Welcome to KONVENS 2016!

We are happy to welcome you at Ruhr-Universität Bochum for this year's Conference on Natural Language Processing (KONVENS) and hope you enjoy your stay! On the following pages you'll find the detailed program for the conference along with some practical information.

Contact us

In case you ever need to contact the conference organizers while away from the venue, we provide a special mobile phone number for you to call or send an SMS to:

+49 (0)1577 2384342

Wi-fi

There is an eduroam signal available in all rooms. If you are not an eduroam user, you can ask for a wi-fi login and password at the registration desk.

Proceedings

You can find the conference proceedings online at:

https://www.linguistics.rub.de/konvens16/proceedings/
Floor plan

Exit to U35 station "Lennershof"

Cafeteria
(drinks and snacks)

Conference Office
Coffee Breaks/Poster Sessions

Presentations

Restrooms

Exit to campus
(dining hall, botanical gardens, etc.)

Note: The cafeteria is open from 7:45 to 16:00 on all conference days.
Detailed Program

Registration opens at 9:00 on all days. Please try to arrive as early as possible so that everyone has enough time to register!

Monday, 19.09.

9:45 – 10:00 Opening

10:00 – 11:00 Invited Talk: Barbara Plank
What to do about non-standard (or non-canonical) language in NLP

11:00 – 11:30 Coffee Break

11:30 – 12:00 Gerold Schneider, Marianne Hundt and Rahel Oppliger
Part-Of-Speech in Historical Corpora: Tagger Evaluation and Ensemble Systems on ARCHER

12:00 – 12:30 Martin Volk, Chantal Amrhein, Noëmi Aepli, Mathias Müller and Phillip Ströbel
Building a Parallel Corpus on the World’s Oldest Banking Magazine

12:30 – 13:00 Fabian Barteld, Ingrid Schröder and Heike Zinsmeister
text-gamma - Inter-annotator agreement for categorization with simultaenous segmentation and transcription-error correction

13:00 – 14:30 Lunch Break

14:30 – 15:00 Maud Ehrmann, Giovanni Colavizza, Yannick Rochat and Frederic Kaplan
Le Temps: diachronic evaluation of NER systems on old newspapers

15:00 – 15:30 Dietmar Schabus, Brigitte Krenn and Friedrich Neubarth
Data-Driven Identification of Dialogue Acts in Chat Messages

15:30 – 16:00 Michael Wojatzki and Torsten Zesch
Stance-based Argument Mining - Modeling Implicit Argumentation Using Stance

16:00 – 16:30 Coffee Break

16:30 – 17:15 GSCL Talk: Doctoral Thesis Award

17:30 – 19:00 GSCL Business Meeting
Tuesday, 20.09.

9:30 – 10:00  Josef Ruppenhofer and Jasper Brandes  
Verifying the robustness of opinion inference

10:00 – 10:30  José Manuel Martínez Martínez, Ekaterina Lapshinova-Koltunski and Kerstin Anna Kunz 
Annotation of Lexical Cohesion in English and German: Automatic and Manual Procedures

10:30 – 11:00  Fabian Simonjetz and Adam Roussel 
Crosslinguistic Annotation of German and English Shell Noun Complexes

11:00 – 11:30  Coffee Break

11:30 – 12:30  Poster Session I
   Daniel Dakota  
Brown clustering for unlexicalized parsing
   Martin Volk, Simon Clematide, Johannes Graën and Phillip Ströbel  
Bi-particle Adverbs, PoS-Tagging and the Recognition of German Separable Prefix Verbs
   Reto Baumgartner  
Morphological analysis and lemmatization for Swiss German using weighted transducers
   Timo Gühring, Nicklas Linz, Rafael Theis and Annemarie Friedrich  
SWAN: an easy-to-use web-based annotation system
   Tobias Roth  
Isolation of Place-Name Forms in Toponymic Data
   Michael Percillier  
Verb lemmatization and semantic verb classes in a Middle English corpus
   Carlota de Benito, Javier Pueyo and Inés Fernández-Ordóñez  
Creating and designing a corpus of rural Spanish
   Dolores Batinić, Sandra Birzer and Heike Zinsmeister  
Creating an extensible, levelled study corpus for learners of Russian
   Tatjana Scheffler and Manfred Stede  
Mapping PDTB-style connective annotation to RST-style discourse annotation
   Polina Panicheva and Olga Mitrofanova  
Developing a Toolkit for Distributional Analysis of Abnormal Collocations in Russian

12:30 – 14:00  Lunch Break

14:00 – 14:30  Christine Köhn, Tobias Staron and Arne Köhn  
Parsing Free-Form Language Learner Data: Current State and Error Analysis

14:30 – 15:00  Sviatlana Höhn, Alain Pfeiffer and Eric Ras  
Challenges of error annotation in native/non-native speaker chat

15:00 – 15:30  Kay Berkling  
Item Presentation in Primers - An Analysis Based on Acquisition Research

17:00 – 18:30  Social Event: German Mining Museum

19:00 –  Conference Dinner at Strätlingshof
Wednesday, 21.09.

9:30 – 10:30  **Invited Talk:**  **John Nerbonne**  
Data from Non-standard Varieties

10:30 – 11:00  Coffee Break

11:00 – 12:00  **Poster Session II**

  * Sebastian Pado, Jan Snojder, Jason Utt and Britta Zeller*  
  Smoothing Syntax-Based Semantic Spaces: Let The Winner Take It All

  * Eva Horch and Ingo Reich*  
  On “Article Omission” in German and the “Uniform Information Density Hypothesis”

  * Rahel Oppliger*  
  Automatic authorship attribution based on character n-grams in Swiss German

  * Adrien Barbaresi*  
  Bootstrapped OCR error detection for a less-resourced language variant

  * Martin Riedl, Tim Feuerbach and Chris Biemann*  
  Running into Brick Walls Attempting to Improve a Simple Unsupervised Parser

  * Kerstin Eckart and Markus Gärtner*  
  Creating Silver Standard Annotations for a Corpus of Non-Standard Data

  * Mirela-Stefania Duma and Wolfgang Menzel*  
  Paragraph Vector for Data Selection in Statistical Machine Translation

  * Harald Lüngen, Michael Beißwenger, Axel Herold, Eric Ehrhardt and Angelika Storrer*  
  Integrating corpora of computer-mediated communication in CLARIN-D: Results from the curation project ChatCorpus2CLARIN

  * Gerhard Jaeger and Pavel Sofroniev*  
  Automatic cognate classification with a Support Vector Machine

  * Daniel Weber and Desislava Zhekova*  
  TweetNorm: Text Normalization on Italian Twitter Data

12:00 – 13:30  Lunch Break

13:30 – 14:00  Yves Scherrer and Nikola Ljubešić  
Automatic normalisation of the Swiss German ArchiMob corpus using character-level machine translation

14:00 – 14:30  Simon Clematide, Karina Frick, Noëmi Aepli and Jean-Phillippe Goldman  
Crowdsourcing Swiss Dialect Transcriptions for Assessing Factors in Writing Variations

14:30 – 15:00  Nikola Ljubešić, Katja Zupan, Darja Fišer and Tomaž Erjavec  
Normalising Slovene data: historical texts vs. user-generated content

15:00 – 15:30  Coffee Break

15:30 – 16:00  Julia Suter, Sarah Ebging and Martin Volk  
Rule-based Automatic Text Simplification for German

16:00 – 16:30  Inès Zribi, Inès Kammoun, Mariem Ellouze, Lamia Hadrich Belguith and Phillippe Blache  
Sentence boundary detection for transcribed Tunisian Arabic

16:30 – 17:00  Alexander Panchenko, Johannes Simon, Martin Riedl and Chris Biemann  
Noun Sense Induction and Disambiguation using Graph-Based Distributional Semantics
Thursday, 22.09.

All workshops/tutorials take place in the same area as the main conference.

IGGSA Workshop on German Sentiment Analysis 2016

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30 – 11:00</td>
<td>Tutorial on Sentiment Analysis (Melanie Siegel)</td>
</tr>
<tr>
<td>11:00 – 11:30</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>11:30 – 13:00</td>
<td>Tutorial – Part II (Melanie Siegel)</td>
</tr>
<tr>
<td>13:00 – 14:30</td>
<td>Lunch Break</td>
</tr>
<tr>
<td>14:30 – 15:00</td>
<td>Workshop on the IGGSA 2016 Shared Task</td>
</tr>
<tr>
<td>14:30 – 15:00</td>
<td>Overview of the IGGSA 2016 Shared Task on Source and Target Extraction from Political Speeches (Josef Ruppenhofer, Julia Maria Struß)</td>
</tr>
<tr>
<td>15:00 – 15:30</td>
<td>Presentation of Participants from Potsdam University (Leonard Kriese)</td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td>Presentation of Participants from Saarland University (Michael Wiegand)</td>
</tr>
<tr>
<td>16:00 – 16:30</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>16:30 – 17:15</td>
<td>Feedback Session</td>
</tr>
</tbody>
</table>

**Tutorial**

With the Web 2.0 most consumer products, cultural and political events are discussed and evaluated in internet forums. These expressions of opinion contain valuable information for companies, governments and other actors: Information about what users and citizens think, where they have difficulties, and how they solve their problems. The sentiment expressions are accessible, but the effort required to read and evaluate them manually is excessive. Opinion mining automatically analyzes and classifies sentiment expressions from publicly available sources. For this task, it is not enough to extract keywords from review texts. More intelligent NLP methods are needed to detect negation scope and arguments of opinion words. Therefore, opinion mining is an interesting research topic for language technologists.

Research on opinion mining so far focused mainly on the English language. Some of the approaches can be directly transferred to the analysis of German sentiment expressions, but many other cannot.

In the tutorial we examine standard approaches to opinion mining. Further, we look at German sentiment expressions and research approaches to handle these.¹ We will take into account and introduce existing resources, such as word lists and corpora.²

The tutorial is aimed at students and researchers with a basic background in NLP and interest in processing German language.

**Instructor**

- Melanie Siegel (Darmstadt University of Applied Sciences)

¹https://sites.google.com/site/iggsahome/publications
²https://sites.google.com/site/iggsahome/downloads
Shared Task

The shared task we ran this summer was a follow up event to the IGGSA-STEMPS: Source, Subjective Expression and Target Extraction from Political Speeches. The first iteration was run in 2014 as part of the GESTALT: GERman Sentiment Analysis shared Task organized as a KONVENS-2014 workshop in Hildesheim, Germany. Our shared task deals with a fine-grained subtask in sentiment analysis: opinion holder and target extraction on political speeches. The 2014 edition of the shared task introduced the first gold standard data of its kind sampled from the publicly available corpus comprising Swiss parliamentary speeches. The task in 2016 was to build on and extend the pilot task held in 2014. The test data for the second iteration of the task contained new unseen data sampled from Swiss parliamentary speeches. The 2014-pilot task saw only one participating system. Therefore, considerable efforts have been made to introduce the following changes so as to facilitate participation in this year’s edition of the shared task:

- Annotation guidelines were revised. In order to make the shared task more accessible, the guidelines were simplified by dropping several linguistic aspects that were felt to be too complex to solve as part of this shared task (e.g. polar facts or compound analysis).

- A sample system developed by participants of the 2014 edition (compliant with the 2016 edition's format) was made publicly available; it is a lexicon-based system which allows users to flexibly test their newly developed opinion role lexicons: https://github.com/miwieg/german-opinion-role-extractor

- An adjudicated version of the 2014 test data (605 sentences with fine-grained annotation) was made available as training data. Thus, unlike in 2014, participants were able to employ supervised learning for system implementation. The quality of the adjudicated gold standard has notably improved of the original test data released in 2014. Observed inter-annotator token-based kappa values in 2014 were 0.39, 0.57, and 0.46 for subjective expressions, sources and targets. On the new test set, the corresponding token-based kappa values are 0.82, 0.89 and 0.72.

Organizers

- Josef Ruppenhofer (IDS Mannheim)
- Julia Maria Struß (University of Hildesheim)
- Michael Wiegand (Saarland University)

Further Information

- http://iggsasharedtask2016.github.io/
- http://sites.google.com/site/iggsahome/

3Josef Ruppenhofer, Julia Maria Struß, Jonathan Sonntag, Stefan Gindl. IGGSA-STEMPS: Shared Task on Source and Target Extraction from Political Speeches, JLCL, Volume 29 (1), 33–46, 2014.
4https://sites.google.com/site/iggsasharedtask/task-1
NLP4CMC III: 3rd Workshop on Natural Language Processing for Computer-Mediated Communication

9:30 – 10:00  Holger Grumt Suárez, Natali Karlova-Bourbonus and Henning Lobin
Semi-automated TEI representation of a discourse-structured blog corpus for German

10:00 – 10:30  Björn Ross, Michael Rist, Guillermo Carbonell, Benjamin Cabrera, Nils Kurowsky and Michael Wojatzki
Measuring the Reliability of Hate Speech Annotations: The Case of the European Refugee Crisis

10:30 – 11:00  Thierry Declerck
Towards the Harmonization and Segmentation of German Hashtags

11:00 – 11:30  Coffee Break

11:30 – 13:00  Round table on future perspectives in NLP4CMC
(based on the results of the EmpiriST shared task on the automatic linguistic annotation of computer-mediated communication and web corpora)

Over the past decade, there has been a growing interest in collecting, processing and analyzing data from genres of social media and computer-mediated communication (CMC) and social media interactions such as chats, blogs, forums, tweets, newsgroups, messaging applications (SMS, WhatsApp), interactions on “social network” sites and on wiki talk pages: As part of large corpora which crawled from the web, CMC data are often regarded as an unloved “bycatch” that proves for linguistic annotation by means of standard natural language processing (NLP) tools that are optimized for edited text; on the other hand, the existence of CMC data in web corpora is relevant for all research and application contexts which require data sets that represent the full diversity of genres and linguistic variation on the web. For corpus-based variational linguistics, CMC discourse is an important resource that closes the “CMC gap” in corpora of contemporary written language and language-in-interaction. With a considerable part of contemporary everyday communication being mediated through CMC technologies, up-to-date investigations of language change and linguistic variation need to be able to include CMC discourse in their empirical analyses.

The workshop provides a platform for the presentation of results and the discussion of ongoing work in adapting NLP tools for processing data from genres of computer-mediated communication (CMC) and in using NLP solutions for building and annotating social media corpora. It is organized by the special interest group “social media / internet-based communication” of the German Society for Language Technology and Computational Linguistics (GSCL). Previous workshops have been held at KONVENS 2014 and as part of the GSCL Conference 2015.

Organizers

- Michael Beißwenger (University of Duisburg-Essen)
- Michael Wojatzki (University of Duisburg-Essen)
- Torsten Zesch (University of Duisburg-Essen)

Further Information

↗️ https://sites.google.com/site/nlp4cmc2016/
Tutorial: Visual Analytics for Computational Linguistics

9:30 – 11:00  General introduction to LingVis. The goals and principles of information visualization (visual variables and interaction techniques) and their cognitive foundations. Demonstrations of current LingVis visualizations.

11:00 – 11:30  Coffee Break

11:30 – 13:00  How to choose a visualization. Survey of existing visualizations relevant for LingVis, with critical analysis by instructors and students.

13:00 – 14:30  Lunch Break

14:30 – 16:00  Hands-On: Repurposing visualizations. How to adapt and modify existing visualizations. Introduction of a sample practice problem with concrete data sets.

16:00 – 16:30  Coffee Break

16:30 – 18:00  Spill-over and conclusion. Discussion of the practice problem, the visualizations and results. Discussion of future directions for LingVis.

The aim of this course is to introduce students to the emerging field of the visualization of linguistic information (LingVis) which combines techniques developed in the fields of Information Visualization (InfoVis) and Visual Analytics with methodology and analyses from theoretical and computational linguistics. As a specialized subfield of information visualization, the visualization of language continues to face particular challenges: Language data is complex, only partly structured and, as with today’s language resources, comes in large quantities. The overall challenge lies in breaking down the multidimensionality into intuitive visual variables that enable an at-a-glance overview of the data and the patterns underlying it. In this course, we aim to introduce students to the basic principles of InfoVis and present concrete use cases of LingVis in both synchronic and diachronic dimensions. A part of the course will include hands-on sessions in which students can experiment with pre-prepared data sets and LingVis software in order to investigate how complex linguistic questions can profit from visual analysis.

Expected level and prerequisites  Introductory. Basic knowledge of linguistics is required. Elementary programming experience will be helpful but is not a priori required.

Instructors

- Annette Hautli-Janisz (University of Konstanz)
- Dominik Sacha (University of Konstanz)

Further Information

http://ling.uni-konstanz.de/pages/home/hautli/teaching/clvis-konvens2016.html
Tutorial: Symbolic Distributional Semantics with the JoBimText Framework

14:30 – 15:30  Methods and Applications in JoBimText
15:30 – 16:00  Access to Semantic Models
16:00 – 16:30  Coffee Break
16:30 – 18:00  Computing Semantic Models

This tutorial gives a theoretical and a hands-on introduction to a symbolic Distributional Semantics framework. The framework contains tools to compute semantic models from the input text, which include multi-word expressions and hypernymies. Furthermore, it provides means for word sense induction and labelling of word senses with extracted hypernym terms. Due to efficient pruning strategies and the usage of Hadoop's MapReduce framework our framework scales to arbitrary large data, e.g. web corpora. Its symbolic approach to similarity computation – instead of numeric vector based approaches like LSA or embeddings – makes the models interpretable.

Target audience  This tutorial targets researchers who want a broad introduction into the field of Distributional Semantics. As JoBimText is a Hadoop application, the tutorial will provide an opportunity to learn about large data processing tools. Furthermore, the target audience will use the framework. This includes how to use JoBimText to compute new models. A main focus will be also on the usage of the JoBimText API to access models. That way, the tutorial is aimed at researchers who want to learn about Distributional Semantics and apply semantic methods into their applications by using an API.

The audience should bring their laptops with Java installed, so that they can try out the methods in practice. If they also want to compute models we recommend having the provided virtual machine (see “Resources” on the website below) installed. For accessing JoBimText models, we will provide an Eclipse project with the necessary libraries provided for the usage.

Instructors
• Martin Riedl (Technische Universität Darmstadt)
• Eugen Ruppert (Technische Universität Darmstadt)

Further Information

https://sites.google.com/site/konvens2016jobintexttutorial/
Social event and dinner

If you are participating in both the social event and conference dinner, you can join a group from the conference venue on Tuesday afternoon to guide you to both locations. If you prefer to go there on your own, you can find travel instructions below.

German Mining Museum

We offer guided tours through the German Mining Museum (available in both German and English), which will lead you through the museum’s underground visitor mine, containing faithfully reconstructed parts of a real mine. We will also visit the viewing platform on the mine's headframe, which provides a fantastic view of Bochum and the Ruhr area.

**When**  Tuesday, 17:00–18:30
**Where**  Am Bergbaumuseum 28, 44791 Bochum

**How to get there**

1. From either the university campus or Bochum Hbf, take the U35 in direction “Riemke Markt” or “Herne / Schloß Strünkede”.
2. Exit at “Deutsches Bergbau-Museum” – the museum is nearby.

**Note**  The average temperature in the mine is only 12°C, so dress warmly!

Strätlingshof

Right after the social event, there will be a conference dinner at Strätlingshof.

**When**  Tuesday, 19:00–
**Where**  Altenbochumer Str. 64, 44803 Bochum

**How to get there**

1. Go to Bochum Hbf, then (from the underground platforms) take either
   - tram 302 in direction “Laer Mitte”, or
   - tram 310 in direction “Witten / Heven Dorf”.
2. Exit at the second stop, “Freigrafendamm”.
3. Walk down the Freigrafendamm (the tree-lined avenue – you can’t miss it!) and take the **second street to the left** (Plüttmannsweg).
4. The restaurant is the last building on the left side.

Alternatively, you can also take the bus 368 to the bus stop “Wirmerstr.”, which is right in front of the restaurant. We recommend you check your individual route with the itinerary planner at http://efa.vrr.de/. However, trams go much more frequently and are therefore the recommended option.
Where to eat

On/near the campus

- **Dining Hall (Mensa).** In the center of the RUB campus, located behind the main lecture hall (Audi-max). The place closest to the conference venue. We will have helpers to guide you there during the lunch breaks. **Only for lunch.**

- **Uni-Center.** Opposite the campus, across the “Ruhr-Universität” station, the Uni-Center contains several places to eat, such as:
  - **Summa cum laude.** Querenburger Höhe 267, 44801 Bochum. [http://www.summacumlaude-bochum.de](http://www.summacumlaude-bochum.de)
  - **Trattoria Falcone.** Querenburger Höhe 282, 44801 Bochum.
  - **Various food stalls** selling fast food, pizza, kebab, etc.

- **Q-West.** [http://q-we.st/](http://q-we.st/)
  On the opposite side of the campus, towards the “N” row of buildings.
  **Important:** During lunch time (11:30–14:30), you **cannot** pay with cash or debit/credit cards – you have to use a university-issued card (guest cards are available from a vending machine) and load it with credit. (On the plus side, you can also re-use the same card in the Mensa and the cafeterias!) Outside of lunch time, you can pay normally here.

- **Le Clochard.** Buscheyplatz 2, 44801 Bochum. [http://www.leclochard.de/](http://www.leclochard.de/)
  Near the Uni-Center and the “Lennershof” station. **Not for lunch – opens at 16:00.**

In the city center

If you’re in the city center, we recommend that you simply go to **Bermuda3Eck** (“bermuda triangle”), Bochum’s renowned quarter of bars and restaurants. It stretches between **Südring** and **Konrad-Adenauer-Platz**, is a short walking distance from the main station (Bochum Hbf), and contains more than 50 bars, cafés, fast food joints and restaurants of all kinds (Italian, Spanish, Indian, Japanese, Vietnamese and more).

Also recommended if you just want to grab a drink after the conference!

- **http://www.bermuda3eck.de/**
C: Main conference venue (ID building)
1: Dining hall (Mensa) – our recommended lunch option
2: Q-West (on-campus restaurant) – careful: during lunch time (11:30–14:30) you need a university-issued card (available from vending machines) to pay here!
3: Uni-Center – restaurants, food stalls, supermarkets, pharmacy, ATMs, …
4: Le Clochard (restaurant) – opens at 16:00
5: Botanical garden – if the weather is nice and you have some time to spare…

You can also find a custom map on Google Maps here…

https://www.linguistics.rub.de/konvens16/map/
Other things to do

• **Botanical Garden of the Ruhr-University.** Located on the university campus, just 350 m away from the conference venue (cf. the supplied area map), the botanical garden is always worth a visit if the weather is nice and you have some time to spare. It features an outside area of 130,000 m², a Chinese garden, and several different greenhouses.
  * Open 9:00–18:00 (greenhouses close at 17:00)
  * Entry is free.
  * http://www.boga.ruhr-uni-bochum.de/ (German only)

• **Art Collections of the Ruhr-University.** Right in the center of the campus, opposite the main lecture hall (Audimax), the art collections are comprised of various museums featuring antiquities (Antikenmuseum) as well as modern and contemporary art (Sammlung Moderne).
  * Open 11:00–17:00 (closed on Mondays)
  * Entry is free.
  * http://www.ruhr-uni-bochum.de/kusa/ (Antikenmuseum; German only)
  * http://kusa-rub-moderne.de/m254 (Sammlung Moderne; German only)

• **Lake Kemnade.** Another nice location during sunny weather, the Lake Kemnade offers plenty of space to walk, sit and relax, or enjoy drinks and cocktails at the beach bar “StrandDeck” (open until 23:00). You can reach the lake with bus lines from either Ruhr-Universität or Bochum Hbf (recommended bus stop: “Witten / Freizeitbad Heveney”), but buses only go rarely (especially in the evenings!), so we strongly recommend you plan your trip in advance.

But wait, there’s more!

For even more things to do in Bochum (including brewery tours, another art museum, a railway museum, industrial monuments, etc.), take a look at the list of suggestions compiled by Ruhr Tourismus: