

Załącznik 1. Projekt ogłoszenia konkursowego

Job type:	PhD student scholarship
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:	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. 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 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. 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. 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 the discourse parser used as a baseline. The evaluation methodology of the shared task will use the methods proposed by the project.
Scholarship:	5 000 PLN per month
Minimum period of contract:	6 months
Maximum period of contract:	2 years
Application deadline:	15 September 2026
Decision up to:	18 September 2026
Position starts on:	1 October 2026

Evaluation criteria:	1. Bachelor or Master's degree in linguistics 2. Familiarity with basic natural language processing methods
Additional evaluation criteria:	1. Knowledge of Polish language corpora 2. Knowledge of discourse representation models 3. Previous experience in discourse annotation projects
Key responsibilities:	participation in all linguistic project activities (analysis of corpora, preparation of annotation guidelines, qualitative evaluation, preparation of the comparative study) under the supervision of the PI and post-doc researchers
Required documents:	1. CV 2. Copy of the degree diploma 3. Motivation letter including all information relevant to the assessment process 4. Certificate of (doctoral) student status
Please submit the documents to:	justyna.kot@ipipan.waw.pl