SHORT TERM SCIENTIFIC MISSION (STSM) – SCIENTIFIC REPORT

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

Action number: CA16105
STSM title: The Poetic Eddas as part of a crowdsourcing experiment in LARA – Old Norse text
STSM start and end date: 21/12/2019 to 02/01/2020
Grantee name: Branislav Bedi

PURPOSE OF THE STSM

The purpose of the STSM was to investigate the feasibility of creating a new type of content for LARA, a platform designed to facilitate crowdsourced creation of multimedia online texts designed for non-native readers. So far, about forty texts have been created, in fourteen languages; these range in size from toy examples a page or two in length to short novels (*Alice in Wonderland* and *Animal Farm* (English) *Kallocain (*Swedish*), Le petit prince (*French*)).* The text created during my STSM was in Old Norse and consisted of the *Völuspá*, a famous poem which is part of the Codex Regius collection of Eddaic poems.

The creation of this material required a combination of expertise in Icelandic studies and software engineering. On the software engineering side, the functionality offered by the current version of the publicly available LARA portal was insufficient. We consequently worked directly with the underlying LARA core engine, extending it in several ways and invoking it through the command-line interface. The resulting LARA texts were then uploaded to the web. In more detail, the new functionality developed for this exercise was as follows:

1. The first task was to tag the text so as to mark each word with its associated lemma. The LARA portal does this by invoking ABLTagger and Nefnir, two state-of-the-art software components recently developed by Icelandic researchers in human language technology. Despite their excellent performance on modern Icelandic, ABLTagger/Nefnir do not do well on Old Norse texts like the Edda. Fortunately, however, we were given access to a manually tagged electronic edition of the Edda created by a colleague of mine at the Arni Magnusson Institute for Icelandic Studies. It was straightforward to write a script which converted the format used into that required for LARA.

2. The model previously used by LARA to attach translations to words also needed to be revised. The portal version only supports a very simple model, where translations are attached to lemma types; thus, each occurrence of a lemma is attached to the same translation. This turned out to be quite unsuitable for the complex poetic language of the Edda. Instead, we followed a suggestion made by Dr Sabina Sestigiani of Swinburne University, Melbourne, and developed a new alternative which attaches translations to surface word *tokens*. The annotator starts by dividing the text into segments, in poems this is typically lines or verses, and uses a spreadsheet to define a translation for each verse; for the Völuspá we used the public domain English translation by Henry Adams Bellows

(1936) listed on Sacred-Texts (https://www.sacred-texts.com/neu/poe/poe03.htm). The LARA core engine then produces a spreadsheet where each segment is presented as three lines, showing the words in the source text at the top, a blank line in the middle, and the segment translation at the bottom. Using the segment translation as a rough guide, the annotator fills in the middle line with the translation to be attached to each individual word. If they wish, they can start by defining default translations for word types, though we did not make use of this feature here.

3. Finally, we decided that it would often be useful to be able to attach explanatory notes to words as well as translations. This is done in a way similar to that used for the lemma-type translations: the LARA core engine creates a spreadsheet with one line for each lemma type, and the annotator fills in the notes where relevant. The core engine then uses the spreadsheet to insert the notes in the appropriate places and also create a linked file listing all the notes.

Of the three pieces of functionality listed above, the first is specific to the marked-up version of the Edda created by the Arni Magnusson Institute, but the script developed can be reused if we decide to create LARA versions of more Eddaic poems. The second and third are generic, and are currently being incorporated into the LARA portal.

The final LARA text, as deployed on the web, allows the user to hover the mouse over any word in the text to hear an audio recording, and also to click on it to get a page which contains a link to grammar information, a linked concordance of the places in the text where the word occurs, and an optional note. There are also links to alphabetical and frequency-ordered indexes and to an alphabetical list of notes, and some images and explanatory text. A caveat is that the grammar information is available only for Modern Icelandic (the Database of Modern Icelandic http://bin.arnastofnun.is/), but because Modern Icelandic has not gone through too many significant changes in language development, some words in Old Norse can also be found in Modern Icelandic and therefore the reader can find their inflection forms on the respective website.

The concrete annotation process primarily required expertise in Icelandic. Word and segment audio recordings were produced using the LiteDevTools audio recording portal. This required a significant amount of effort, especially the segment recording; reading of Old Norse is nontrivial, as there are certain ellipses and consonant combinations that no longer exist in Modern Icelandic. Preparation of the translations required carefully study of the meaning of individual words that either do not exist in Modern Icelandic or have changed meaning. Also, the original text included many references to other poems and Norse mythology in general. These references had to be studied in order to first understand and then select a suitable translation of individual words. Despite the fact that we had a good translation of the text in English, this was a loose translation and often included explanations or description of references rather than translation of the meaning of words and verses. It was still useful because it often suggested which way to go and made translation at a word level more efficient. As well as utilising internet sources, books from a library in Iceland were used to support the translation of individual words. The time spent on studying resources not only improved the understanding of the text and consequently creating a more meaningful translation of individual words, but also helped to create a better list of references and description of words, names and phrases used in the poem, and facilitated the creation of a names index.

In summary, the purpose of the STSM was achieved; our intuitive impression is that the Völuspá is now significantly more accessible to the general reading public. Even though there are many translations of this poem in different languages, and a considerable critical and explanatory literature, we think we are the first people to provide an interactive resource that allows people enhance their understanding of Old Norse by reading. A useful feature from this point of view is that the translation is hidden and the reader can use the functionality of LARA only to display it when needed. This is especially important for students of Medieval Studies in Old Norse and other reading public seriously interested in this language.

Our next step will be to make the LARA version of the *Völuspá* available through the popular Goodreads platform (90M members). We are particularly interested in finding people who have knowledge of Old Norse and may wish to record the text, which is currently pronounced mostly according to the conventions of modern Icelandic. Goodreads has many Scandinavian members, and the *Völuspá* is not a long text, so this does not seem an unrealistic ambition. We will also track social interaction around the poem in terms of online discussion, recommendations and other social network indicators. We will carry out the experiment during the Spring semester of 2020 and aim to write it up as an article for a conference, which could be in the area of language/cultural learning or crowdsourcing.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSM

During the STSM 21 December 2019 – 2 January 2020 (13 days), the following work was carried out:

- Reviewing the tagged text of Poetic Eddas (Elder Eddas) and selecting the poem of Völuspá, which describes the creation and the destruction of the world as portrayed in Norse mythology.
- Development of a script to convert the manually tagged version of the text into LARA format.
- Development of an enhanced LARA word token model.
- Development of a LARA mechanism to allow attaching of notes to words.
- Division of the poem text into segments, so that source text could be aligned with English translations.
- Filling in word translations for use in the word token model. This was the most laborious task, and involved studying literature on names and additional references in the original poems.
- Recording of word and segment audio using the LiteDevTools platform.
- Adding explanatory notes, other text and visual images.
- Using the above resources to compile the LARA text, and posting it on the web.
- (Soon) posting the online version of the poem to the Goodreads website, tracking resulting online interaction, and attempting to recruit people interested in contributing Old Norse audio recordings.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

This section lists the main results of the STSM. First of all, the most important result is a complete marked-up text of Völuspá in an HTML format with translation of individual words and segments including audio recordings for both. In addition, they also include a complete concordance page with references to 1) the grammar information, 2) the frequency of words, 3) the alphabetical order of words, 4) the examples of sentences in which that particular word occurs in the text, and 5) the index of names that would enable the reader to click on the displayed name or word and get additional explanation. The results moreover include a first-time hands-on experience with a new translation model based on surface token translation, that enables the translator to use the already existing segment translation to guide their translation of individual words. The main result is a fully interactive text in Old Norse that includes various kinds of information in English translation about words, names, expression and references to Norse mythology, which is also possible to listen to. This text has been designed to serve educational purpose in the group of students in Icelandic Medieval Studies.

FUTURE COLLABORATIONS (if applicable)

A) Working on another poem from the Poetic Edda, the *Hávamál*. This may require another STSM because the work behind creating such a text in LARA requires considerable attention, study and preparation.

B) Continued work with the LARA team on expanding the available context in various language including Old Norse and further development of the new functionality developed during the course of this STSM, in particular the novel translation model and the inclusion of index name information.

C) Based on the above STSM, another kind of future collaboration will centre around the organisation of a second hands-on workshop for creating language-learning materials in LARA. This will most probably be in Basel as part of an annual meeting of lecturers of Icelandic abroad.

D) We are currently in the process of writing a scientific article describing how we have used the Völuspá as an example of the preservation and transition of knowledge to a new generation of student in Old Norse and Medieval Studies. We will submit an abstract to the upcoming IASS 2020 conference in Vilnius.