

LARA: An extensible open source platform for learning languages by reading

Branislav Bédi¹, Matt Butterweck², Cathy Chua³, Johanna Gerlach⁴, Birgitta Björg Guðmarsdóttir⁵, Hanieh Habibi⁶, Bjartur Örn Jónsson⁷, Manny Rayner⁸, Sigurður Vigfússon⁹

Abstract

LARA (Learning And Reading Assistant) is an open source platform that supports conversion of plain texts into an interactive multimedia form designed to support second and foreign language (L2) learners. We illustrate the open source aspects using collaborative work carried out during a six-week summer project at the Árni Magnússon Institute for Icelandic Studies. Three BA students extended the platform in different directions in cooperation with other members of the international LARA team. The three subprojects were respectively concerned with adding automatically generated flashcards, adding multimedia versions of poetic texts in the archaic language Old Norse; and extending LARA to allow inclusion of a sign language content in Icelandic Sign Language (ÍTM). All three reached successful conclusions.

Keywords: Archaic language, flashcards, interactive text, open source, sign language.

1. Introduction

LARA¹⁰ (Learning and Reading Assistant; Akhlaghi et al., 2019) is a collaborative open source¹¹ project, active since mid-2018, whose goal is to develop tools that support conversion of plain texts into an interactive multimedia form designed to support development of L2 language skills by reading. The basic approach is in line with Krashen's influential Theory of Input (Krashen 1982), suggesting that language learning proceeds most successfully when learners are presented with interesting and comprehensible L2 material in a low-anxiety situation. LARA implements this abstract programme by providing concrete assistance to L2 learners, making texts more comprehensible to help them develop their reading, vocabulary and pronunciation skills. In particular, LARA texts include translations and human-recorded audio attached to words and sentences, and a personalised concordance constructed from the learner's reading history. The learner, just by clicking or hovering on a word, is always in a position to answer three questions: what does it mean, what does it sound like, and where have I seen it before. Figure 1 shows an example. Related platforms, from which we have adapted some ideas, include Learning With Text¹² and Clilstore¹³. LARA however offers considerably more functionality. In particular, generation of learner-specific concordances is as far as we know unique to LARA. The LARA tools are made available through a free portal, divided into two layers. The core LARA engine consists of a suite of Python modules, which can also be run stand-alone from the command-line. These are

¹ The Árni Magnússon Institute for Icelandic Studies, Reykjavík, Iceland; branislav.bedi@arnastofnun.is

² Independent researcher, Germany, MButterweck@t-online.de

³ University of Adelaide, Adelaide, Australia, Cathy.Chua@adelaide.edu.au

⁴ University of Geneva, Geneva, Switzerland, Johanna.Gerlach@unige.ch

⁵ University of Iceland, Reykjavík, Iceland, bbg14@hi.is

⁶ University of Geneva, Geneva, Switzerland; hanieh.habibi@unige.ch

⁷ University of Iceland, Reykjavík, Iceland, boj8@hi.is

⁸ University of Geneva, Switzerland, Emmanuel.Rayner@unige.ch

⁹ The Communication Centre for Deaf and Hard of Hearing, Reykjavík, Iceland, siggivig@gmail.com

¹⁰ <https://www.unige.ch/collector/lara/>

¹¹ <https://sourceforge.net/projects/collector-lara/>

¹² <https://sourceforge.net/projects/lwt/>

¹³ <http://multidict.net/clilstore/>

accessed through a web layer implemented in PHP. There is comprehensive online documentation.¹⁴

The screenshot displays a web interface for a LARA document. At the top, there is a navigation bar with a 'Table of Contents' link and a green circle labeled '1' pointing to it. Below this is a text pane containing French text from 'Le petit prince'. A green circle labeled '2' points to the phrase 'il y a' in the text. A green circle labeled '3' points to a speaker icon in the text. Below the text is an illustration of a man looking through a telescope. The lower pane shows a concordance for 'il y avoir', with a green circle labeled '4' pointing to the title. The concordance lists several sentences with 'il y a' highlighted in red. A green circle labeled '5' points to a back-arrow icon at the end of one of the concordance lines. A green box highlights the first three lines of the concordance.

Table of Contents

1

2

3

4

5

Figure 1. Example of a LARA document: *Le petit prince*. The user navigates using the controls at the top (1). The text is in the upper pane. Clicking on a word displays information about it in the lower pane. Here, the user has just clicked on part of the multiword *il y a* = “there is” (2), showing an automatically generated concordance (4). Hovering the mouse over a word plays audio and shows a popup translation. Clicking on a loudspeaker plays audio for the preceding sentence (3). The back-arrows (5) link each line in the concordance to its context of occurrence. A link to the document can be found on the LARA content page¹⁵.

¹⁴ <https://www.issco.unige.ch/en/research/projects/collector/LARADoc/build/html/index.html>

¹⁵ <https://www.unige.ch/collector/lara-content/>

In this paper, we will concentrate on the open source aspects. We illustrate using work carried out during a six-week summer project at the Árni Magnússon Institute for Icelandic Studies, where three BA students, who had previously not worked with LARA, extended the platform in different directions with some assistance from other members of the LARA team. The three subprojects were respectively concerned with adding automatically generated flashcards; creating multimedia versions of poetic texts written in the archaic language Old Norse; and extending LARA to allow inclusion of sign language content. In sections 2 to 4, we briefly sketch the three subprojects. The final section summarises and concludes.

2. Extending LARA with flashcards

A common suggestion we have received from LARA users is that it would be useful to make the platform more interactive and include functionality that allows learners to test their understanding of a text. The first subproject, carried out by a student who had just completed a BSc in computer science, addressed this idea by adding capabilities to create flashcards automatically extracted from a LARA text. A member of the core LARA team first wrote a toy version of the flashcard module in Python, showing how to extract the necessary information from the internalised form. The student then worked autonomously, except for a couple of requests for low-level functions to obtain other types of internal information. Finally, the flashcard module was incorporated into the web layer by another member of the core LARA team, working together with the student.

The final version of the flashcard module supports five different kinds of flashcards. A new set of flashcards is generated from the text each time the functionality is accessed; the choice of examples is random, but the “distractors” (incorrect answers) are constrained so that they are made as similar as possible to the correct answer. An example of the simplest kind of flashcard is shown in Figure 2. Examples of other kinds of flashcards are shown in the sections on Old Norse and sign language further down.

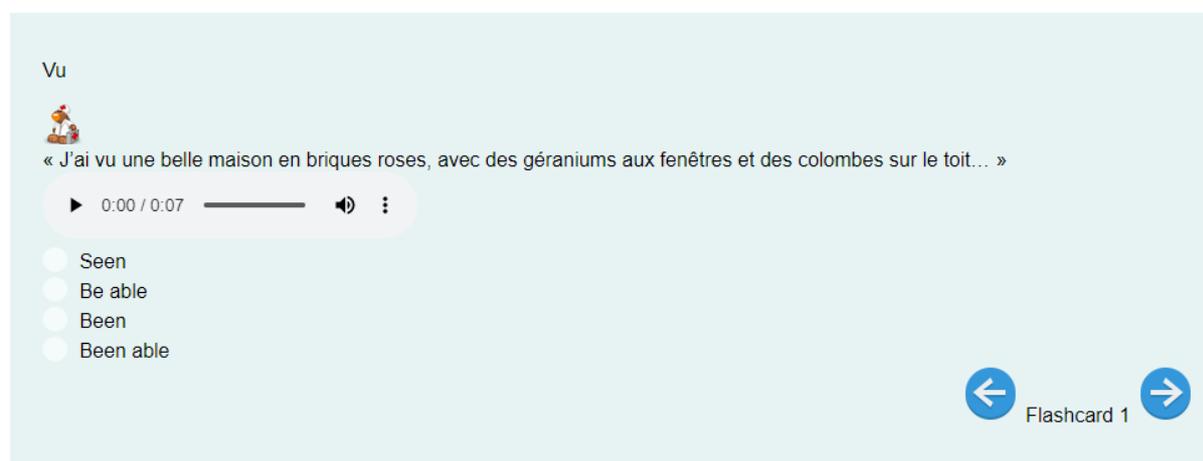


Figure 2. Example of LARA flashcard for *Le petit prince*. The student has to pick the most appropriate translation for the word presented at the top, “Vu”, out of the four alternatives. The context for the word is presented in both text and audio form.

3. Using LARA for Old Norse

In most countries, students at middle schools are required to read classic works of literature that play an important part in the relevant country’s cultural history: English children read Shakespeare, French children Molière, etc. The archaic language of the texts is in general not fully comprehensible to the students without some explanation. Our second subproject was

designed to see if LARA could provide assistance in this kind of situation. In Iceland, the appropriate culture referent is the Poetic Edda, a poem-cycle first written down in the late 13th century but composed earlier. The Edda is written in Old Norse, the language spoken in Iceland between the 8th and 14th centuries, from which Modern Icelandic has developed. Old Norse is much closer to Modern Icelandic than English is to Old English, but still displays substantial differences: the grammar is not exactly the same, many words have shifted in meaning or have different spelling, and some have disappeared.

The screenshot displays the LARA interface for the Old Norse text *Völuspá*. At the top, there are navigation controls (1) and an audio player (2). The main content area shows the text in both original Old Norse (3) and modern orthography (4). Words in red (5) are linked to informative notes. A word, *skein*, is highlighted, and its information page for the lemma *skína* (6) is displayed. A runic symbol (7) is also shown. Below the text, there are links to grammar information (8), a lexicon entry (8), a concordance (9), notes (10), and frequency and alphabetical indexes (11).

Figure 3. Example of LARA document in Old Norse (*Völuspá*). showing navigation controls (1); recorded audio for each verse (2); text in both original (3) and modern orthography (4); words in red linked to informative notes (5); clicking on a word, here *skein*, displays the information page for the lemma, here *skína* (6); runic symbol (7) displays translation of verse; automatically generated links (8) to online resources; automatically generated concordance (9); list of notes (10); frequency and alphabetical indexes (11). Hovering the mouse over a word plays audio and shows a popup translation (not featured).

The student responsible for the subproject, who had just completed her second year of a BA degree in Icelandic, used the LARA portal to create versions of the two best-known Edda

poems, *Völuspá* and *Hávamál*; in contrast to the other subprojects, this did not involve developing any new platform functionality. The poems are annotated with glosses in Modern Icelandic and read with adapted Old Norse pronunciation. Key words and phrases, most often names of gods and places, are linked to explanatory notes. An interesting aspect concerns *kennings*, poetic phrases typically of two or three not necessarily contiguous words characteristic of the Edda and related Old Norse poems. These could successfully be handled by the multiword annotation scheme illustrated in Figure 1, a use of this mechanism we had not anticipated. An example of a page from an Edda text is shown in Figure 3, and a ‘sentence with gap’ flashcard in Figure 4. This work is described at greater length in another paper from this volume.

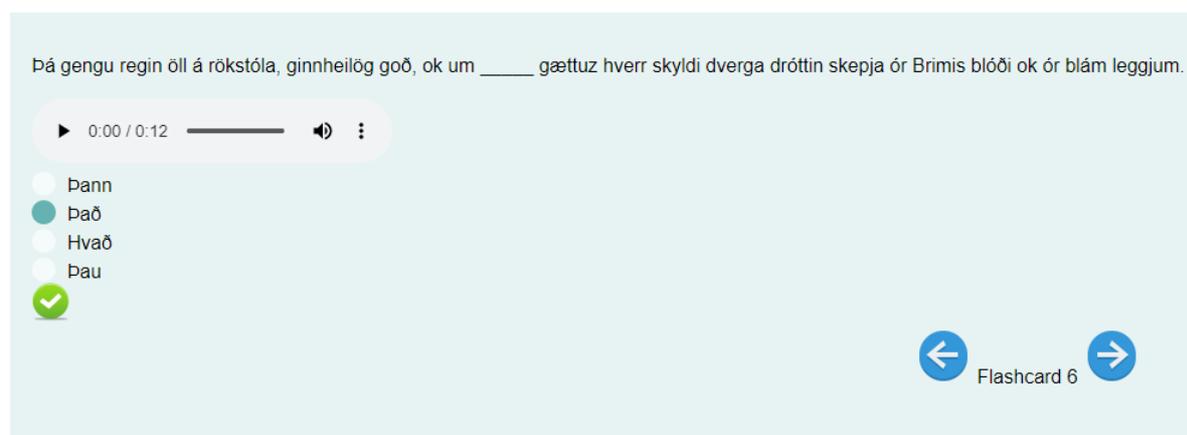


Figure 4. Example of LARA flashcard for *Völuspá*. The student has to find the missing word in the incomplete verse presented at the top. After they have answered correctly, they can listen to the whole verse using the audio control.

4. Adding sign language to LARA documents

The theme of the third subproject was signed language. Here, the intention was to create an initial version of a LARA text designed for Deaf learners. The starting point was an existing LARA text for an Icelandic children’s story, *Tína fer í fri*; this had been constructed for a previous experiment (Bédi et al 2019), where it had been used by beginner/intermediate L2 learners in an Icelandic-for-foreigners course. In the current project, we repurposed the text so that it could be used by Deaf signers of ÍTM (*íslenskt táknmál*, Icelandic sign language) who wished to strengthen their Icelandic reading skills. Like all sign languages, ÍTM has no grammatical connection with the surrounding oral/aural language, here Icelandic. It is thus by no means assured that native signers of ÍTM will have strong reading skills in Icelandic, which for them is a second language.

As with the other two subprojects, core members of the LARA team did a small amount of preparatory work, generalising the treatment of multimedia so that word and sentence annotations could be supplied in video as well as audio form. The rest of the project was performed by the student, who had just completed a BA degree in ÍTM and translation. He created signed videos for the words and sentences in the text, using the online recording tool integrated with LARA, after which the LARA platform scripts downloaded and linked everything together to create the final document. The signed video extension was incorporated into the flashcard module developed during the first subproject. Examples of LARA pages and flashcards for ÍTM are shown in Figures 5 and 6. This work will be presented at greater length elsewhere.

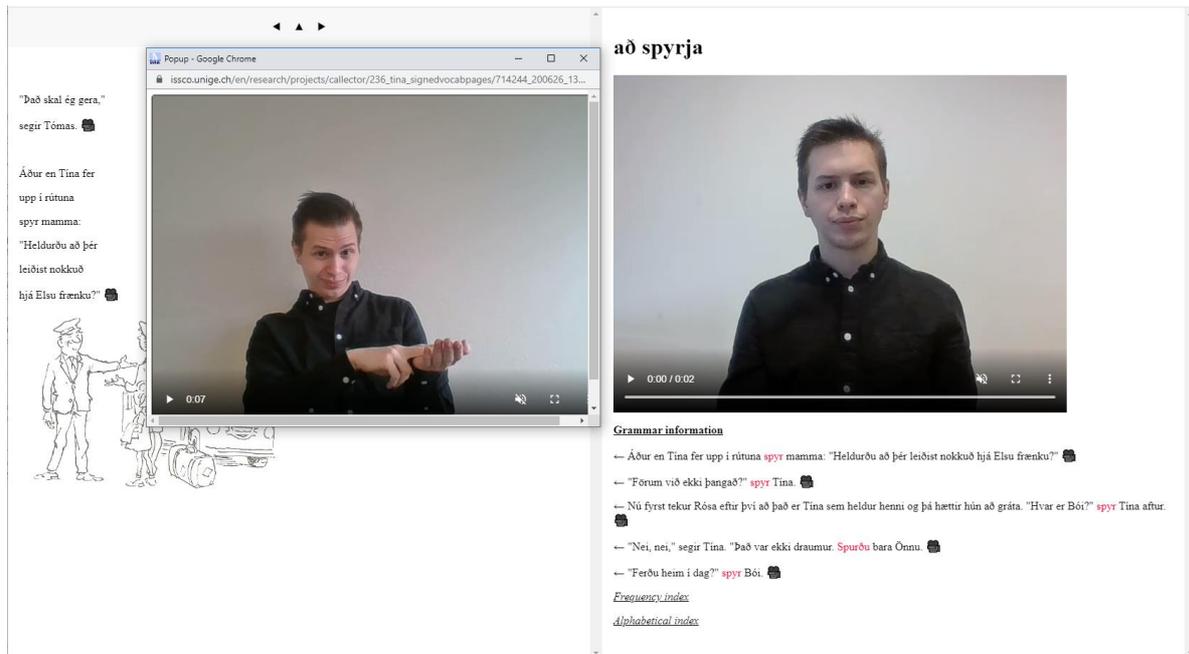


Figure 5. Example of LARA document in Icelandic with ÍTM annotation. Controls are similar to figures 1 and 3. Concordance pages contain signed video for the word in question (right); clicking on a camera icon pops up a signed video for the preceding sentence (left).

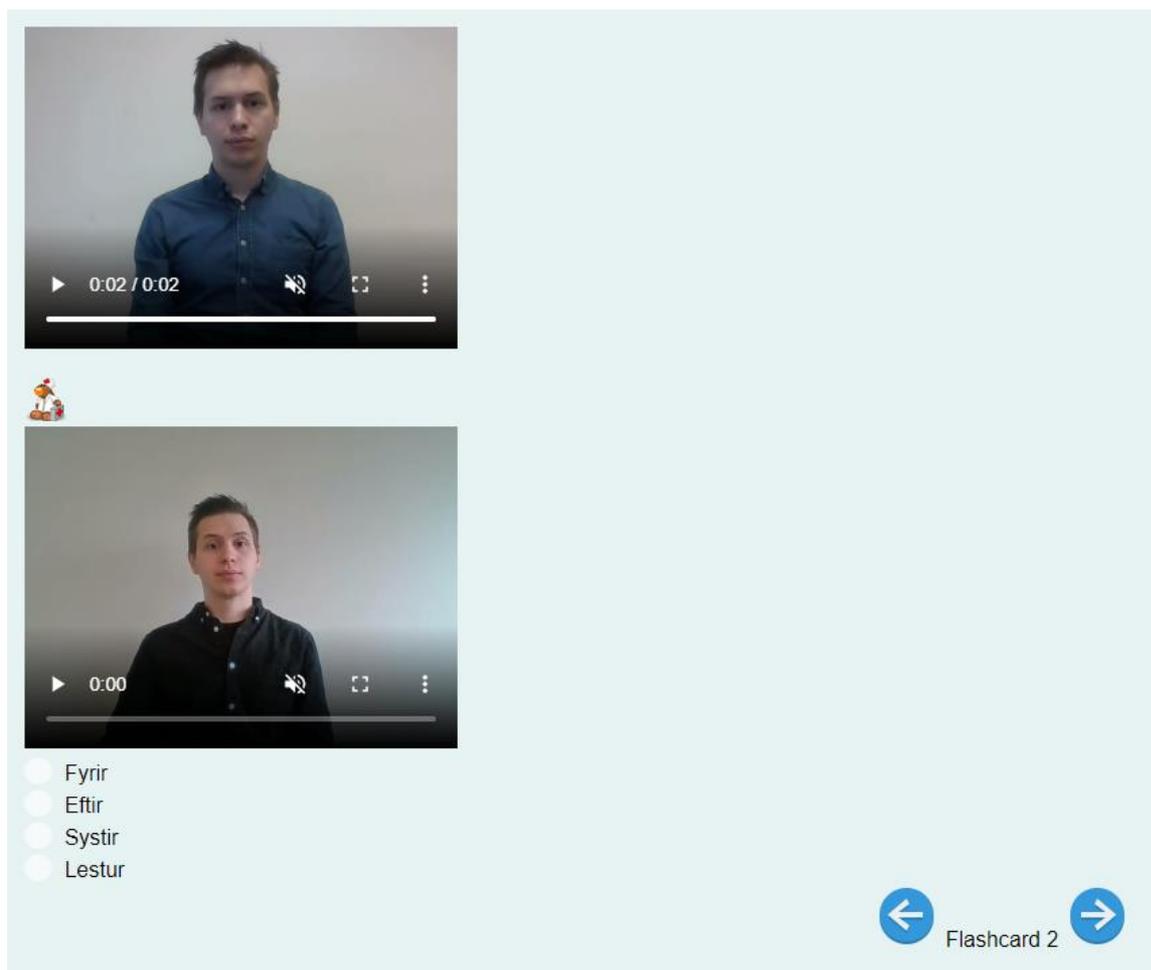


Figure 6. Example of flashcard for signed LARA text. The upper video is the question and the lower video is the context.

5. Summary and further directions

We have briefly described three projects where students without previous exposure to LARA extended it in different directions over a six-week period. Initially, we were not certain that the students would be able to complete this ambitious program in such a short time. We are encouraged by the successful result and are already planning further collaboration. If you are interested in developing other open source extensions to LARA and would like assistance, please feel free to contact us.

6. Acknowledgements

Work in Iceland was funded by The Icelandic Centre for Research (RANNÍS). Work in Switzerland was funded by the Swiss National Science Foundation and the University of Geneva.

References

- Akhlaghi, E., Bédi, B., Butterweck, M., Chua, C., Gerlach, J., Habibi, H., Ikeda, J., Rayner, M., Sestigiani, S., and Zuckermann, G. (2019). Overview of LARA: A learning and reading assistant. In Proceedings of SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education, pages 99–103, 2019.
- Bédi, B., Chua, C., Habibi, H., Martinez-Lopez, R., and Rayner, M. (2019). Using LARA for language learning: a pilot study for Icelandic. *CALL and complexity—short papers from EUROCALL 2019*, page 33, 2019.
- Krashen, S.D. (1982). *Principles and Practice in Second Language Acquisition*. First Internet Edition July 2009. Oxford: Pergamon.