

## **Project A05: Contextualised metrics of linguistic creativity in literary and non-literary text**

**Project leaders:** Berenike Herrmann, Sina Zarriß

### **Project Summary**

In this interdisciplinary project situated between linguistics and literary studies, we will investigate how linguistic creativity can be identified and measured across literary and non-literary genres with corpus-based and mixed methods. We ask how the originality and success of linguistic signs vary depending on textual cues and are mediated by a spectrum of popular and artistic genre contexts such as travel blogs, popular fiction and literary novels. Focusing on the thematic domain of spatial descriptions across genres (e.g. „the squares devour a huge amount of drinking water to stay green“; „wood was chopped down in the mountain forests, and lay there clean, dead and fragrant“) we aim to develop data-driven metrics and models that automatically identify creative signs at sentence, paragraph and text level.

In order to conceptualise and model linguistic creativity, we will develop a mixed-method-setup that integrates different stylistic, experimental, and computational methods: we will use intuitive ratings of readers to account for literariness and creativity in usage across genres, philological expert annotations of textual literariness (foregrounding devices), as well as genre-sensitive language models and corpus-based metrics to measure and classify creativity. We will then leverage these for the annotation of texts and rating studies, which will feed back into computational modeling and the development of new methods for computational stylistics. A central goal of the project is to extend and integrate existing resources in literary and computational stylistics with recent language model- and embedding-based methods from NLP that can incorporate feature-driven knowledge of foregrounding and rhetorical devices. The main contributions of the project will be made within the areas of Computational Linguistics and Computational Literary Studies, but we will also collaborate extensively with other projects in the CRC targeting questions of core linguistics, studying whether our corpus- and rating-based metrics generalize to specific linguistic phenomena.

### **Open Positions**

#### **PhD position 1 (65%)**

*Profile:* The ideal candidate has a master in German philology, literary studies, digital humanities or a related field, with some experience in either stylistics, rhetoric, and structuralist approaches to reading or in computational text analysis, stylometry and corpus-based literary studies. A (near) native speaker command of German is a pre-requisite as well as a strong interest in formal and computational approaches to literary language and discourse.

*Main research focus within the project:* The dissertation topic will focus on developing a genre-sensitive operationalisation of linguistic creativity in literary texts, for a corpus of literary texts to conduct readers' rating studies, expert annotations, and text-level computational stylistic analysis. Work will thus include conceptualisation and operationalisation of literary creativity; creation of de-description corpora; design and realisation of annotation studies; design and realisation of rating studies; work on computational and stylometric metrics of linguistic creativity for literary descriptions. Where relevant, the research will be synchronized with the other PhD in the project (Computational Linguistics).

### **PhD position 2 (100%)**

*Profile:* The ideal candidate has a master in Computational Linguistics or a related field, some practical experience in working with language models and machine learning in NLP and a strong interest in linguistically motivated approaches to computational language modeling.

*Main research focus within the project:* The dissertation topic will focus on developing sentence embedding models that disentangle style and content, based on language models for different literary and non-literary genres. Work will thus include fine-tuning German language models on different genres and developing sentence embedding models; creation of description corpora and work on detection of descriptions; computational experimentation on creativity metrics; work on case studies and collaborations with other projects in the CRC.

**For further information please contact the project leaders:**

berenike.herrmann@uni-bielefeld.de, sina.zarriess@uni-bielefeld.de