

Proceedings of the 15th Conference on Natural Language Processing

(KONVENS 2019)

October 9 – 11, 2019

German Society for Computational Linguistics & Language Technology
Friedrich-Alexander-Universität Erlangen-Nürnberg
Edited by the Chair of Computational Corpus Linguistics

Contents

Introduction	iii
Program Committee	iii
Acknowledgements	iv
Organizers	iv
KONVENS 2019 Papers	1
Long Papers	1
BERT for Named Entity Recognition in Contemporary and Historic German <i>Kai Labusch, Clemens Neudecker and David Zellhöfer</i>	1
A Supervised Learning Approach for the Extraction of Sources and Targets from German Text <i>Michael Wiegand, Margarita Chikobava and Josef Ruppenhofer</i>	10
A Descriptive Analysis of a German Corpus Annotated with Opinion Sources and Targets <i>Michael Wiegand, Leonie Lapp and Josef Ruppenhofer</i>	20
Automated Assessment of Language Proficiency on German Data <i>Edit Szügyi, Sören Etler, Andrew Beaton and Manfred Stede</i>	30
A Probabilistic Morphology Model for German Lemmatization <i>Christian Wartena</i>	40
To Act Or Not To Act - Annotating and Classifying Email Regarding Necessary Action <i>Veronika Hintzen and Alexander Fraser</i>	50
AkkuBohrHammer vs. AkkuBohrhammer: Experiments towards the Evaluation of Compound Splitting Tools for General Language and Specific Domains <i>Anna Häfty, Ulrich Heid, Anna Moskvina, Julia Bettinger, Michael Dorna and Sabine Schulte im Walde</i>	59
Neural classification with attention assessment of the implicit-association test OMT and prediction of subsequent academic success <i>Dirk Johannßen and Chris Biemann</i>	68
Towards Multimodal Emotion Recognition in German Speech Events in Cars using Transfer Learning <i>Deniz Cevher, Sebastian Zepf and Roman Klinger</i>	79
Visualising and evaluating the effects of combining active learning with word embedding features <i>Maria Skeppstedt, Rafal Rzepka, Kenji Araki and Andreas Kerren</i>	91
Creating Information-maximizing Natural Language Messages Through Image Captioning-Retrieval <i>Fabian Karl, Mikko Lauri and Chris Biemann</i>	101

German End-to-end Speech Recognition based on DeepSpeech <i>Aashish Agarwal and Torsten Zesch</i>	111
Label Propagation of Polarity Lexica on Word Vectors <i>Harald Koppen and Ritavan</i>	120
Detecting the boundaries of sentence-like units in spoken German <i>Josef Ruppenhofer and Ines Rehbein</i>	130
Dependency Trees for Greenlandic <i>Eckhard Bick</i>	140
Metaphor detection for German Poetry <i>Ines Reinig and Ines Rehbein</i>	149
Does BERT Make Any Sense? Interpretable Word Sense Disambiguation with Contextualized Embeddings <i>Gregor Wiedemann, Steffen Remus, Avi Chawla and Chris Biemann</i>	161
Determining Response-generating Contexts on Microblogging Platforms <i>Jennifer Fest, Arndt Heilmann, Oliver Hohlfeld, Stella Neumann, Jens Helge Reelfs, Marco Schmitt and Alina Vogelgesang</i>	171
Extraction and Classification of Speech, Thought, and Writing in German Narrative Texts <i>Luise Schricker, Manfred Stede and Peer Trilcke</i>	183
Argumentative Relation Classification as Plausibility Ranking <i>Juri Opitz</i>	193
Predicting Semantic Labels of Text Regions in Heterogeneous Document Images <i>Somtochukwu Enendu, Johannes Scholtes, Jeroen Smeets, Djoerd Hiemstra and Mariet Theune</i>	203
Evaluating Off-the-Shelf NLP Tools for German <i>Katrin Ortmann, Adam Roussel and Stefanie Dipper</i>	212
Short Papers	223
Towards a gold standard corpus for detecting valencies of Zulu verbs <i>Gertrud Faafß and Sonja Bosch</i>	223
“Konservenglück in Tiefkühl-Town” – Das Songkorpus als empirische Ressource interdisziplinärer Erforschung deutschsprachiger Poptexte <i>Roman Schneider</i>	229
Combining embedding methods for a word intrusion task <i>Finn Årup Nielsen and Lars Kai Hansen</i>	237
Deep learning for Free Indirect Representation <i>Annelen Brunner, Ngoc Duyen Tanja Tu, Lukas Weimer and Fotis Jannidis</i>	241
Representing document-level semantics of biomedical literature using pre-trained embedding models: Novel assessments <i>Jon Stevens, Brandon Punturo, Derek Chen, Mike Kim and Jacob Zimmer</i>	246

Deep-EOS: General-Purpose Neural Networks for Sentence Boundary Detection <i>Stefan Schweter and Sajawel Ahmed</i>	251
How Does Visual Complexity Influence Predictive Language Processing in a Situated Context? <i>Özge Alacam, Wolfgang Menzel and Tobias Staron</i>	256
Teacher-Student Learning Paradigm for Tri-training: An Efficient Method for Unlabeled Data Exploitation <i>Yash Bhalgat, Zhe Liu, Pritam Gundecha, Jalal Mahmud and Amita Misra</i>	262
Kaleidoscope Abstracts	267
Generic Web Content Extraction with Open-Source Software <i>Adrien Barbaresi</i>	267
Ein Tool zur Visualisierung des Gebrauchs von Schreibvarianten <i>Peter M. Fischer and Christian Lang</i>	269
Identification of Reading Absorption in User-Generated Book Reviews <i>Piroska Lendvai, Simone Rebora and Moniek Kuijpers</i>	271
Sketches of a graphical user interface for word alignment annotation <i>Maria Skeppstedt, Magnus Ahltorp, Gunnar Eriksson and Rickard Domeij</i>	273
Predicting default and non-default aspectual coding: Impact and density of information features <i>Michael Richter and Tariq Yousef</i>	275
A New German Reddit Corpus <i>Andreas Blombach, Natalie Dykes, Stefan Evert, Philipp Heinrich, Besim Kabashi and Thomas Proisl</i>	278
GermEval 2019 Papers	280
Task 1 Papers	280
GermEval 2019 Task 1: Hierarchical Classification of Blurbs <i>Steffen Remus, Rami Aly and Chris Biemann</i>	280
Multi-Label Classification of Blurbs with SVM Classifier Chains <i>Franz Bellmann, Lea Bunzel, Christoph Demus, Lisa Fellendorf, Olivia Gräupner, Qiuyi Hu, Tamara Lange, Alica Stuhr, Jian Xi, Dirk Labudde and Michael Spranger</i>	293
Multi-Label Multi-Class Hierarchical Classification using Convolutional Seq2Seq <i>Venkatesh Umaashankar and Girish Shanmugam S</i>	300
Enriching BERT with Knowledge Graph Embeddings for Document Classifica- tion <i>Malte Ostendorff, Peter Bourgonje, Maria Berger, Julian Moreno-Schneider, Georg Rehm and Bela Gipp</i>	307

COMTRAVO-DS team at GermEval 2019 Task 1 on Hierarchical Classification of Blurbs	
<i>David S. Batista and Matti Lyra</i>	315
Convolutional Neural Networks for Classification of German Blurbs	
<i>Erdan Genc, Louay Abdalgawad, Viorel Morari and Peter Kluegl</i>	323
TwistBytes - Hierarchical Classification at GermEval 2019: walking the fine line (of recall and precision)	
<i>Fernando Benites</i>	328
The HUIU Contribution to the GermEval 2019 Shared Task 1	
<i>Melanie Andresen, Melitta Gillmann, Jowita Grala, Sarah Jablotschkin, Lea Roseler, Eleonore Schmitt, Lena Schnee, Katharina Straka, Michael Vauth, Sandra Kübler, Heike Zinsmeister</i>	336
Label Frequency Transformation for Multi-Label Multi-Class Text Classification	
<i>Raghavan A K, Venkatesh Umaashankar and Gautham Krishna Gudur</i>	341
Logistic Regression and Naive Bayes for Hierarchical Multi-label Classification at GermEval 2019 - Task 1	
<i>Kristian Rother and Achim Rettberg</i>	347
Task 2 Papers	354
Overview of GermEval Task 2, 2019 Shared Task on the Identification of Offensive Language	
<i>Julia Maria Struß, Melanie Siegel, Josep Ruppenhofer, Michael Wiegand and Manfred Klenner</i>	354
InriaFBK Drawing Attention to Offensive Language at Germeval2019	
<i>Michele Corazza, Stefano Menini, Elena Cabrio, Sara Tonelli and Serena Villata</i>	366
German Hatespeech classification with Naive Bayes and Logistic Regression - hshl at GermEval 2019 - Task 2	
<i>Kristian Rother and Achim Rettberg</i>	372
FraunhoferSIT at GermEval 2019: Can Machines Distinguish Between Offensive Language and Hate Speech? Towards a Fine-Grained Classification	
<i>Inna Vogel and Roey Regev</i>	377
FoSIL - Offensive language classification of German tweets combining SVMs and deep learning techniques	
<i>Florian Schmid, Justine Thielemann, Anna Mantwill, Jian Xi, Dirk Labudde and Michael Spranger</i>	382
2019 GermEval Shared Task on Offensive Tweet Detection h_da submission	
<i>Isabell Börner, Midhad Blazevic, Maximilian Komander and Margot Mieskes</i>	387

HAU at the GermEval 2019 Shared Task on the Identification of Offensive Language in Microposts: System Description of Word List, Statistical and Hybrid Approaches	
<i>Johannes Schäfer, Tom De Smedt and Sylvia Jaki</i>	391
UPB at GermEval-2019 Task 2: BERT-Based Offensive Language Classification of German Tweets	
<i>Andrei Paraschiv and Dumitru-Clementin Cercel</i>	398
hpiDEDIS at GermEval 2019: Offensive Language Identification using a German BERT model	
<i>Julian Risch, Anke Stoll, Marc Ziegele and Ralf Krestel</i>	405
The HUIU Contribution for the GermEval Shared Task 2	
<i>Melanie Andresen, Melitta Gillmann, Jowita Grala, Sarah Jablotschkin, Lea Röseler, Eleonore Schmitt, Lena Schnee, Katharina Straka, Michael Vauth, Sandra Kübler and Heike Zinsmeister</i>	411
TUWienKBS19 at GermEval Task 2, 2019: Ensemble Learning for German Offensive Language Detection	
<i>Joaquín Padilla Montani and Peter Schüller</i>	418
RGCL at GermEval 2019: Offensive Language Detection with Deep Learning	
<i>Alistair Plum, Tharindu Ranasinghe, Constantin Orăsan, Ruslan Mitkov</i>	423
FKIE - Offensive Language Detection on Twitter at GermEval 2019	
<i>Theresa Krumbiegel</i>	429
bertZH at GermEval 2019: Fine-Grained Classification of German Offensive Language using Fine-Tuned BERT	
<i>Tim Graf and Luca Salini</i>	434

Introduction

The Conference on Natural Language Processing (“Konferenz zur Verarbeitung natürlicher Sprache”, KONVENS) aims at offering a broad perspective on current research and developments within the interdisciplinary field of natural language processing. It allows researchers from all disciplines relevant to this field of research to present their work.

The 15th KONVENS “Bridging the gap between NLP and human understanding” took place October 9–11, 2019, at Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany. It was the first KONVENS in an annual schedule, following KONVENS 2018 in Vienna, Austria, and preceding KONVENS 2020 in Zürich, Switzerland. KONVENS 2019 was collocated with the GermEval Workshop and a Statistics Tutorial.

The conference was organized by the Chair of Computational Corpus Linguistics led by Prof. Dr. Stefan Evert.

Program Committee

Barbaresi, Adrien	Pado, Sebastian
Biemann, Chris	Pinkal, Manfred
Bollmann, Marcel	Pirker, Hannes
Buchberger, Ernst	Quasthoff, Uwe
Butt, Miriam	Rehbein, Ines
Clematide, Simon	Rehm, Georg
Crysmann, Berthold	Ruppenhofer, Josef
De William, Ernesto Luca	Sasaki, Felix
Dipper, Stefanie	Schäfer, Roland
Heid, Ulrich	Scherrer, Yves
Heiden, Serge	Schmid, Helmut
Klenner, Manfred	Schmidt, Thomas
Klinger, Roman	Schulte im Walde, Sabine
Kordoni, Valia	Stede, Manfred
Krenn, Brigitte	Tanguy, Ludovic
Kübler, Sandra	Versley, Yannick
Lapshinova-Koltunski, Ekaterina	Weskott, Thomas
Mehler, Alexander	Wolska, Magdalena
Menzel, Wolfgang	Zesch, Torsten
Nakov, Preslav	Zinsmeister, Heike
Neumann, Günter	

Acknowledgements

Invited speakers:

Gemma Boleda (Universitat Pompeu Fabra, Barcelona): Computational linguistics and linguistic theory

Daisuke Bekki (Ochanomizu University) and Hitomi Yanaka (RIKEN Center for Advanced Intelligence Project): Hybrid natural language understanding: neural network, logic and beyond

Special thanks:

Prof. Dr. Günther Görz, FAU Erlangen-Nürnberg

Prof. Dr. Joachim Hornegger, President of FAU Erlangen-Nürnberg

Prof. Dr. Rainer Trinczek, Dean of the Faculty of Humanities, Social Sciences, and Theology

Financial support:

Dr. German Schweiger-Stiftung

Luise Prell Stiftung

Gastprofessorenprogramm der FAU, gefördert durch das Bayerische Staatsministerium für Wissenschaft und Kunst

Organizers

Blombach, Andreas

Dykes, Natalie

Evert, Stefan

Greiner, Paul

Griebel, Tim

Heinrich, Philipp

Kabashi, Besim

Proisl, Thomas

Schorr, Tanja