

KONVENS 2022

**Proceedings of the 18th Conference on Natural Language
Processing/Konferenz zur Verarbeitung natürlicher
Sprache (KONVENS 2022)**

12-15 September, 2022
University of Potsdam
Potsdam, Germany

Introduction

The papers of these proceedings have been presented at the 18th edition of KONVENS (Konferenz zur Verarbeitung natürlicher Sprache/Conference on Natural Language Processing). KONVENS is a conference series on computational linguistics established in 1992 that was held biennially until 2018 and has been held annually since. 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 18th KONVENS took place on-site from September 12 to September 15, 2022 at University of Potsdam. The KONVENS main conference was accompanied by a workshop, a shared task (GermEval), two tutorials and a ‘PhD Day’. In addition, this year’s edition hosted a career networking event. In total these proceedings contain 21 papers (10 long, 11 short).

Many thanks to all who submitted their work to KONVENS and to our board of reviewers for supporting us greatly with evaluating the submissions. Moreover we would like to thank University of Potsdam for providing the conference rooms, all people involved in organisation, and our sponsors. Without their support KONVENS 2022 would not have been possible.

Robin Schaefer
Xiaoyu Bai
Manfred Stede
Torsten Zesch

People

Local Organization

Mohammad Yeghaneh Abkenar, Xiaoyu Bai, Annemarie Friedrich (GSCL), Freya Hewett, René Knaebel, Robin Schaefer, Manfred Stede

Program Chairs

Xiaoyu Bai, Robin Schaefer, Manfred Stede, Torsten Zesch

Program Committee

Adrien Barbaresi, Maria Berger, Marcel Bollmann, Peter Bourgonje, Ernst Buchberger, Berthold Crysmann, Stefanie Dipper, Stephanie Evert, Jana Götze, Anke Holler, Roman Klinger, Valia Kordoni, Brigitte Krenn, Udo Kruschwitz, Ekaterina Lapshinova-Koltunski, Katja Markert, Alexander Mehler, Clemens Neudecker, Rainer Osswald, Simon Ostermann, Simone Paolo Ponzetto, Ines Rehbein, Georg Rehm, Josef Ruppenhofer, Felix Sasaki, Roland Schäfer, Tatjana Scheffler, Yves Scherrer, David Schlangen, Helmut Schmid, Gerold Schneider, Roman Schneider, Sabine Schulte im Walde, Maria Skeppstedt, Manfred Stede, Henning Wachsmuth, Magdalena Wolska, Sina Zarriß, Torsten Zesch, Heike Zinsmeister

Satellite Events

2nd Workshop on Computational Linguistics for Political Text Analysis

Organizers: Ines Rehbein, Christopher Klamm, Simone Ponzetto, Gabriella Lapesa

Text Complexity DE Challenge 2022 (GermEval)

Organizers: Salar Mohtaj, Babak Naderi, Sebastian Möller

Text to talk: foundations of interactive language modeling for conversational AI and talking robots (Tutorial)

Organizers: Andreas Liesenfeld, Ada Lopez, Mark Dingemanse

Retico – An Introduction to Building Incremental Dialogue Systems in Python (Tutorial)

Organizers: Thilo Michael, Maike Paetzel-Prüsmann, Jana Götze, David Schlangen

PhD Day

Organized by local organizers

Invited Talks

Malvina Nissim: In Other Words. Models and Evaluation for Text Style Transfer

Whenever we write about something, we make a choice (consciously or not) on how we do it. For example, I can write about a series I watched while I was COVID-bound at home like this: ‘I viewed it and I believe it is a high quality program.’ but also like this: ‘I’ve watched it and it is AWESOME!!!!’. The content is (approximately) the same, but the style I’ve used is different: informal in the second formulation, much more formal in the first one. In the larger field of Natural Language Generation, text style transfer is, broadly put, the task of converting a text of one style (for example informal) into another (for example formal) while preserving its content. How can models be best trained for this task? What can be expected of a system performing text style transfer? And what does it mean to do it well, especially given the broad range of rewriting possibilities? In this talk I will present various strategies to model the task of style transfer under different conditions and I will discuss insights from both human and automatic evaluations. Chiefly, through the analysis of both modelling and evaluation and through engagement with audience, I will also reflect on the nature, the definition, and the the future of the task itself.

Henning Wachsmuth: Generation of Subjective Language. Chances and Risks

Research on natural language generation has made tremendous advances in the last years, due to powerful neural language models, such as BART, T5, and GPT-3. While generation technologies have been studied extensively for fact-oriented applications such as machine translation and customer service chatbots, they are recently also employed increasingly for creating and modifying subjective language – from the encoding of human beliefs in newly produced text to the debiasing of corpora and the transfer of subjective style characteristics of human-written texts. This brings up the question whether there are generation tasks that we should refrain from doing research on, due to the ethical issues they may entail. In this talk, I will give an overview of recent research on the generation of subjective language and present selected approaches in detail, covering the areas of computational argumentation, media framing, and social bias mitigation. On this basis, I will discuss both the chances for humans and society emerging from respective generation technologies and the ethical risks that come with their application. The interaction of chances and risks defines a red line that, I argue, should not be crossed without important reasons.

Table of Contents

Data Augmentation for Intent Classification of German Conversational Agents in the Finance Domain <i>Sophie Rentschler, Martin Riedl, Christian Stab and Martin Rückert</i>	1
MONAPipe: Modes of Narration and Attribution Pipeline for German Computational Literary Studies and Language Analysis in spaCy <i>Tillmann Dönicke, Florian Barth, Hanna Varachkina and Caroline Sporleder</i>	8
Lemma Hunting: Automatic Spelling Normalization for CMC Corpora <i>Eckhard Bick</i>	16
DocSCAN: Unsupervised Text Classification via Learning from Neighbors <i>Dominik Stammach and Elliott Ash</i>	21
Modelling Cultural and Socio-Economic Dimensions of Political Bias in German Tweets <i>Aishwarya Anegundi, Konstantin Schulz, Christian Rauh and Georg Rehm</i>	29
Adapting GermaNet for the Semantic Web <i>Claus Zinn, Marie Hinrichs and Erhard Hinrichs</i>	41
Assessing the Linguistic Complexity of German Abitur Texts from 1963–2013 <i>Noemi Kapusta, Marco Müller, Matilda Schauf, Isabell Siem and Stefanie Dipper</i>	48
Measuring Faithfulness of Abstractive Summaries <i>Tim Fischer, Steffen Remus and Chris Biemann</i>	63
Sentiment Analysis on Twitter for the Major German Parties during the 2021 German Federal Election <i>Thomas Schmidt, Jakob Fehle, Maximilian Weissenbacher, Jonathan Richter, Philipp Gottschalk and Christian Wolff</i>	74
Do gender neutral affixes naturally reduce gender bias in static word embeddings? <i>Jonas Wagner and Sina Zarriëß</i>	88
Improved Open Source Automatic Subtitling for Lecture Videos <i>Robert Geislinger, Benjamin Milde and Chris Biemann</i>	98
Constructing a Derivational Morphology Resource with Transformer Morpheme Segmentation <i>Łukasz Knigawka</i>	104
Improved Opinion Role Labelling in Parliamentary Debates <i>Laura Bamberg, Ines Rehbein and Simone Paolo Ponzetto</i>	110
ABSINTH : A small world approach to word sense induction <i>Victor Zimmermann and Maja Hoffmann</i>	121

This isn't the bias you're looking for: Implicit causality, names and gender in German language models	129
<i>Sina Zarrieß, Hannes Gröner, Torgrim Solstad and Oliver Bott</i>	
Evaluation of Automatic Speech Recognition for Conversational Speech in Dutch, English, and German: What Goes Missing?	135
<i>Alianda Lopez, Andreas Liesenfeld and Mark Dingemanse</i>	
Semantic Role Labeling for Sentiment Inference: A Case Study	144
<i>Manfred Klenner and Anne Göhring</i>	
Building an Extremely Low Resource Language to High Resource Language Machine Translation System from Scratch	150
<i>Flammie A Pirinen and Linda Wiecheteck</i>	
More Like This: Semantic Retrieval with Linguistic Information	156
<i>Steffen Remus, Saba Anwar, Gregor Wiedemann, Fynn Petersen-Frey, Seid Muhie Yimam and Chris Biemann</i>	
TopicShoal: Scaling Partisanship Using Semantic Search	167
<i>Sami Diaf and Ulrich Fritsche</i>	
Bye, Bye, Maintenance Work? Using Model Cloning to Approximate the Behavior of Legacy Tools	175
<i>Piush Aggarwal and Torsten Zesch</i>	