

ACL-08: HLT

**Workshop on
Parsing German
(PaGe-08)**

Proceedings of the Workshop

June 20, 2008
The Ohio State University
Columbus, Ohio, USA

Production and Manufacturing by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

Sponsorship provided by



©2008 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 978-1-932432-15-2

Introduction

Welcome to the ACL Workshop on Parsing German, the first of what we hope will be a long and fruitful series of workshops on this topic.

German possesses an interesting set of configurational properties on the syntactic level which make it far less flexible with respect to word order than other free word order languages. Analyses of these properties, which have formed a part of the traditional syntax of German since the early 19th century, only re-entered the mainstream of generative linguistics research within the last twenty years or so. In computational linguistics, however, their realization has varied quite widely: “topological fields” in HPSG-style analyses, multiple parse trees, special constraints on liberation in constraint-based dependency-style analyses, various hybrid “deep/shallow” approaches, and agnostic parameter estimation over graphs. This variation can also acutely be felt in the annotation of German treebanks. Many corpora have historically elected to annotate only a few of the different senses of the term “constituent” inherent to German syntax, resulting in standards that make German appear either more like English or more like Czech.

The aim of this workshop was to provide a forum for theoretical discussion as well as a shared task, based on the TIGER and TueBa-D/Z German treebanks, for these various approaches to make their case on empirical grounds. This combination we believe to be essential to balancing the considerations of what structure merits learning versus the ease with which it can be learned. Both treebanks are annotated collections of German newspaper text on similar topics. They are annotated with POS, morphology, phrase structure, and grammatical functions. TueBa-D/Z additionally uses topological fields to describe fundamental word order restrictions in German clauses. The treebanks differ significantly in their annotation schemes, however: while TIGER relies on crossing branches to describe long distance relationships, TueBa-D/Z uses pure tree structures with designated labels for long distance relationships. Additionally, the annotation in TIGER is flat on the phrasal level while TueBa-D/Z annotates phrasal structure more hierarchically.

A report on the results of this year’s shared task can be found in the final paper of these proceedings.

Organizers:

Sandra Kübler, Indiana University (USA)
Gerald Penn, University of Toronto (Canada)

Program Committee:

Berthold Crysman, University of Bonn (Germany)
Amit Dubey, University of Edinburgh (UK)
Anette Frank, University of Heidelberg (Germany)
Erhard Hinrichs, University of Tübingen (Germany)
Julia Hockenmeier, University of Illinois (USA)
Laura Kallmeyer, University of Tübingen (Germany)
Frank Keller, University of Edinburgh (UK)
Wolfgang Menzel, University of Hamburg (Germany)
Stefan Müller, Free University of Berlin (Germany)
Stefan Oepen, University of Oslo (Norway)
Helmut Schmid, University of Stuttgart (Germany)
Gerold Schneider, University of Zürich (Switzerland)
Hans Uszkoreit, University of the Saarland (Germany)
Josef van Genabith, Dublin City University (Ireland)

Table of Contents

<i>Lexicalised Parsing of German V2</i>	
Yo Sato	1
<i>Parse Selection with a German HPSG Grammar</i>	
Berthold Crysmann	9
<i>Part-of-Speech Tagging with a Symbolic Full Parser: Using the TIGER Treebank to Evaluate Fips</i>	
Yves Scherrer	16
<i>Revisiting the Impact of Different Annotation Schemes on PCFG Parsing: A Grammatical Dependency Evaluation</i>	
Adriane Boyd and Detmar Meurers	24
<i>Parsing German with Latent Variable Grammars</i>	
Slav Petrov and Dan Klein	33
<i>Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines</i>	
Anna Rafferty and Christopher D. Manning	40
<i>A Dependency-Driven Parser for German Dependency and Constituency Representations</i>	
Johan Hall and Joakim Nivre	47
<i>The PaGe 2008 Shared Task on Parsing German</i>	
Sandra Kübler	55

Workshop Program

Friday, June 20, 2008

- 8:45–9:00 Opening Remarks
- 9:00–9:30 *Lexicalised Parsing of German V2*
Yo Sato
- 9:30–10:00 *Parse Selection with a German HPSG Grammar*
Berthold Crysmann
- 10:00–10:30 *Part-of-Speech Tagging with a Symbolic Full Parser: Using the TIGER Treebank to Evaluate Fips*
Yves Scherrer
- 10:30–11:00 Break
- 11:00–12:00 Invited Talk by Wolfgang Menzel
- 12:00–12:30 *Revisiting the Impact of Different Annotation Schemes on PCFG Parsing: A Grammatical Dependency Evaluation*
Adriane Boyd and Detmar Meurers
- 12:30–14:00 Lunch
- Shared Task**
- 14:00–14:30 *Parsing German with Latent Variable Grammars*
Slav Petrov and Dan Klein
- 14:30–15:00 *Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines*
Anna Rafferty and Christopher D. Manning
- 15:00–15:30 *A Dependency-Driven Parser for German Dependency and Constituency Representations*
Johan Hall and Joakim Nivre
- 15:30–16:00 Break
- 16:00–16:30 *The PaGe 2008 Shared Task on Parsing German*
Sandra Kübler

Friday, June 20, 2008 (continued)

16:30–17:30 Panel Discussion