

Natural language and AI. New perspectives for linguistic studies

In den letzten Jahren hat die Künstliche Intelligenz (KI) in vielen Disziplinen erhebliche Fortschritte erzielt. Die Integration von KI-Anwendungen, insbesondere von *Large Language Models* (LLM), in linguistische Studien hat neue Horizonte für die Analyse natürlicher Sprache eröffnet. Von der Morphosyntax über die Semantik bis hin zur Variationslinguistik bieten diese Technologien Linguistinnen und Linguisten die Möglichkeit, komplexe sprachliche Phänomene zu untersuchen. Diese Entwicklung hat dazu geführt, dass sprachliche Aufgaben wie Textgenerierung, Übersetzung und Korpusannotation in einem Maße automatisiert werden können, das zuvor unvorstellbar war.

Die Anwendung der KI in linguistischen Untersuchungen offenbart aber auch Herausforderungen. Eine bedeutende Problematik liegt beispielsweise in der Notwendigkeit, angemessene Trainingsdaten für KI-Modelle bereitzustellen, die eine breite Vielfalt sprachlicher Phänomene und Strukturen abdecken. Oftmals sind diese Daten unvollständig, ungleichmäßig oder sogar fehlerhaft, was die Zuverlässigkeit der KI-Systeme beeinträchtigen kann. Ein weiteres Hindernis ist die Tatsache, dass KI-Modelle auf bereits existierenden Daten trainiert werden, was dazu führen kann, dass sie implizite Bias übernehmen, die in den Trainingsdaten vorhanden sind. Dies kann zu Verzerrungen in den Ergebnissen führen und die Neutralität und Objektivität sprachwissenschaftlicher Analysen beeinträchtigen. Des Weiteren scheint Sprachvariation eine Herausforderung darzustellen. KI-Modelle müssen in der Lage sein, diese Vielfalt zu erkennen und angemessen zu verarbeiten. Dies ist oft schwierig, da die Modelle möglicherweise auf bestimmte sprachliche Muster oder Normen beschränkt sind.

Im Rahmen des geplanten Workshops werden verschiedene Anwendungen der KI (wie ChatGPT, DialoGPT, Meena, BlenderBot usw.) diskutiert, die für sprachwissenschaftliche Studien von Bedeutung sein können. Durch die Zusammenführung von theoretischen Ansätzen der Linguistik und modernen KI-Methoden werden die Potenziale und Herausforderungen dieser Technologien für die linguistische Forschung ausgelotet.

Im Fokus des Workshops stehen insbesondere – jedoch nicht ausschließlich – folgende Forschungsfragen:

- Wie können KI-Anwendungen genutzt werden, um Grammatikalitätsurteile in verschiedenen Sprachen zu untersuchen und zu vergleichen? Welche Anwendungen kommen der muttersprachlichen Intuition am nächsten?
- Inwiefern können *Large Language Models* (LLM) zur automatischen Korpusannotation und -analyse eingesetzt werden, um sprachliche Variation zu identifizieren und zu verstehen?
- Welche Rolle spielen semantische Modelle und neuronale Netzwerke bei der Übersetzung zwischen Sprachen mit unterschiedlicher syntaktischer Struktur?
- Wie können mithilfe von NLP-Techniken die Bedeutung und die Verwendung von Sprachvariation in verschiedenen sozialen Kontexten untersucht werden?

Linguistinnen und Linguisten verschiedener theoretischer Ausrichtungen sind eingeladen, an dieser Arbeitsgruppe teilzunehmen und empirische Studien zu präsentieren, die sich mit aktuellen Entwicklungen und zukünftigen Perspektiven im Bereich der natürlichen Sprachverarbeitung und der KI für linguistische Studien befassen. Sprache des Workshops ist Englisch.

Abstracts für 20-minütige Vorträge (plus 10 Minuten Diskussion) müssen im .doc- und im PDF-Format eingereicht werden. Sie sollten eine Seite nicht überschreiten (ohne Literaturverzeichnis), in einer Schriftgröße von mindestens 12pt verfasst sein und Ränder von 1 Zoll/2,5 cm aufweisen. Jede(r) Autor(in) kann höchstens ein Abstract als Einzelautor(in) und ein Abstract als Koautor(in) einreichen.

Abstracts sind spätestens bis zum 01. September 2024 per E-Mail einzureichen an:

catasso@uni-wuppertal.de

thomas.scharinger@uni-jena.de

Ansprechpartner:

Nicholas Catasso (catasso@uni-wuppertal.de)

Thomas Scharinger (thomas.scharinger@uni-jena.de)