# Discourse Processing

# Synthesis Lectures on Human Language Technologies

Editor
**Graeme Hirst,** ***University of Toronto***

Synthesis Lectures on Human Languages Technologies is edited by Graeme Hirst of the University of Toronto. The series consists of 50- to 150-page monographs on topics relating to natural language processing, computational linguistics, information retrieval, and spoken language understanding. Emphasis is on important new techniques, on new applications, and on topics that combine two or more HLT subfields.

Discourse Processing
Manfred Stede
2011

Bitext Alignment
Jörg Tiedemann
2011

Linguistic Structure Prediction
Noah A. Smith
2011

Learning to Rank for Information Retrieval and Natural Language Processing
Hang Li
2011

Computational Modeling of Human Language Acquisition
Afra Alishahi
2010

Introduction to Arabic Natural Language Processing
Nizar Y. Habash
2010

Cross-Language Information Retrieval
Jian-Yun Nie
2010

Automated Grammatical Error Detection for Language Learners
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2010

Data-Intensive Text Processing with MapReduce
Jimmy Lin and Chris Dyer
2010

Semantic Role Labeling
Martha Palmer, Daniel Gildea, and Nianwen Xue
2010

Spoken Dialogue Systems
Kristiina Jokinen and Michael McTear
2009

Introduction to Chinese Natural Language Processing
Kam-Fai Wong, Wenjie Li, Ruifeng Xu, and Zheng-sheng Zhang
2009

Introduction to Linguistic Annotation and Text Analytics
Graham Wilcock
2009

Dependency Parsing
Sandra Kübler, Ryan McDonald, and Joakim Nivre
2009

Statistical Language Models for Information Retrieval
ChengXiang Zhai
2008

Discourse Processing

Manfred Stede

ISBN: 978-3-031-01016-3 paperback
ISBN: 978-3-031-02144-2 ebook

DOI 10.1007/978-3-031-02144-2

A Publication in the Springer Nature series
*SYNTHESIS LECTURES ON ADVANCES IN AUTOMOTIVE TECHNOLOGY*

Lecture #15
Series Editor: Graeme Hirst, *University of Toronto*
Series ISSN
Synthesis Lectures on Human Language Technologies
Print 1947-4040 Electronic 1947-4059

# Discourse Processing

Manfred Stede
University of Potsdam

*SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #15*

## ABSTRACT

*Discourse Processing* here is framed as marking up a text with structural descriptions on several levels, which can serve to support many language-processing or text-mining tasks. We first explore some ways of assigning structure on the document level: the logical document structure as determined by the layout of the text, its genre-specific content structure, and its breakdown into topical segments. Then the focus moves to phenomena of local coherence. We introduce the problem of coreference and look at methods for building chains of coreferring entities in the text. Next, the notion of coherence relation is introduced as the second important factor of local coherence. We study the role of connectives and other means of signaling such relations in text, and then return to the level of larger textual units, where tree or graph structures can be ascribed by recursively assigning coherence relations. – Taken together, these descriptions can inform text summarization, information extraction, discourse-aware sentiment analysis, question answering, and the like.

## KEYWORDS

text structure, document structure, topic segmentation, coreference, anaphora resolution, coherence relation, discourse parsing

# Contents

# Acknowledgments

The author is grateful to the anonymous reviewers and to the series editor for their thoughtful and constructive suggestions to improve an earlier version of the manuscript. Also, thanks are due to several readers who provided comments on earlier versions of individual sections: Heike Bieler, Markus Egg, Thomas Hanneforth, Manfred Klenner, Constantin Orasan, Maite Taboada, and Sebastian Varges.

Manfred Stede
November 2011