

# Contents

Preface v

## I Linguistic analyses of text resources

*Extending the TIGER query language with universal quantification* 3  
Torsten Marek, Joakim Lundborg and Martin Volk

*Exploring automatic theme identification: a rule-based approach* 15  
Lara Schwarz, Sabine Bartsch, Richard Eckart and Elke Teich

*Finding canonical forms for historical German text* 27  
Bryan Jurish

*Computing distance and relatedness of medieval text variants from German* 39  
Stefanie Dipper and Bettina Schrader

*Data structures for the analysis of regional language variation* 53  
Birgit Kellner, Timm Lehmborg, Ingrid Schröder and Kai Wörner

*Local syntactic tagging of large corpora using weighted finite state transducers* 65  
Jörg Didakowski

## II Extraction of lexical knowledge from text resources

*Towards improved text understanding with WordNet* 81  
Christiane Fellbaum, Peter Clark and Jerry Hobbs

*Influence of accurate compound noun splitting on bilingual vocabulary extraction* 91  
Marcin Junczys-Dowmunt

*A taxonomy of task-related dialogue actions: the cases of tutorial and collaborative planning dialogue* 105  
Magdalena Wolska and Mark Buckley