SHORT TERM SCIENTIFIC MISSION (STSM) – SCIENTIFIC REPORT

The STSM applicant submits this report for approval to the STSM coordinator

Action number: CA16105
STSM title: Preparation of a crowdsourcing experiment within the purview of the "Learning materials through crowdsourcing: teachers, perspectives & scenarios" workshop for WG1
STSM start and end date: 01/09/2018 to 15/09/2018
Grantee name: Jaka Čibej

PURPOSE OF THE STSM

(max.500 words)

The motivation for the proposed STSM is a meeting organized by Work Group 1 in Gothenburg, Sweden, between the 5th and 7th December 2018. The meeting is inviting WG1 members (and if funding allows, people outside WG1) to join a hands-on workshop on the topic of "Learning materials through crowdsourcing: teachers, perspectives & scenarios" (more info at https://spraakbanken.gu.se/eng/wg1-dec18-gbg). Aside from presentations and discussions on relevant crowdsourcing topics, a crucial point of the meeting is also to provide a hands-on exercise in crowdsourcing for the group members, which will involve a small-scale crowdsourcing experiment and introduce all the steps involved in the process of setting up a crowdsourcing experiment (technical insight, data format, data upload, requirements, etc.).

To make the crowdsourcing experiment manageable within the context of the three-day meeting, the topic of the crowdsourcing task was narrowed down to English multi-word expressions in second language learning context. Even before the meeting, the participants will be asked to complete a crowdsourcing (mini-)task based on English data, i.e. act as a "crowd". The collected data will then be discussed and analyzed as part of the meeting (involving group work).

The main focus of the STSM was to prepare everything necessary for the smooth execution of the hands-on tutorial and practical exercises at the WG1 meeting in December. In particular, the goals of the STSM involved the following: (a) prepare the necessary data that is to be used in the crowdsourcing task aimed at the participants of the meeting; (b) prepare the necessary technical infrastructure for the microtask; (c) test the crowdsourcing workflow and prepare the final version of the crowdsourcing experiment for the meeting.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

(max.500 words)

The STSM involved the following steps:

Funded by the Horizon 2020 Framework Programme
of the European Union

Step 1: We exported all available data from a database of English multi-word expressions annotated with their corresponding CEFR-level. We selected a limited number of expressions to include in the experiment. The maximum number of expressions was determined in order to keep the experiment manageable as part of the WG1 meeting in December 2018. Expressions were then grouped in different combinations, each representing a single microtask to be solved by the crowdsourcers.

Step 2: The selected expressions were then converted into .JSON, a PyBossa-compatible format.

Step 3: A project was created in the PyBossa platform (the local installation used at the Centre for Language Resources and Technologies at the University of Ljubljana). The selected microtasks were uploaded into the project.

Step 4: A suitable, user-friendly microtask interface was developed in PyBossa. Several different microtasks designs were tested in order to determine the optimal approach in terms of general user-friendliness, minimum number of clicks required, methodological suitability, etc. The final design was also tested on real users and confirmed to be adequate for the experiment.

Step 5: Annotation guidelines were developed, as well as instructions on how to register on the PyBossa platform. A testing phase for annotation was started in order to determine the length and feasibility of the crowdsourcing task in line with the allotted time planned at the meeting. The testing phase confirmed that the preliminary calculations were correct and that the task is feasible with the available time and number of crowdsourcers.

Step 6: The output data from the testing phase of the PyBossa project was preliminary analyzed through a custom-developed script.

Step 7: Guidelines were additionally refined in line with the comments provided by the crowdsourcers in the testing phase.

Step 8: The final crowdsourcing task infrastructure for the pre-meeting crowdsourcing task was set up. The links and instructions on how to participate in the task will be sent out to the workshop participants by the end of September 2018.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

(max. 500 words)

The developed methodology for selecting multi-word expressions and preparing the data can be re-used for similar crowdsourcing experiments in other languages.

The STSM resulted in a successful set-up of the infrastructure for the crowdsourcing experiment to be carried out before and at the WG1 workshop in Gothenburg (5-7 December 2018).

The annotation guidelines will be made publicly available and can be translated into other languages for similar experiments. Other resources will also be made freely available, such as the developed templates for the microtask interface and the scripts for data preparation and the analysis of the crowdsourcing results.

In addition, if the experiment proves successful, it will result in a database of English multi-word expressions that can be used as a language resource in second language teaching.

FUTURE COLLABORATIONS (if applicable)

(max.500 words)

Possible future collaborations include assistance with other microtasks and help with a new PyBossa installation for Språkbanken. Similar experiments for other languages may also provide opportunities for collaboration.