



Institute of Computer Science, Polish Academy of Sciences
Jana Kazimierza 5, 01-248 Warszawa, Poland

JOB OFFER

Job type:	Employment contract, post-doc position
Project:	Universal Discourse: a multilingual model of discourse relations financed within NCN MAESTRO grant No: 2023/50/A/HS2/00559
Project leader:	Maciej Ogrodniczuk
Project description:	<p>The project intends to create a unified description of discourse relations (at the level of discourse markers, relation arguments and types) in the multilingual setting by harmonizing current corpus-based discourse representation models. In turn, the computational properties of the newly created description will be verified by creating prototypes of a multilingual discourse parser, fine-tuned on existing large language models with the harmonized dataset.</p> <p>The first phase of the proposed research will be the development of a multilingual ontology of discourse relations modelled after ISO 24617-8 standard and containing usage examples excerpted from existing discourse corpora and literature. Speakers of at least 10 European languages with linguistic backgrounds will be consulted to maintain sufficient coverage of constructs. The ontology will be used to create annotation instructions for human annotators to carry out the phase of corpus</p> <p>harmonization. In this step, a set of existing corpora following various discourse representation theories will be reannotated to provide a harmonized cross-lingual dataset for further analyses and computational work. The harmonization will be carried out manually by proficient users of corresponding languages with a strong linguistic foundation.</p> <p>The data will be then used to create prototypes of multilingual discourse parsers focusing on two main sub-tasks: identification of discourse relation realization types and the sense classification of explicit and implicit relations. The prototypes will be implemented by fine-tuning existing large language models with newly produced data and will be evaluated quantitatively and qualitatively.</p> <p>In the final step, the project results will be disseminated. The ontology of discourse relations will be integrated into one of the ontological repositories. The resulting datasets will be made available in a common search interface. The dataset and model will be used as input for a comparative study and as training data for the shared task on multilingual discourse parsing planned to be co-located with one of the major natural language processing conferences with the prototype of</p> <p>the discourse parser used as a baseline. The evaluation methodology of the shared task will use the methods proposed by the project.</p>
Salary:	140 000 PLN per year (project cost)
Minimum period of contract:	6 months

This position is funded in whole by National Science Centre, Poland, Grant MAESTRO No. 2023/50/A/HS2/00559,
Universal Discourse: a multilingual model of discourse relations (UDISC), dr hab. Maciej Ogrodniczuk





Institute of Computer Science, Polish Academy of Sciences
Jana Kazimierza 5, 01-248 Warszawa, Poland

Maximum period of contract:	3 years
Application deadline:	6 June 2025
Decision up to:	9 June 2025
Position starts on:	1 July 2025
Evaluation criteria:	<ol style="list-style-type: none"> 1. PhD in linguistics, obtained up to 7 years from the date of employment 2. Knowledge of discourse representation models, proven by (co-)authorship of at least one publication related to discourse phenomena 3. Previous experience in discourse annotation projects, proven by participation in at least one annotation project/task related to annotation of discourse relations
Key responsibilities:	<ul style="list-style-type: none"> – participation in the development of the multilingual ontology of discourse, relations and in the corpus-based analysis of the existing discourse descriptions, – participation in the creation of the discourse annotation guidelines, – adjudication of discourse annotation, – qualitative evaluation of the discourse parsers, – carrying out the comparative study based on corpora annotated with discourse relations, – co-authoring of the project articles and/or the monograph on Universal Discourse
Required documents:	<ol style="list-style-type: none"> 1. CV 2. Copy of the PhD diploma 3. List of publications related to analysis of discourse phenomena 4. List of research projects/tasks related to annotation of discourse relations
Please submit the documents to:	justyna.kot@ipipan.waw.pl