

# Short-Term Scientific Mission Grant - APPLICATION FORM<sup>1</sup>-

**Action number: CA18209**

**Applicant name: Lucía Pitarch Ballesteros**

## Details of the STSM

Title: Population of LLOD cloud with Deep learning approaches: Metaphor conceptualization and multilingual lexical relation acquisition.

Start and end date: 01/09/2023 to 01/10/2023

## Goals of the STSM

Purpose and summary of the STSM.

(max.200 word)

This STSM includes two main goals. The first one consists on validating the multilingual version of BATS (Bigger Analogy Test Set) dataset, created in U.C.4.1.3. task, by conducting relation acquisition and cross-lingual transfer learning experiments with it. The outcome could then be then represented as Linked Data, and could be used to enrich the LLOD (Linguistic Linked Open Data) cloud, by adding the lexical and semantic category entailed between pairs of words. This first goal would encompass the following Nexus Linguarum tasks: 3.3. Multilingual approaches, 2.1. Knowledge extraction, 3.2. Deep learning and 4.1.3. multilingual BATS dataset creation.

The second goal of the STSM consists on exploring the existing Linked Data models to represent metaphors. It englobes WG1 and task 2.1. Knowledge extraction. This goal stands as the initial point for future research in metaphor identification and conceptualization in structured data format.

The STSM would benefit from the host institution (Centre for Translation Studies, Austria) by being the heart of the development of multilingual BATS dataset, being involved in key Lexical Relation Acquisition studies, and previous experience working with metaphors. While the applicant would provide experience in the use of PTLMs (Pretrained Language Models) and Prompt Engineering for their exploitation.

<sup>1</sup> This form is part of the application for a grant to visit a host organisation located in a different country than the country of affiliation. It is submitted to the COST Action MC via-e-COST. The Grant Awarding Coordinator coordinates the evaluation on behalf of the Action MC and informs the Grant Holder of the result of the evaluation for issuing the Grant Letter.

**COST Association AISBL**

Avenue du Boulevard – Bolwerklaan 21 | 1210 Brussels, Belgium  
T +32 (0)2 533 3800 | office@cost.eu | www.cost.eu



Funded by  
the European Union

## Working Plan

Description of the work to be carried out by the applicant.

(max.500 word)

Goal 1: BATS validation through relation acquisition and crosslingual transfer experiments

- Initial learning phase on the used approaches by the host group in Lexical Acquisition and Crosslingual transfer
- Problem statement, available models for the mentioned tasks, and BATS dataset analysis.
- Experimental exploration of the use of BATS data on different multilingual models such as Chat-GPT, BLOOM, multilingual BERT or XLM-Roberta, through zero and few shot settings and different prompts.
- Analyse the viability of adding the outcome novel sets of lexical relation triples extracted through our approach to further enrich the LLOD cloud.

Goal 2: Metaphor conceptualization through LLOD.

- Analysis of the available resources in the LLOD field to conceptualise metaphors.
- Analysis of the available datasets for multilingual metaphor identification and mapping to source domains.
- Run initial experiments on multilingual metaphor identification and source domain mapping.

## Expected outputs and contribution to the Action MoU objectives and deliverables.

Main expected results and their contribution to the progress towards the Action objectives (either research coordination and/or capacity building objectives) and deliverables.

(max.500 words)

Experimental results from goal 1 could be presented as contributions in a conference or journal. Additionally, BATS dataset would be validated through relation acquisition and cross-lingual transfer learning experiments. This should serve as feedback for a second, curated version of the multilingual BATS dataset created in task 4.1.3, that might be represented as Linked Data.

The results of the explorative analysis of metaphor representation with LLOD (goal 2) should be twofold: pointing at the limitations and possibilities of their conceptualization as LLOD, as well as providing them in a structured way which can be then exploited by Language Models to further enhance their identification, mapping, and processing. This will serve as the basis for a future model to represent metaphor as Linked Data and their linkage with other LLOD resources.

The outcomes of this STSM would contribute to the following Action's deliverables: D12 (final activity reports, particularly from WG2, WG3, and WG4) and D13 (scientific papers).