

Report on the outcomes of a Short-Term Scientific Mission¹

Action number: 18209

Grantee name: Anna Bączkowska

Details of the STSM

Title: **Offensive language translation in film discourse and its integration with LLOD methods: OntoLex-FrAC and VarTrans modules**

Start and end date: 02/04/2022 to 28/04/2022

Description of the work carried out during the STSM

Description of the activities carried out during the STSM. Any deviations from the initial working plan shall also be described in this section.

(max. 500 words)

Grantee enters max 500 word summary here.

During my research stay at the University of Vienna I met on a regular basis with my Host, Dr Dagmar Gromann in the Centre for Translation Studies. We discussed and decided on a number of theoretical foundations of the project, including: the precise definition of offensive language encoded by swearwords, the linguistic typology of functions played by swearwords, a list of swearwords included in our study (which was extended twice stemming from some shifts in the ultimate theoretical framework adapted), the methodology of data retrieval, the definition of context for the swearwords under inspection, the possible applicability of the study to LLOD. We have also searched for several versions of the data – subtitles for the romantic comedy “Bridget Jones’s Diary”, in the English original version, Polish professional and fansubbing versions and German professional and fansubbing versions. After some revision, we decided on the final versions of the material. We have had a trial annotation task with the English professional version in English independently to check interrater agreement, next adjusted the original taxonomy of functions accordingly, and started annotation proper again, which will be finished in May. After that, we are planning to have an online meeting to discuss the interrater agreement again. We have exemplified and decided to utilize OntoLex-Lemon in combination with vartrans for the multilingual representation of swearwords, representing functions as a separate OntoLex-Lemon resource, and to represent information from the respective subtitle corpora in OntoLex-Frac, including

¹ This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

frequency information. Moreover, I have undergone several trainings in the LLOD conducted for me by the Host, both the theoretical aspects and the hands-on analysis with the use of select software available online. We have also started writing a joint paper (ca. 1500 words) which will present our original approach to the analysis of swearwords in film subtitles, decided on the possible publication venues and roughly discussed the scope of a joint presentation, to be given at an international conference organised by the COST ACTION NexusLinguarum in Vilnius, on the 23-24th of September 2022. Overall, we have decided to write two joint papers, one to be submitted to 'Rasprave' (Q2) as a post-conference publication, and one to "Language Resources and Evaluation" (Q1). On the 8th of April, I also gave a talk at the Centre of Translation on the translation of offensive language in film subtitles: **"Im/politeness and offensive language in professional and nonprofessional subtitles"**.

The tools and resources we are using are below:

For OntoLex-Lemon <https://www.w3.org/2019/09/lexicog/>), for vartrans (just one format): Elena Montiel-Ponsoda, Jorge Gracia, Guadalupe Aguado-de-Cea, and Asunción Gómez-Pérez. 2011. Representing translations on the semantic web. In Proceedings of the 2nd International Conference on Multilingual Semantic Web - Volume 775 (MSW'11). CEUR-WS.org, Aachen, DEU, 25–37. For OntoLex-Frac: <https://acoli-repo.github.io/ontolex-frac/>

I have found this STSM a very fruitful research stay, even though not all planned tasks have been fully covered by us (mostly due to some technical problems with the subtitle data). Nevertheless, we have managed to do a great deal and have already planned future joint measures.

Description of the STSM main achievements and planned follow-up activities

Description and assessment of whether the STSM achieved its planned goals and expected outcomes, including specific contribution to Action objective and deliverables, or publications resulting from the STSM. Agreed plans for future follow-up collaborations shall also be described in this section.

(max. 500 words)

Grantee enters max 500 word summary here.

Our main achievements include:

- Writing about 1500 words of our joint paper
- Introductory analysis of the English original film dialogues – annotating the text and checking interrater agreement
- Designing a linguistic taxonomy of swearwords implemented in the automatic extraction of swearwords
- Proposing and defining key concepts within the theoretical framework adopted: swearwords and their functions
- Planning further measures that comprise:
 - a. A joint paper for 'Rasprave' (finishing the current one); Q2
 - b. A joint paper for 'Language Resources and Evaluation', Q1
 - c. A joint presentation at the NexusLinguarum Vilnius conferences in September 2022
 - d. Further joint projects:
 - (i). a STSM visit of the Host at the University of Gdańsk in 2023;
 - (ii). a second STSM visit of the grant Holder in 2022 or 2023 to finish the current research aims.
- Dissemination of research on swearwords and LLOD at a guest talk I gave on the 8th of April 2022.

Follow-up activities:

- Writing and publishing two joint papers

- Giving a presentation at an international conference
- Further joint measures through STSM, both in Poland and in Austria
- My further training in LLOD