

21st Conference on Natural Language Processing/Konferenz zur Verarbeitung natürlicher Sprache (KONVENS 2025)

Volume 1 - Long and Short Papers

Hildesheim, Germany
9-12 September 2025

ISBN: 979-8-3313-2674-6

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© (2025) by the Association for Computational Linguistics
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY. (<https://creativecommons.org/licenses/by/4.0/>)

Printed with permission by Curran Associates, Inc. (2026)

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

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
Web: www.proceedings.com

Contents

Methods

- Surprisal in Action: A Comparative Study of LDA and LSA for Keyword Extraction 1
J. Nathanael Philipp, Max Kölbl, Michael Richter
- Learn to pick the winner: Black-box ensembling for textual and visual question answering . . . 12
Yuxi Xia, Klim Zaporozets, Benjamin Roth
- LRMs are not thinking straight: Unreliability of thinking trajectories 27
Jhouben Cuesta-Ramirez, Samuel Beaussant, Mehdi Mounsif

Applications

- Adaption and Evaluation of Generative Large Language Models for German Medical Information Extraction 33
Sören Spiegel, Seid Muhie Yimam, Philipp Breitfeld, Frank Ückert
- ZEFYS2025: A German Historical Newspaper Dataset for Named Entity Recognition and Entity Linking 48
Sophie Schneider, Ulrike Förstel, Kai Labusch, Jörg Lehmann, Clemens Neudecker
- Generating Search-Engine-Optimized Headlines for Sports News 59
Frank Zalkow, Benedikt Schäfer, Thomas Moissl, Jonas Bücherl, Kerstin Markl, Sebastian Bothe, Francois Duchateau, Julia Dollase, Patric Kabus, Daniel Steinigen, Oliver Schmitt, Fabian Küch

New Resources

- Predicting Functional Content Zones in German Source-Dependent Argumentative Essays: Experiments on a Novel Dataset 66
Xiaoyu Bai, Manfred Stede
- SocCor: A Multimodal-based Multilingual Soccer Corpus for Text Data Analytics 80
Paul Löhr, Jannik Strötgen
- A Survey of Idiom Datasets for Psycholinguistic and Computational Research 90
Michael Flor, Xinyi Liu, Anna Feldman

Discourse and Semantics

- Function Words as Stable Features for German Opinion Articles Classification 101
Amelie Schmidt-Colberg, Simon Burkard, Anne Grohnert, Michael John
- LLM-based Classification of Grounding Acts in German 112
Milena Belosevic, Hendrik Buschmeier
- Efficient and Effective Coreference Resolution for German 128
Fynn Petersen-Frey, Hans Ole Hatzel, Chris Biemann

Hate Speech

- FASCIST-O-METER: Classifier for Neo-fascist Discourse Online 138
Rudy Alexandro Garrido Veliz, Martin Semmann, Chris Biemann, Seid Muhie Yimam
- Conditioning Large Language Models on Legal Systems? Detecting Punishable Hate Speech . 154
Florian Ludwig, Torsten Zesch, Frederike Zufall

HICC: A Dataset for German Hate Speech in Conversational Context	168
<i>Lars Schmid, Pius von Däniken, Patrick Giedemann, Don Tuggener, Judith Bühler, Maria Kamenowski, Katja Girschick, Dirk Baier, Mark Cieliebak</i>	
Translation and Multilinguality	
Evaluating the Feasibility of Using ChatGPT for Cross-cultural Survey Translation	177
<i>Danielly Sorato, Diana Zavala-Rojas</i>	
Information Divergence in Translation and Interpreting: Findings from Same-Source Texts . . .	192
<i>Maria Kunilovskaya, Sharid Loáiciga, Ekaterina Lapshinova-Koltunski</i>	
SEAS: Sentence Extraction and Alignment from Subtitles	203
<i>Josh Stephenson, Libby Barak</i>	
Methods and Applications for German and English	
German Aspect-based Sentiment Analysis in the Wild: B2B Dataset Creation and Cross-Domain Evaluation	213
<i>Jakob Fehle, Niklas Donhauser, Udo Kruschwitz, Nils Constantin Hellwig, Christian Wolff</i>	
Automatic Creation of Marginalia	228
<i>Aaron Lang, Robin Jegan, Andreas Henrich</i>	
Localization of English Affective Narrative Generation to German	241
<i>Johannes Schäfer, Sabine Weber, Roman Klinger</i>	
Multimodal Docker Unified UIMA Interface: New Horizons for Distributed Microservice-Oriented Processing of Corpora using UIMA	257
<i>Daniel Bundan, Giuseppe Abrami, Alexander Mehler</i>	
Systematic Review of Linguistic Characteristics in Profiling and Automated Detection of Autistic Speech	269
<i>Charlotte Bellinghausen, Andreas Riedel</i>	
Towards a Cross-Dialectal Dictionary for Low German (Low Saxon)	282
<i>Christian Chiarcos, Janine Siewert, Tabea Gröger, Christian Fäth</i>	
Applying an Information-theoretic Approach for Automatic Identification of German Multi-word Expressions	295
<i>Sergei Bagdasarov, Elke Teich</i>	
Hit or Be Hit: Tests of (Pre)Compositional Abilities in Vision and Language Models	306
<i>Mădălina Zgreabă, Albert Gatt, Pablo Mosteiro</i>	
Hybrid Feature-Embedding Models for Robust AI Text Detection	318
<i>Kasper Thomas Gartside Knudsen, Christian Hardmeier</i>	
Using LLMs for experimental stimulus pretests in linguistics. Evidence from semantic associations between words and social gender	326
<i>Christian Lang, Franziska Kretzschmar, Sandra Hansen</i>	
PETapter: Leveraging PET-style classification heads for modular few-shot parameter-efficient fine-tuning	333
<i>Jonas Rieger, Mattes Ruckdeschel, Gregor Wiedemann</i>	
Large Language Model Data Generation for Enhanced Intent Recognition in German Speech . .	349
<i>Theresa Pekarek Rosin, Burak Can Kaplan, Stefan Wermter</i>	
Detecting Sexism and Its Severity in German Online Comments: Modeling Annotation Subjectivity with BERT and mBERT	360
<i>Melanie Woodrow, Margot Mieskes</i>	