

# **20th Conference on Natural Language Processing/Konferenz zur Verarbeitung natürlicher Sprache (KONVENS 2024)**

Vienna, Austria  
10 – 13 September 2024

ISBN: 979-8-3313-0484-3

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2024) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Large Language Models as Evaluators for Scientific Synthesis</i> Julia Evans, Jennifer D’Souza and Sören Auer .....	1
<i>A Crosslingual Approach to Dependency Parsing for Middle High German</i> Cora Haiber .....	23
<i>Complexity of German Texts Written by Primary School Children</i> Jammila Laâguidi, Dana Neumann, Ronja Laarmann-Quante, Stefanie Dipper and Mihail Chifligarov .....	32
<i>Exploring Automatic Text Simplification for Lithuanian</i> Justina Mandravickaitė, Egle Rimkiene, Danguolė Kalinauskaitė and Danguolė Kotryna Kapkan .....	40
<i>Word alignment in Discourse Representation Structure parsing</i> Christian Obereder and Gabor Recski .....	50
<i>Evaluating and Fine-Tuning Retrieval-Augmented Language Models to Generate Text with Accurate Citations</i> Vinzent Penzkofer and Timo Baumann .....	57
<i>Discourse Parsing for German with new RST Corpora</i> Sara Shahmohammadi and Manfred Stede .....	65
<i>Revisiting the Phenomenon of Syntactic Complexity Convergence on German Dialogue Data</i> Yu Wang and Hendrik Buschmeier .....	75
<i>Estimating Word Concreteness from Contextualized Embeddings</i> Christian Wartena .....	81
<i>Using GermaNet for the Generation of Crossword Puzzles</i> Claus Zinn, Marie Hinrichs and Erhard Hinrichs .....	89
<i>Leveraging Cross-Lingual Transfer Learning in Spoken Named Entity Recognition Systems</i> Moncef Benaïcha, David Thulke and Mehmet Ali Tuğtekin Turan .....	98
<i>Exploring Data Acquisition Strategies for the Domain Adaptation of QA Models</i> Maurice Falk, Adrian Ulges and Dirk Krechel .....	106
<i>CO-Fun: A German Dataset on Company Outsourcing in Fund Prospectuses for Named Entity Recognition and Relation Extraction</i> Neda Foroutan, Markus Schröder and Andreas Dengel .....	117
<i>GERestaurant: A German Dataset of Annotated Restaurant Reviews for Aspect-Based Sentiment Analysis</i> Nils Constantin Hellwig, Jakob Fehle, Markus Bink and Christian Wolff .....	123
<i>How to Translate SQuAD to German? A Comparative Study of Answer Span Retrieval Methods for Question Answering Dataset Creation</i> Jens Kaiser and Agnieszka Falenska .....	134
<i>Few-Shot Prompting for Subject Indexing of German Medical Book Titles</i> Lisa Kluge and Maximilian Kähler .....	141

<i>Binary indexes for optimising corpus queries</i>	
Peter Ljunglöf, Nicholas Smallbone, Mijo Thoresson and Victor Salomonsson . . . . .	149
<i>An Improved Method for Class-specific Keyword Extraction: A Case Study in the German Business Registry</i>	
Stephen Meisenbacher, Tim Schopf, Weixin Yan, Patrick Holl and Florian Matthes .	159
<i>Tabular JSON: A Proposal for a Pragmatic Linguistic Data Format</i>	
Adam Roussel . . . . .	166
<i>Semiautomatic Data Generation for Academic Named Entity Recognition in German Text Corpora</i>	
Pia Schwarz . . . . .	173
<i>Redundancy Aware Multiple Reference Based Gainwise Evaluation of Extractive Summarization</i>	
Mousumi Akter and Santu Karmaker . . . . .	182
<i>Fine-grained quotation detection and attribution in German news articles</i>	
Fynn Petersen-Frey and Chris Biemann . . . . .	196
<i>Decoding 16th-Century Letters: From Topic Models to GPT-Based Keyword Mapping</i>	
Phillip Benjamin Ströbel, Stefan Aderhold and Ramona Roller . . . . .	209
<i>Analysing Effects of Inducing Gender Bias in Language Models</i>	
Stephanie Gross, Brigitte Krenn, Craig Lincoln and Lena Holzwarth . . . . .	222
<i>OMoS-QA: A Dataset for Cross-Lingual Extractive Question Answering in a German Migration Context</i>	
Steffen Kleinle, Jakob Prange and Annemarie Friedrich . . . . .	231
<i>Role-Playing LLMs in Professional Communication Training: The Case of Investigative Interviews with Children</i>	
Don Tuggener, Teresa Schneider, Ariana Huwiler, Tobias Kreienbühl, Simon Hischier, Pius Von Däniken and Susanna Niehaus . . . . .	249
<i>Features and Detectability of German Texts Generated with Large Language Models</i>	
Verena Irrgang, Veronika Solopova, Steffen Zeiler, Robert M. Nickel and Dorothea Kolossa . . . . .	264
<i>Lex2Sent: A bagging approach to unsupervised sentiment analysis</i>	
Kai-Robin Lange, Jonas Rieger and Jonas Rieger . . . . .	281
<i>Discourse-Level Features in Spoken and Written Communication</i>	
Hannah J. Seemann, Sara Shahmohammadi, Manfred Stede and Tatjana Scheffler . .	292
<i>Version Control for Speech Corpora</i>	
Vlad Dumitru, Matthias Boehm, Martin Hagmüller and Barbara Schuppler . . . . .	303
<i>Querying Repetitions in Spoken Language Corpora</i>	
Elena Frick, Henrike Helmer and Dolores Lemmenmeier-Batinić . . . . .	309
<i>Exploring Phonetic Features in Language Embeddings for Unseen Language Varieties of Austrian German</i>	
Lorenz Gutscher and Michael Pucher . . . . .	317

<i>A Multilingual Dataset of Adversarial Attacks to Automatic Content Scoring Systems</i> Ronja Laarmann-Quante, Christopher Chandler, Noemi Incirkus, Vitaliia Ruban, Alona Solopov and Luca Steen .....	326
<i>Towards Improving ASR Outputs of Spontaneous Speech with LLMs</i> Karner Manuel, Julian Linke, Mark Kröll, Barbara Schuppler and Bernhard C. Geiger	339
<i>OneLove beyond the field - A few-shot pipeline for topic and sentiment analysis during the FIFA World Cup in Qatar</i> Christoph Rauchegger, Sonja Mei Wang and Pieter Delobelle .....	349
<i>Linguistic and extralinguistic factors in automatic speech recognition of German atypical speech</i> Eugenia Rykova and Mathias Walther .....	358
<i>LLM-based Translation Across 500 Years. The Case for Early New High German</i> Martin Volk, Dominic P. Fischer, Patricia Scheurer, Raphael Schwitter and Phillip B. Ströbel .....	368