Conference Agenda
34th Tagung der Computerlinguistik-Studierenden
Ruhr University in Bochum

Thursday, May 22

8:45 - 9:30	GB 5/160	Reception and Attendees registration		
9:30 - 10:00		Opening word		
10:00 - 11:30		Keynote presentation by Dr. Ronja Laarmann-Quante: What computational linguists can do with spelling errors		
11:30 - 13:00	Campus navigation and Lunch			
13:00 - 13:40	GB 5/160	Training for the unexpected – approaching universal phone recognition with whisper (Jacob Suchardt)	**Workshop** Typst: The new LaTeX? (Jakob Moser)	GB 03/149
13:50 - 14:20		The language of the weather forecast on the British Isles – a standardized variety? (Katharina C. Hummel)	Automated scoring of a German written elicited imitation test (Mihail Chifligarov)	
14:30 - 15:00		Interpreting prosody: variation in theoretical assumptions and its impact on computational models (Liliia Maiorova)		
15:00 - 16:00	Coffee break			
16:00 - 17:30	HGB 30	Keynote presentation by Prof. Dr. Stefanie Dipper: Juice of life, power current and the Elijah two-act play – annotation and analysis of metaphors in German sermons		
19:00 - 21:45	GB 5/160	✨Game Night✨ Unwind with drinks, snacks, casual chat, and board games – bring your favourites or enjoy ours!		

Friday, May 23

<table>
<tr><td>9:45 - 10:00</td><td rowspan="2">GB 5/160</td><td colspan="3">Briefing</td></tr>
<tr><td>10:00 - 11:30</td><td colspan="3">Keynote presentation by Prof. Dr. Sina Zarrieß:

Challenging large and small language models with simple linguistic tasks</td></tr>
<tr><td>11:40 - 12:10</td><td rowspan="2">GB 5/160</td><td>Analyse akademischer Sprache mit Hilfe von Frequent Frames
(Ilka Plesse)</td><td>Comparing masked language model scoring metrics
(Nele Mastracchio and Anna Stein)</td><td rowspan="2">GB 03/149</td></tr>
<tr><td>12:20 - 12:50</td><td>Trier Digital -
Wandel sichtbar machen
(Carina Zander)</td><td>What the Heck is Quantum NLP?
(Adrian Mülthaler)</td></tr>
<tr><td>12:50 - 14:00</td><td colspan="4">Lunch</td></tr>
<tr><td>14:00 - 15:30</td><td>GB 5/160</td><td colspan="3">Keynote presentation by Prof. Dr. Torsten Zesch:

Automatic Analysis of Learner Language to Support Learners</td></tr>
<tr><td>15:40 - 16:10</td><td>GB 5/160</td><td>Klassifizierung der Lieder aus dem Allgemeinen Deutschen Kommersbuch nach ihren Kategorien
(Nune Arazyan)</td><td>Question Authority:
The Gold Standard
(Jakob Schuster)</td><td>GB 03/149</td></tr>
<tr><td>18:00 - 21:45</td><td>MC OpenSpace</td><td colspan="3">✨Quiz’ n’ chat Night✨

Join us for a relaxing mingle, find yourself a team and challenge your wits together!
The Quiz begins at 18:30.</td></tr>
</table>

Saturday, May 24
all day – in the GABF 04/411 lecture hall

9:45 - 10:00	Briefing
10:00 - 11:30	Keynote presentation by Prof. Dr. Bilal Zafar: On Trustworthiness of LLMs
11:40 - 12:10	**Lightning talks:** ❖ Bridging laughter and language models (Ariana Acosta) ❖ Efficient knowledge injection for extreme multi-label classification in the library domain (Katja Konermann) ❖ Five things they don't want you to know about pre-training language models (Katja Konermann and Anina Klaus)
12:10 - 13:00	Lunch
13:00 - 14:30	Career talk with RUB Alumni
14:40 - 15:00	**Lightning talks:** ❖ Building and analysing a corpus of right-wing extremist German rock music (Carlotta Schneeberger) ❖ Manual implementation of an automated scoring of score 3 error types of a German written elicited imitation test (Jammila Laâguidi)
15:00 - 15:20	Coffee break
15:20 - 15:50	**Lightning talks:** ❖ OCR in times of LLMs (Luca Scavone and Aaron Henning) ❖ "Adidas is known for their rugged and durable designs, while Nike's designs are often criticized for their lack of practicality in harsh conditions" - how to manipulate LLM opinions through data poisoning? (Valentin Höpfl) ❖ Changing the Emotions of a Beatles' Song (Misha Sonkin)
15:50 - 16:50	Closing word and Selection of the next year's host

Presentation abstracts:

Ilka Plesse
Analyse akademischer Sprache mit Hilfe von Frequent Frames

Frequent Frames bestehen in ihrer einfachsten Form aus zwei häufig zusammen vorkommenden Wörtern, zwischen denen sich ein Zielwort befindet. In diesem Vortrag wird untersucht, wie sich Frequent Frames nutzen lassen, um das Register der akademischen Sprache in Bezug auf Syntax und Phraseologie zu charakterisieren. Die Ergebnisse, die auf der Grundlage verschiedener Korpora gewonnen wurden werden für die Analyse verglichen. Es werden Korpora, welche aus Abstracts akademischer Journalbeiträge bestehen, mit Ergebnissen aus Korpora von Nachrichtenartikeln verglichen, die sich an eine breite Bevölkerung richten. Darüber hinaus wird untersucht, welche Art von Frames die Merkmale komplexer akademischer Sprache am besten wiedergibt. Dazu werden kontinuierliche und diskontinuierliche Frames unterschiedlicher Breite verwendet. Alle Analysen werden sowohl für das Englische als auch für das Deutsche durchgeführt, um den Wert dieser Berechnungen aus einer sprachübergreifenden Perspektive zu bewerten. Die Vermutung liegt nahe, dass sich mithilfe der Frequent Frames Analyse Aspekte und Phrasen identifizieren lassen, die spezifisch in der akademischen Sprache auftreten.

Mihail Chifligarov
Automated scoring of a German written elicited imitation test

---

In this talk, a recent workshop paper will be presented, focusing on the automated scoring of a German Written Elicited Imitation Test. The test assesses procedural knowledge in German as a foreign language by asking learners to reproduce briefly displayed sentences in writing. Responses are scored on a 5-point scale, where grammatical accuracy is weighted more heavily than lexical correctness. Two models are compared: a rule-based system built on hand-crafted scoring rules, and a neural model trained on stimulus-response pairs. Both achieve strong agreement with human ratings, but different strengths are observed. The rule-based model performs better on unseen sentences and extreme scores, while the neural model shows advantages in mid-range scores. By presenting both the methodology and key findings, this talk aims to highlight the potential of combining rule-based structure with neural flexibility in developing robust automated scoring systems.

Nele Mastracchio and Anna Stein
Comparing masked language model scoring metrics

Unlike sentences from auto-regressive language models, sentences from masked language models cannot be scored using the chain rule since some of the tokens are masked. To counter this problem, Salazar 2021 proposed a pseudo-loglikelihood (PLL) scoring method where tokens are successively masked with the rest of the sentence serving as context. Building on this, Kauf & Ivanova 2023 suggest a modified masking scheme to address problems with OOV words in Salazar's PLL metric. Sentence length should not affect PLL scores, however, the modified masking scheme shows unexpected correlations between sentence length and perplexity. Nonetheless, the modified scheme shows unexpected sentence length effects.
We extend the study by Kauf & Ivanova 2023 using the training data and models from the Babylm challenge. We present a comparison between the original PLL and the newly proposed PLL metrics with a focus on the sentence length effects.

Liliia Maiorova

Interpreting prosody:
variation in theoretical assumptions and its impact on computational models

---

This pet project looks at how different interpretations of prosody influence the ways it is separated from speech. Many researchers are now building models that learn prosodic patterns directly from audio, often without using labeled data. These models are used for tasks like emotion recognition, voice conversion, and speech synthesis. However, their success largely depends on how prosody is defined and represented — and definitions and interpretations of prosody differ greatly from one study to another, often in ways that seem chaotic or inconsistent.
The presentation addresses several key questions: What exactly is prosody, and in what ways do differing definitions impact the performance and outcomes of computational models? What methods are commonly used to identify and represent prosodic features? And why is the involvement of linguists crucial in the development of such systems?

Nune Arazyan

Klassifizierung der Lieder aus dem Allgemeinen Deutschen Kommersbuch
nach ihren Kategorien

Seit mehreren Jahrhunderten werden auf Veranstaltungen von Studierendenverbindungen Lieder aus dem Allgemeinen Deutschen Kommersbuch gesungen. Dieses Liederbuch, das Mitte des 19. Jahrhunderts erstmals erschien und seitdem in zahlreichen Auflagen überarbeitet wurde, enthält eine Sammlung von Liedern, die für verbindungsstudentische Anlässe traditionell von Bedeutung sind. Die darin enthaltenen Lieder sind (mit leichten Abweichungen je nach Auflage) in unterschiedliche, historisch und kulturell begründete Kategorien eingeteilt. Diese Einteilung durch die Herausgebenden wirft die Frage auf, welche spezifischen Merkmale und Gemeinsamkeiten die Lieder innerhalb der Kategorien aufweisen, die eine klare Kategorisierung ermöglichen.
Das Ziel dieser Arbeit ist es, diese Merkmale durch eine automatisierte Feature-Extraktion zu identifizieren – sowohl innerhalb der einzelnen Kategorien als auch kategorieübergreifend. Basierend auf diesen extrahierten Merkmalen wurde eine Klassifizierung der Lieder vorgenommen und evaluiert. Die Inhalte des Allgemeinen Deutschen Kommersbuches sowie die Kategorisierungen der Lieder wurden mittels Webscraping aus der digitalisierten und transkribierten Auflage 55-58 von Wikisource entnommen.

Jakob Schuster

Question authority: the Gold Standard

Standard NLP protocol relies heavily on gathering task-specific human-made annotations and aggregating them into a single truth label - the gold standard. However, a growing body of research raises doubts, whether solely relying on the gold standard sufficiently supports the strong conclusions it is often used to justify. This talk discusses the gold standard on several levels: the annotation landscape, the diversity of human language, conventions of academic evaluation, and effects of fine-tuning on LLMs.

Katharina C. Hummel

The language of the weather forecast on the British Isles – a standardized variety?

The weather determines the preparations, possibilities and risks of people’s daily professional and private activities and thus forecasts are to be accessible and comprehensible. Accessibility is achieved through digital media formats, and comprehensibility depends on the linguistic features employed by the presenters.

Based upon an examination of a corpus of written and transcribed spoken fore-casts by five broadcasters in the United Kingdom and Ireland utilizing keywords, word sketches, n-grams and concordances, the presenter argues that these lin-guistic features systematically differ from syntax structures and the expression of weather events and uncertainty in everyday language use. This allows for iden-tifying weather forecasts in larger text collections and considering them a stand-ardized register of English. The results are linked to existing research findings on weather event expression, as well as the connection of science and mass media and language use in weather forecasts to provide insight into the specific distri-bution of patterns. The findings have implications for language teaching.

Jacob Suchardt

Training for the unexpected – approaching universal phone recognition with whisper

---

Natural Language Processing has seen immense improvements in recent years, yet these are mainly reserved for "high-resource" — particularly English and Indo-European — languages. In contrast, citing a "data bottleneck" as the cause, the vast majority of the 7000+ global languages is still largely excluded from these technologies, engendering a "research and performance disparity" which threatens to reinforce itself continuously.
Speech-to-Text (STT) holds potential to address this data bottleneck as speech recordings of LRLs are more frequently available despite textual data scarcity. However, current STT systems rarely cater to LRLs and are furthermore dependent on orthographic representations, while LRLs may have neither orthographical nor writing conventions. To further documentation, research, and development of LRLs, we attempt to develop a universal speech recognition system by shifting the focus from orthographic to phonetic representations and fine-tuning the state-of-the-art Whisper model to generate transcriptions in the International Phonetic Alphabet directly from spoken language.

Carina Zander

Trier Digital - Wandel sichtbar machen

Dieses interdisziplinäre Projekt untersucht den städtischen Wandel in Trier durch die Integration digitaler Methoden mit geisteswissenschaftlicher Forschung. Das von Masterstudierenden der Digital Humanities durchgeführte Projekt verfolgt einen multiperspektivischen Ansatz zur Stadtentwicklung, der räumliche, visuelle und diskursive Dimensionen einbezieht. Ein besonderer Schwerpunkt liegt auf der Computerlinguistik: Durch quantitative Textanalyse, Themenmodellierung und Emotionserkennung untersuchte das Team die rhetorischen und affektiven Strukturen der kommunalen politischen Kommunikation. Die Analyse öffentlich zugänglicher Quellen, wie der offiziellen Zeitung der Stadt, ermöglichte die Identifizierung thematisch bedeutsamer Orte, wiederkehrender diskursiver Muster und emotionaler Dynamiken, die den öffentlichen Diskurs prägen. Ergänzt durch historische Kartographie und visuelles Material zeigt das Projekt, wie digitale Werkzeuge latente Strukturen in städtischen Erzählungen aufdecken und ein nuanciertes Verständnis der politischen und kulturellen Prozesse bieten können, die den städtischen Wandel vorantreiben. Diese Arbeit ist ein Beispiel für das kritische Potenzial der digitalen Geisteswissenschaften an der Schnittstelle von Sprache, Raum und Gesellschaft.

Jakob Moser, workshop

Typst: The new LaTeX?

This workshop is a hands-on introduction to the typing system Typst (https://typst.app), an alternative to LaTeX. The participants will explore how to create basic documents and presentations, and play around with some more advanced features, like scripting. By the end of this workshop, you should be

able to write your own paper in Typst. No previous knowledge is necessary, although the introductory slides will only be funny if you fought against LaTeX at least once.

Adrian Mülthaler
What the Heck is Quantum NLP?

Have you ever wondered what quantum physics and natural language have in common? In this presentation, the author explores how quantum-inspired models — yes, quantum! — can be used for natural language processing. As large language models become more powerful, so do concerns about their scalability, interpretability, and efficiency. The presentation includes a brief introduction to quantum computing before focusing on the DisCoCat framework (short for Distributional Compositional Categorical), which combines vector space semantics with grammar using category theory — and can be described as 'quantum native.' From there, the presenter will move on to a discussion of how its simulation complexity can be examined (based on a bachelor's thesis), offering some insight into whether NLP could benefit from the use of quantum computers.