SHORT TERM SCIENTIFIC MISSION (STSM) SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

Action number: 44518
STSM title: Testing and evaluation of intergenerational language game that fosters language learning and collects language data through explicit crowdsourcing
STSM start and end date: 03/06/2019 to 07/06/2019
Grantee name: Marianne Grace Zshornack Araneta

BACKGROUND AND PURPOSE OF THE STSM:

There are many non-standardized languages and dialects with no standard orthography, or at the very least, their own corresponding references to negotiate toward standardization. This poses a problem for the preservation of the language, especially for those languages with not only an aging, but also decreasing population of users. This becomes even more problematic when these non-standardized languages lack family transmission. In order to preserve such languages, it is therefore necessary to achieve the following: (1) foster interest in order to encourage learning of these languages, and at the same time, (2) find a means to gather data on these non-standardized languages.

A way to achieve both objectives is to use gamified learning and crowdsourcing techniques. This was explored during a hackathon organized by enetCollect in Brussels early this year. The game that was initially developed through the platform rpgboss aimed to use a role-playing game (RPG) in order to encourage language learning and collect language-related data within the game's coherent narrative. The game also included non-linguistic activities in order to sustain user interest. The prototype was thereby a mix of concepts in NLP, linguistics, language learning, crowdsourcing, and game design and development.

The primary goal of the STSM was to continue and refine the work that had already begun during the hackathon. To accomplish such a goal, the STSM had the following objectives:

1. To further develop the narrative and conceptualize more activities embedded within the game, such as:
   a. Activities geared toward language data collection and language learning
   b. Activities or obstacles within the game that focus on skills or tasks unrelated to language transmission in order to provide users with a variety of tasks and sustain their interest
   c. Other activities that can evaluate the crowdsourced data and also the level of language learning
2. To construct a mid-fi prototype of the game within the new platform "RPGMaker", as it allows more options than the previous one (rpgboss); this also includes studying the RPGMaker platform in order to exploit all the functionalities of the game
3. To establish an infrastructure for continued collaboration among group members

DESCRIPTION OF WORK CARRIED OUT DURING THE STSM

The schedule of the work was as follows:

3 June 2019
- Met with Karën Fort, Host Institution Supervisor as well as team supervisor, and Alice Millour, team member, in order to discuss the following:
  - Progress of game development on the new platform (RPG Maker)
  - Existing game narrative and areas to develop
  - Organize objectives and schedule for the next few days
- Worked with Alice Millour in order to build the framework of the game
  - Alice focused on finding a means to collect language data within the game and fix server-related issues
  - I focused on establishing Github, the agreed-upon communication channel for collaboration (including the parameters) and later on, started refining existing games to fit within a coherent narrative

4 June 2019
- After a debriefing, continued work with Alice, where she focused on back-end issues (i.e. server communication, collection of data across different servers, testing)
- Researched on the RPG Maker platform
  - Tutorial on game development within the platform
  - Looked for plug-ins that could allow more activities / tasks within the game
- Continued developing the narrative, this time focusing on non-linguistic activities given the limitations of the RPG Maker platform

5 June 2019
- Attended a seminar on Language Adaptation Experiments at Ecole Normale Supérieure
- Discussed with Alice different evaluation methods (in terms of game activities) that could be used within the game (in terms of crowdsourced data, level of language learning, and player engagement)
- Encoded game on RPG Maker (% Alice and Yann-Alan Pilatte, another member of the hackathon team)
- Fixed the bugs that emerged while developing the game (% Alice)
- Continued research on valuable plug-ins and activities in the RPG Maker platform; also took a look at research on game design and establishing coherent game narratives (% Marianne)

6 June 2019
- Continued encoding the game on RPG Maker and fixing bugs (% Alice)
- Elaborated interactions within the game, more particularly wrote the scripts and tasks per activity (% Marianne)
- Updated issues on Github, where the project is made private and accessible only to the members of the original team during the Hackathon

7 June 2019
- Met with a first year Master student at the University of Sorbonne to discuss the concept, as it was relevant to her research on Mauritian creole
- Had a debriefing with Karën on the week's accomplishments and future work
- Revised the narrative flowchart in order to gain a better view of the storyline and the activities (linguistic, non-linguistic, evaluation) developed so far

DESCRIPTION OF THE MAIN RESULTS OBTAINED

The following were the main results (and insights) obtained during the five-day STSM:
1. Established a communication channel for (mainly remote) collaboration that allows us to:
   a. See list of tasks and assign them accordingly
   b. Monitor problems (such as bugs)
   c. Share codes (e.g. plug-ins) and documentation securely
   d. Give updates on work completed
2. Explored the RPG Maker platform in terms of:
   a. Plug-ins and built-in features (and therefore activities) within the platform
   b. Means to gather data encoded by the players (toward crowdsourcing)
3. Developed a complete and coherent narrative (with a corresponding flowchart) for one setting of the game; such narrative included the following:
   a. Linguistic activities
   b. Non-linguistic activities
   c. Evaluation activities

   Such narrative incorporated concepts in game design (such as having tokens/keys, levels within the game, easter eggs).

As a team (Karën, Alice, and myself), we were able to confirm the following—and even gather additional—insights related to crowdsourcing language data using a gamified platform:
- Crowdsourcing language data of a non-standardized language is a challenge to evaluate, especially since there is no corpus that can be used as a basis; in this light, the primary objective would then be to collect such data and make sure that the crowdsourced material is neither random nor nonsensical. We were able to come up with certain activities that could check for this, but this is not as extensive, as we had to make it coherent with the game narrative.
- We also wanted to find a means to collect "original" data, where there is no single answer to a question. For example, there are certain linguistic activities that require the strict identification (i.e. translation) of a particular world, such as the French *fleur* for the English *flower*. This does not produce original data, and limits to certain answers, perhaps with only variations in spelling. While this kind of activity is purposeful, it is also important to include tasks that can elicit a variety of answers to a single query. This allows a greater breadth of crowdsourced data.
- While we were able to conceptualize various tasks that could have crowdsourced more language data, we made sure that they did not compromise the coherence of the game narrative. As established during the hackathon, the game is not strictly gamification; it is primarily a game that has a mix of non-linguistic and linguistic activities that crowdsource language data. Going in this direction could best sustain player interest and allow a variety of activities within the game.
- We also had to work within the limitations of the platform used. We first had to choose an appropriate platform given server limitations. We then had to find a way to collect the data that the user encoded in the platform.

FUTURE COLLABORATIONS (if applicable)

During the STSM, we were able to plan for future collaborations as the team aims to develop a working game. The team would have to initially work remotely to develop a mid-fi prototype that could eventually be tested on students (testing % Marianne). This should have been accomplished by August. The next course of action would be to have another STSM in October where members of the team (perhaps Alice and Marianne) could meet again and work on developing the game based on the test results, as well as working to publish a paper. Another option would be to take part in another hackathon to find the time and space to expand the game as a group. This method was very effective during the first hackathon, and another one could most definitely benefit the project.