair

David Farwell, Computing Research Laboratory, New Mexico State University, Chair

Donna Byron, The Ohio State University

Nicoletta Calzolari, Istituto di Linguistica Computazionale-CNR, Pisa

Claire Cardie, Cornell University

Dan Cristea, University of Iasi

Rodolfo Delmonte, Ca' Foscari University

Antonio Ferrández, University of Alicante

Udo Hahn, University of Freiburg

Graeme Hirst, University of Toronto

Nancy Ide, Vassar College

Steve Maiorano, University of Sheffield

Ruslan Mitkov, University of Wolverhampton

Massimo Poesio, University of Essex

Carol Van Ess-Dykema, DoD

Renata Vieira, PIPCA Unisinos, Brazil

Bonnie Webber, University of Edinburgh

CONFERENCE WEBSITE:

<http://www.hlt.utdallas.edu/~cristina/workshop>

PREFACE

With the advent of the Internet and the predominant representation of on-line information in textual form, the quality and robustness of natural language understanding systems becomes critical. A major challenge in building robust systems that interpret text, extract information, summarize and answer questions is the need to identify the entities referred to by pronouns or other referring expressions. Such textual references are a pervasive phenomenon that interact with virtually every aspect of semantic and discourse analysis. While reference resolution algorithms have been applicable to processing discourse and dialogue for some time, more recently they have become important tools for text-based Information Extraction systems, Question Answering systems and Summarization systems.

The objective of this ACL workshop on Reference Resolution and its Applications has been to bring together researchers whose work focuses on the resolution of reference in natural language and to provide a forum for presenting and discussing the results of recent innovative work. The contributions in this proceeding cover a range of topics including:

- new or extended algorithms for reference resolution,
- new algorithms for resolving ellipted anaphors,
- multidocument personal name resolution,
- resolving references to referents in preexisting databases,
- reference resolution for improving document clustering,
- reference resolution for improving name recognition,
- reference resolution for the identification of discourse structure.

The workshop organizers would like to thank the program committee for their efforts in reviewing the submissions. We would also like to thank the participants for their contributions and their generous collaboration in both the preparation of the proceedings and the activities of the workshop itself.

Sanda Harabagiu & David Farwell (chairs)
July, 2004

Table of Contents

PREFACE	iii
PRELIMINARY PROGRAM	v
TABLE OF CONTENTS	vii
<i>Multi-Document Person Name Resolution</i> Michael Fleischman and Eduard Hovy	1
<i>Cross Document Co-Reference Resolution Applications for People in the Legal Domain</i> Christopher Dozier and Thomas Zielund	9
<i>Event Clustering on Streaming News Using Co-Reference Chains and Event Words</i> June-Jei Kuo and Hsin-Hsi Chen	17
<i>Coreference Resolution for Information Extraction</i> Dmitry Zelenko, Chinatsu Aone and Jason Tibbetts	24
<i>Applying Coreference to Improve Name Recognition</i> Heng Ji and Ralph Grishman	32
<i>Using Word Similarity Lists for Resolving Indirect Anaphora</i> Caroline Gasperin and Renata Vieira	40
<i>Discourse-New Detectors for Definite Description Resolution: A Survey and a Preliminary Proposal</i> Massimo Poesio, Olga Uryupina, Renata Vieira, Mijail Alexandrov-Kabadjov and Rodrigo Goulart	47
<i>Dynamic Centering</i> Daniel Hardt	55
<i>Ellipsis Resolution for Disguised Agent</i> Shigeko Nariyama	63
<i>Reference Resolution over a Restricted Domain: References to Documents</i> Andrei Popescu-Belis and Denis Lalanne	71
<i>BioAR: Anaphora Resolution for Relating Protein Names to Proteome Database Entries</i> Jung-Jae Kim and Jong C. Park	79
<i>Ellipsis Resolution by Controlled Default Unification for Multi-modal and Speech Dialog Systems</i> Michael Streit and HansUlrich Krieger	87
<i>An Algorithm for Resolving Individual and Abstract Anaphora in Danish Texts and Dialogues</i> Costanza Navarretta	95
<i>Topic Identification in Chinese Based on Centering Model</i> Ching-Long Yeh and Yi-Chun Chen	103
AUTHOR INDEX	110

PRELIMINARY PROGRAM

Sunday, July 25

- 9:00-9:20 Welcome, Description of a Reference Game
- 9:20-10:00 *Multi-Document Person Name Resolution*
Michael Fleischman and Eduard Hovy
- 10:00-10:30 COFFEE BREAK
- 10:30-11:10 *Cross Document Co-Reference Resolution Applications for People in the Legal Domain*
Christopher Dozier and Thomas Zielund
- 11:10-11:50 *Event Clustering on Streaming News Using Co-Reference Chains and Event Words*
June-Jei Kuo and Hsin-Hsi Chen
- 11:50-12:00 Results (1) Reference Game
- 12:00-13:30 LUNCH
- 13:30-14:10 *Coreference Resolution for Information Extraction*
Dmitry Zelenko, Chinatsu Aone and Jason Tibbetts
- 14:10-14:50 *Applying Coreference to Improve Name Recognition*
Heng Ji and Ralph Grishman
- 14:50-15:30 *Using Word Similarity Lists for Resolving Indirect Anaphora*
Caroline Gasperin and Renata Vieira
- 15:30-16:00 COFFEE BREAK
- 16:00-16:10 Results (2) Reference Game
- 16:10-16:50 *Discourse-New Detectors for Definite Description Resolution: A Survey and a Preliminary Proposal*
Massimo Poesio, Olga Uryupina, Renata Vieira, Mijail Alexandrov-Kabadjov and Rodrigo Goulart
- 16:50-17:30 *Dynamic Centering*
Daniel Hardt

Monday, July 26

- 9:00-9:20 Description of Reference Game 2
- 9:20-10:00 *Ellipsis Resolution for Disguised Agent*
Shigeiko Nariyama
- 10:00-10:30 COFFEE BREAK
- 10:30-11:10 *Reference Resolution over a Restricted Domain: References to Documents*
Andrei Popescu-Belis and Denis Lalanne
- 11:10-11:50 *BioAR: Anaphora Resolution for Relating Protein Names to Proteome Database Entries*
Jung-Jae Kim and Jong C. Park
- 11:50-12:00 Results (1) Reference Game 2
- 12:00-13:30 LUNCH
- 13:30-14:10 *Ellipsis Resolution by Controlled Default Unification for Multi-modal and Speech Dialog Systems*
Michael Streit and HansUlrich Krieger
- 14:10-14:50 *An Algorithm for Resolving Individual and Abstract Anaphora in Danish Texts and Dialogues*
Costanza Navarretta
- 14:50-15:30 *Topic Identification in Chinese Based on Centering Model*
Ching-Long Yeh and Yi-Chun Chen
- 15:30-16:00 COFFEE BREAK
- 16:00-16:10 Results (2) Reference Game 2
- 16:10-17:00 Discussion and Wrap-up