

KONVENS 2023

**The 19th Conference on Natural Language Processing
(KONVENS 2023)**

Proceedings of the Conference

September 18 - 22, 2023

©2023 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 979-8-89176-029-5

Message from the Organizers

KONVENS (Konferenz zur Verarbeitung natürlicher Sprache/Conference on Natural Language Processing) is a conference series on computational linguistics established in 1992. It is held annually in Germany, Austria, and Switzerland. KONVENS is organized under the auspices of the German Society for Computational Linguistics and Language Technology, the Special Interest Group on Computational Linguistics of the German Linguistic Society, the Austrian Society for Artificial Intelligence and SwissText.

The papers in these proceedings were presented at the 19th edition of KONVENS, which took place from September 19 to September 21, 2023, at Technische Hochschule Ingolstadt (THI) in Germany. The KONVENS 2023 main conference received over 50 submissions, which were peer-reviewed by three reviewers each. After the review, 22 papers were accepted for presentation at the conference and archived in these proceedings. Several non-archival works have also been presented at the conference. The KONVENS main conference was accompanied by several complementary events: three workshops, a shared task on speaker attribution in newswire and parliamentary debates, a tutorial on learning from task instructions and an event for PhD students. In addition, KONVENS 2023 featured the GSCL Awards for bachelor's and master's theses.

Many thanks to all who submitted their work to KONVENS and our board of reviewers for supporting us greatly with evaluating the submissions. Moreover, we would like to thank Technische Hochschule Ingolstadt, everyone involved in organizing and conducting the conference, and, of course, our sponsors. Without their support, KONVENS 2023 would not have been possible.

Munir Georges (General Chair)

Aaricia Herygers (Local Chair)

Annemarie Friedrich (Program Co-Chair)

Benjamin Roth (Program Co-Chair)

Sponsors

We extend our heartfelt gratitude to our sponsors for their generous support of the KONVENS conference. Their commitment has been instrumental in the success of this event, enabling a vibrant platform for knowledge sharing and innovation.

Gold Sponsors

Bundesdruckerei



Careers in the Bundesdruckerei Group

Tradition and innovation in harmony, that's what Bundesdruckerei is all about. Secure identities, secure data and infrastructures are our core expertise, enabling us to shape digital transformation in a responsible manner. What do we need? We need great people full of ideas and energy who give their best every day. What you can expect is a wide range of activities and countless opportunities for further development.

Secure. Satisfying. Career.

We are all about digital data and identities. With innovative products and advanced technologies. Our customers and the public particularly appreciate the fact that we provide security and trust. We have been providing security "Made in Germany" since 1868. Do innovative products and new concepts excite you too? Are you keen to push new developments? Are you looking to get progress moving and shape the future? Then we should get together.

Start your career at the Bundesdruckerei Group

Join our team and find the job you're looking for in our job vacancies:

<https://www.bundesdruckerei.de/de/karriere/jobs>

Bosch Center for Artificial Intelligence



The Bosch Center for Artificial Intelligence (BCAI) is leading among the industrial AI research centers in Europe. BCAI develops modern AI technologies for Bosch products and services "invented for life." Together with experts from all over the globe, Bosch conducts cutting edge research in applied AI.

<https://www.bosch-ai.com/>

Copper Sponsors



Organizing Committee

Munir Georges (General Chair)
Aaricia Herygers (Local Chair)
Annemarie Friedrich (Program Co-Chair)
Benjamin Roth (Program Co-Chair)

Program Committee

Adrien Barbaresi (Berlin-Brandenburgische Akademie der Wissenschaften)
Agnieszka Faleńska (Universität Stuttgart)
Alexander Fraser (Ludwig-Maximilians-Universität München)
Alexander Mehler (Goethe-Universität Frankfurt)
Andrea Horbach (FernUniversität Hagen)
Anke Holler (Universität Göttingen)
Asad Sayeed (University of Gothenburg)
Barbara Plank (Ludwig-Maximilians-Universität München)
Barbara Schuppler (Technische Universität Graz)
Bernhard Fisseni (Universität Duisburg-Essen)
Bernhard Schröder (University of Duisburg-Essen)
Brigitte Krenn (Österreichisches Forschungsinstitut für Artificial Intelligence)
Casey Kennington (Boise State University)
Cerstin Mahlow (Zürcher Hochschule für Angewandte Wissenschaften)
Chris Biemann (Universität Hamburg)
Christian Chiarcos (Universität zu Köln)
Christian Wartena (Hochschule Hannover)
Clemens Neudecker (Staatsbibliothek zu Berlin - Preußischer Kulturbesitz)
David Schlangen (Universität Potsdam)
Debayan Banerjee (Universität Hamburg)
Ekaterina Lapshinova-Koltunski (Universität Hildesheim)
Eva Maria Vecchi (Universität Stuttgart)
Gabriella Lapesa (Universität Stuttgart)

Georg Rehm (Deutsches Forschungszentrum für Künstliche Intelligenz)
Gertrud Faaß (Universität Hildesheim)
Heike Zinsmeister (Universität Hamburg)
Hendrik Schuff (Bosch Center for Artificial Intelligence, Universität Stuttgart)
Henning Wachsmuth (Leibniz Universität Hannover)
Irina Nikishina (Universität Hamburg)
Josef Ruppenhofer (IDS Mannheim)
Katharina Kann (University of Colorado Boulder)
Kerstin Jung (Universität Stuttgart)
Kilian Evang (Heinrich-Heine-Universität Düsseldorf)
Korbinian Riedhammer (Technische Hochschule Nürnberg)
Manfred Stede (Universität Potsdam)
Marc Schulder (Universität Hamburg)
Marcel Bollmann (Linköping University)
Margot Mieskes (Hochschule für Angewandte Wissenschaften Darmstadt)
Maria Berger (Ruhr-Universität Bochum)
Melanie Andresen (Universität Stuttgart)
Michael Roth (Universität Stuttgart)
Nicolai Erbs (DB Fernverkehr)
Odette Scharenborg (Delft University of Technology)
Peter Bourgonje (Universität Potsdam)
Rainer Osswald (Heinrich-Heine-Universität Düsseldorf)
Roman Klinger (Universität Stuttgart)
Saba Anwar (Universität Hamburg)
Sabine Schulte im Walde (Universität Stuttgart)
Sebastian Pado (Universität Stuttgart)
Seid Muhie Yimam (Universität Hamburg)
Simone Ponzetto (Universität Mannheim)
Sophie Henning (Bosch Center for Artificial Intelligence, Ludwig-Maximilians-Universität München)
Tanja Samardzic (Universität Zürich)
Tatjana Scheffler (Ruhr-Universität Bochum)
Tobias Bocklet (Technische Hochschule Nürnberg)

Torsten Zesch (FernUniversität Hagen)

Udo Kruschwitz (Universität Regensburg)

Valia Kordoni (Humboldt-Universität zu Berlin)

Yves Scherrer (University of Helsinki)

Table of Contents

<i>Is Prompt-Based Finetuning Always Better than Vanilla Finetuning? Insights from Cross-Lingual Language Understanding</i>	
Bolei Ma, Ercong Nie, Helmut Schmid and Hinrich Schuetze	1
<i>Comparing Pre-Training Schemes for Luxembourgish BERT Models</i>	
Cedric Lothritz, Saad Ezzini, Christoph Purschke, Tegawendé Bissyandé, Jacques Klein, Isabella Olariu, Andrey Boytsov, Clément LeFebvre and Anne Goujon	17
<i>LLpro: A Literary Language Processing Pipeline for German Narrative Texts</i>	
Anton Ehrmantraut, Leonard Konle and Fotis Jannidis	28
<i>Classifying Noun Compounds for Present-Day Compositionality: Contributions of Diachronic Frequency and Productivity Patterns</i>	
Maximilian M. Maurer, Chris Jenkins, Filip Miletić and Sabine Schulte im Walde	40
<i>From Qualitative to Quantitative Research: Semi-Automatic Annotation Scaling in the Digital Humanities</i>	
Fynn Petersen-Frey, Tim Fischer, Florian Schneider, Isabel Eiser, Gertraud Koch and Chris Bie-mann	52
<i>Data and Approaches for German Text simplification – towards an Accessibility-enhanced Communication</i>	
Thorben Schomacker, Michael Gille, Marina Tropmann-Frick and Jörg von der Hülls	63
<i>Steps towards Addressing Text Classification in Low-Resource Languages</i>	
Maximilian Weißenbacher and Udo Kruschwitz	69
<i>Toward a Multilingual Connective Database: Aligning German/French Concessive Connectives</i>	
Peter Bourgonje, Sophia Rauh and Karolina Zaczynska	77
<i>Factuality Detection using Machine Translation – a Use Case for German Clinical Text</i>	
Mohammed Bin Sumait, Aleksandra Gabryszak, Leonhard Hennig and Roland A. Roller	85
<i>Linking Danish Parser Output to a Central Word Repository - From Morphosemantic Disambiguation to Unique Identifiers</i>	
Eckhard Bick	93
<i>Automatic Dictionary Generation: Could Brothers Grimm Create a Dictionary with BERT?</i>	
Hendryk Weiland, Maike Behrendt and Stefan Harmeling	102
<i>Towards UkrainianWordNet: Incorporation of an Existing Thesaurus in the Domain of Physics</i>	
Melanie Siegel, Maksym Vakulenko and Jonathan Baum	121
<i>Pointer Networks: A Unified Approach to Extracting German Opinions</i>	
Julia Wunderle, Jan Pfister and Andreas Hotho	127
<i>Exploring Automatic Text Simplification of German Narrative Documents</i>	
Thorben Schomacker, Tillmann Dönicke and Marina Tropmann-Frick	139
<i>Are idioms surprising?</i>	
Jonas Nathanael Philipp, Michael Richter, Erik Daas and Max Kölbl	149

<i>Answer Candidate Type Selection: Text-To-Text Language Model for Closed Book Question Answering Meets Knowledge Graphs</i>	
Mikhail Salnikov, Maria Lysyuk, Pavel Braslavski, Anton Razzhigaev, Valentin A. Malykh and Alexander Panchenko	155
<i>HS-EMO: Analyzing Emotions in Hate Speech</i>	
Johannes Schäfer and Elina Kistner	165
<i>Evaluating Data Augmentation Techniques for the Training of Luxembourgish Language Models</i>	
Isabella Olariu, Cedric Lothritz, Tegawendé Bissyandé and Jacques Klein	174
<i>Information Type Classification with Contrastive Task-Specialized Sentence Encoders</i>	
Philipp Seeberger, Tobias Bocklet and Korbinian Riedhammer	180
<i>German Text Embedding Clustering Benchmark</i>	
Silvan Wehrli, Bert Arnrich and Christopher Irrgang	187
<i>Aspect-Based Sentiment Analysis as a Multi-Label Classification Task on the Domain of German Hotel Reviews</i>	
Jakob Fehle, Leonie Münster, Thomas Schmidt and Christian Wolff	202
<i>Political claim identification and categorization in a multilingual setting: First experiments</i>	
Urs Zaberer, Sebastian Pado and Gabriella Lapesa	219

Conference Program

Tuesday, September 19, 2023

8:45–9:00 *Opening Remarks*

9:00–10:00 *Keynote 1 by Marc Schulder: Being a computational linguist in sign language research*

10:00–10:30 *Coffee Break*

10:30–12:00 **Session 1: Paper Presentations & Sponsor Talk**

10:30–10:45 *Is Prompt-Based Finetuning Always Better than Vanilla Finetuning? Insights from Cross-Lingual Language Understanding*

Bolei Ma, Ercong Nie, Helmut Schmid and Hinrich Schuetze

10:45–11:00 *Comparing Pre-Training Schemes for Luxembourgish BERT Models*

Cedric Lothritz, Saad Ezzini, Christoph Purschke, Tegawendé Bissyandé, Jacques Klein, Isabella Olariu, Andrey Boytsov, Clément LeFebvre and Anne Goujon

11:00–11:15 *LLpro: A Literary Language Processing Pipeline for German Narrative Texts*

Anton Ehrmanntraut, Leonard Konle and Fotis Jannidis

11:15–11:30 *Classifying Noun Compounds for Present-Day Compositionality: Contributions of Diachronic Frequency and Productivity Patterns*

Maximilian M. Maurer, Chris Jenkins, Filip Miletić and Sabine Schulte im Walde

11:30–11:45 *From Qualitative to Quantitative Research: Semi-Automatic Annotation Scaling in the Digital Humanities*

Fynn Petersen-Frey, Tim Fischer, Florian Schneider, Isabel Eiser, Gertraud Koch and Chris Biemann

11:45–12:00 *Sponsor Presentation by Bosch*

12:00–14:00 *Lunch Break*

Tuesday, September 19, 2023 (continued)

14:00–14:40 Session 2: Poster Session 1 Flashlights & Sponsor Talk

14:00–14:20 *Poster Session 1 Flashlights: Advertisement*

14:20–14:40 *Sponsor Presentation by Bundesdruckerei*

14:40–16:00 *Poster Session 1 Projects & PhD Thesis Presentations*

Data and Approaches for German Text simplification – towards an Accessibility-enhanced Communication

Thorben Schomacker, Michael Gille, Marina Tropmann-Frick and Jörg von der Hülls

Steps towards Addressing Text Classification in Low-Resource Languages

Maximilian Weißenbacher and Udo Kruschwitz

15:30–16:00 *Coffee Served*

16:00–17:00 Session 3: Fun Networking Session

17:00–17:15 *Thierry Declerck Memorial*

17:15–17:30 *Break*

17:30–19:00 *GSCL Members Meeting*

19:00– *Welcome Reception*

Wednesday, September 20, 2023

9:00–10:00 *Keynote 2 by Ivana Kruijff-Korbayová*

10:00–10:30 *Coffee Break*

10:30–12:00 *Thesis Presentations*

12:00–14:00 *Lunch Break*

14:00–14:45 **Session 4: Paper Presentations**

14:00–14:15 *Toward a Multilingual Connective Database: Aligning German/French Concessive Connectives*
Peter Bourgonje, Sophia Rauh and Karolina Zaczynska

14:15–14:30 *Factuality Detection using Machine Translation – a Use Case for German Clinical Text*
Mohammed Bin Sumait, Aleksandra Gabryszak, Leonhard Hennig and Roland A. Roller

14:30–14:45 *Poster Session 2 Flashlights*

14:40–16:00 **Poster Session 2 Projects & PhD Thesis Presentations**

Linking Danish Parser Output to a Central Word Repository - From Morphosemantic Disambiguation to Unique Identifiers
Eckhard Bick

Automatic Dictionary Generation: Could Brothers Grimm Create a Dictionary with BERT?
Hendryk Weiland, Maike Behrendt and Stefan Harmeling

Towards UkrainianWordNet: Incorporation of an Existing Thesaurus in the Domain of Physics
Melanie Siegel, Maksym Vakulenko and Jonathan Baum

Wednesday, September 20, 2023 (continued)

Pointer Networks: A Unified Approach to Extracting German Opinions

Julia Wunderle, Jan Pfister and Andreas Hotho

Exploring Automatic Text Simplification of German Narrative Documents

Thorben Schomacker, Tillmann Dönicke and Marina Tropmann-Frick

Are idioms surprising?

Jonas Nathanael Phlipp, Michael Richter, Erik Daas and Max Kölbl

Answer Candidate Type Selection: Text-To-Text Language Model for Closed Book Question Answering Meets Knowledge Graphs

Mikhail Salnikov, Maria Lysyuk, Pavel Braslavski, Anton Razzhigaev, Valentin A. Malykh and Alexander Panchenko

HS-EMO: Analyzing Emotions in Hate Speech

Johannes Schäfer and Elina Kistner

Evaluating Data Augmentation Techniques for the Training of Luxembourgish Language Models

Isabella Olariu, Cedric Lothritz, Tegawendé Bissyandé and Jacques Klein

Information Type Classification with Contrastive Task-Specialized Sentence Encoders

Philipp Seeberger, Tobias Bocklet and Korbinian Riedhammer

15:30–16:00 *Coffee Served*

16:00–18:00 *Social Event*

18:00– *Conference Dinner*

Thursday, September 21, 2023

9:00–10:00 *Keynote 3 by Hinrich Schuütze: Glot500: Scaling Multilingual Corpora and Language Models to 500 Languages*

10:00–10:30 *Coffee Break*

10:30–11:30 **Session 5: Paper Presentations**

10:30–10:45 *German Text Embedding Clustering Benchmark*
Silvan Wehrli, Bert Arnrich and Christopher Irrgang

10:45–11:00 *Aspect-Based Sentiment Analysis as a Multi-Label Classification Task on the Domain of German Hotel Reviews*
Jakob Fehle, Leonie Münster, Thomas Schmidt and Christian Wolff

11:00–11:15 *Political claim identification and categorization in a multilingual setting: First experiments*
Urs Zaberer, Sebastian Pado and Gabriella Lapesa

11:15–11:30 *Policy Domain Prediction from Party Manifestos with Adapters and Knowledge Enhanced Transformers*
Hsiao-Chu Yu, Ines Rehbein and Simone Paolo Ponzetto

11:30–12:00 *GSCL Thesis Awards & Closing*

