

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, and Sigurður Vigfússon⁹

Abstract. Learning and Reading Assistant (LARA) is an open source platform that enables conversion of plain texts into an interactive multimedia form designed to support second- and foreign-language (L2) learners. In this workshop, 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 undergraduate level 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 the inclusion of sign language content in Icelandic sign language – Íslenskt TákMál (ÍTM). All three reached successful conclusions.

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

1. The Árni Magnússon Institute for Icelandic Studies, Reykjavík, Iceland; branislav.bedi@arnastofnun.is; <https://orcid.org/0000-0001-7637-8737>
2. Independent researcher, Hagen, Germany; mbutterweck@t-online.de
3. University of Adelaide, Adelaide, Australia; cathy.chua@adelaide.edu.au; <https://orcid.org/0000-0002-5462-2441>
4. University of Geneva, Geneva, Switzerland; johanna.gerlach@unige.ch; <https://orcid.org/0000-0002-3371-4021>
5. University of Iceland, Reykjavík, Iceland; bbg14@hi.is; <https://orcid.org/0000-0002-5032-6303>
6. University of Geneva, Geneva, Switzerland; hanieh.habibi@unige.ch; <https://orcid.org/0000-0001-7602-2933>
7. University of Iceland, Reykjavík, Iceland; boj8@hi.is
8. University of Geneva, Geneva, Switzerland; emmanuel.rayner@unige.ch; <https://orcid.org/0000-0002-6002-2434>
9. The Communication Centre for Deaf and Hard of Hearing, Reykjavík, Iceland; siggivig@gmail.com; <https://orcid.org/0000-0002-7258-1209>

How to cite: Bédi, B., Butterweck, M., Chua, C., Gerlach, J., Guðmarsdóttir, B. B., Habibi, H., Jónsson, B. Ó, Rayner, M., & Vigfússon, S. (2020). LARA: an extensible open source platform for learning languages by reading. In K.-M. Frederiksen, S. Larsen, L. Bradley & S. Thoušný (Eds), *CALL for widening participation: short papers from EUROCALL 2020* (pp. 27-35). Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.48.1160>

1. Introduction

LARA¹⁰ (Akhlaghi et al., 2019) is a collaborative open source¹¹ project, active since mid-2018, whose goal is to develop tools that enable 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 (1982) influential theory of input, 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 over 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 accessed through a web layer implemented in PHP¹⁴. There is comprehensive online documentation¹⁵.

In this paper, we will concentrate on the open source aspects. We illustrate this 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.

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/>

14. <https://www.php.net/>

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

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 page¹⁶

◀ ▶ ▶ Table of Contents (1)

Ça ne pouvait pas m'étonner beaucoup. Je savais bien qu'en dehors des grosses planètes comme la Terre, Jupiter, Mars, Vénus, auxquelles on a donné des noms, il y en a des centaines d'autres qui sont quelquefois si petites qu'on a beaucoup de mal à les apercevoir au télescope. Quand un astronome découvre l'une d'elles, il lui donne pour nom un numéro. Il l'appelle par exemple : « l'astéroïde 3251. » (3)

J'ai de sérieuses raisons de croire que la planète d'où venait le petit prince est l'astéroïde B 612. Cet astéroïde n'a été aperçu qu'une fois au télescope, en 1909, par un astronome turc.

il y avoir (4)

← Et en effet, sur la planète du petit prince, **il y avait** comme sur toutes les planètes, de bonnes herbes et de mauvaises herbes.

← Or **il y avait** des graines terribles sur la planète du petit prince... c'étaient les graines de baobabs.

← **Il y avait**, sur une étoile, une planète, la mienne, la Terre, un petit prince à consoler !

← **Il n'y a** pas de tigres sur ma planète, avait objecté le petit prince, et puis les tigres ne mangent pas l'herbe.

← Mais **il n'y a** personne à juger ! - On ne sait pas, lui dit le roi.

← **Il n'y en a** qu'un. (5)

← **Il n'y a** rien à comprendre, dit l'allumeur. La consigne c'est la consigne. Bonjour.

← Ici c'est le désert. **Il n'y a** personne dans les déserts.

← **Il y a** une fleur... je crois qu'elle m'a apprivoisé...

← **Il me semblait** même qu'**il n'y eût** rien de plus fragile sur la Terre.

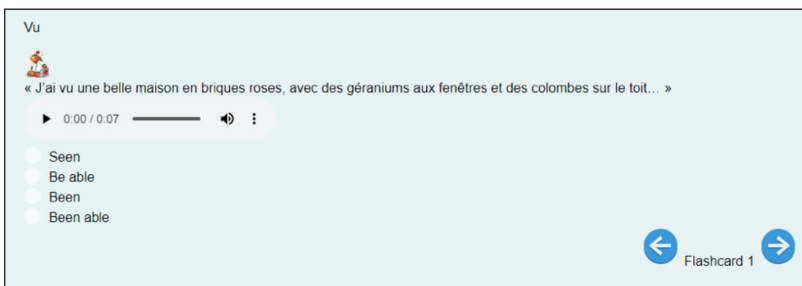
16. <https://www.unige.ch/collector/lara-content/>

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 Bachelor of Science 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 spellings, and some have disappeared.

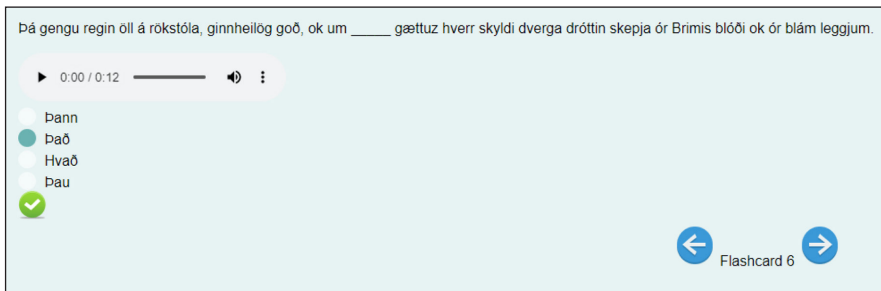
Figure 3. Example of LARA document in Old Norse (*Völuspá*) showing navigation controls (1); recorded audio for each verse (2); text in both original (3) and modern orthography (4); words in red are linked to informative notes (5); clicking on a word, here *skein*, displays the information page for the 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 [Bédi et al. \(2020, this volume\)](#).

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 sign 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 story 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 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 and audio forms. 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 Figure 5 and Figure 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 opens a signed video for the preceding sentence (left)

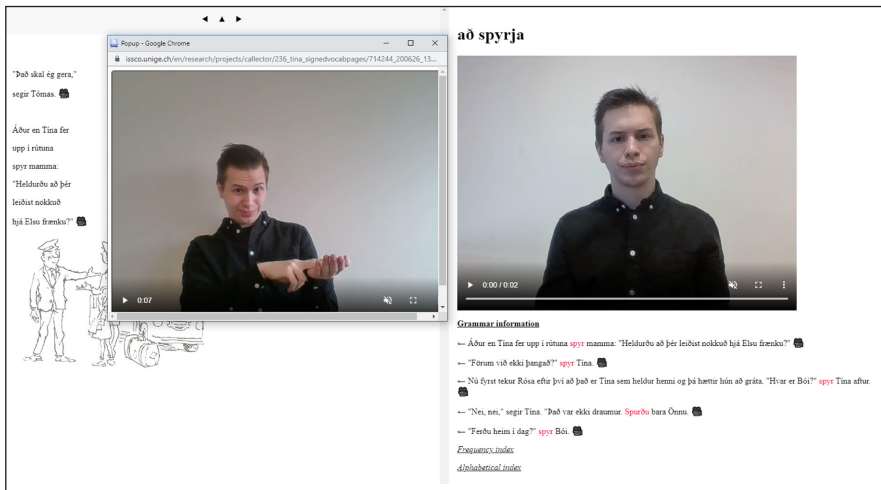
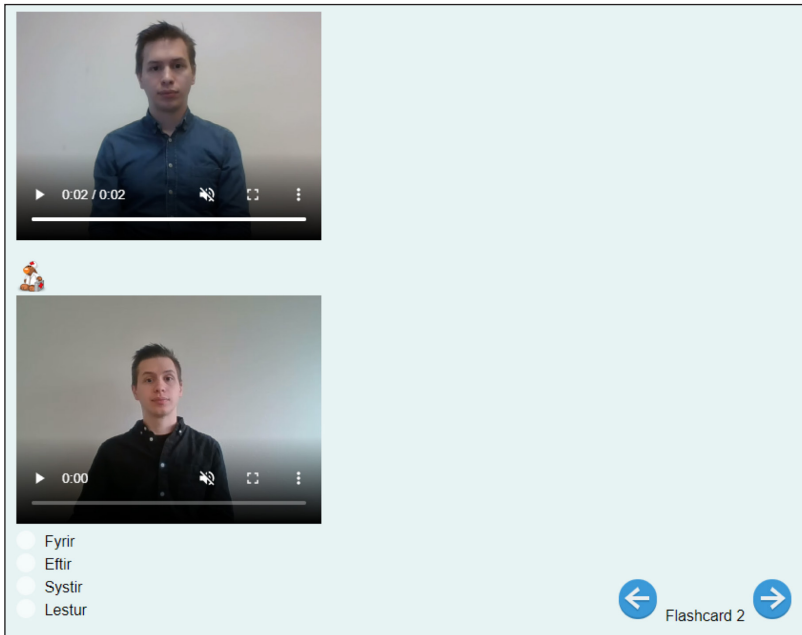


Figure 6. Example of flashcard for signed LARA text: the upper video is the question, and the lower video is the context



5. Summary and further directions

We have briefly described three summer projects where students without previous exposure to LARA extended it in different directions over a six-week period. This ambitious program was completed in half the time that was originally planned; encouraged by the successful results, we envisage further collaboration with the same and new collaborators. If you are interested in developing other open source extensions to LARA and need assistance, please feel free to contact us at the addresses given above.

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., & 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* (pp. 99-103). <https://doi.org/10.21437/slate.2019-19>
- Bédi, B., Bernharðsson, H., Chua, C., Guðmarsdóttir, B. B., Habibi, H., & Rayner, M. (2020). Constructing an interactive Old Norse text with LARA. In K.-M. Frederiksen, S. Larsen, L. Bradley & S. Thouéšny (Eds), *CALL for widening participation: short papers from EUROCALL 2020* (pp. 27-35). Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.48.1159>
- Bédi, B., Chua, C., Habibi, H., Martinez-Lopez, R., & Rayner, M. (2019). Using LARA for language learning: a pilot study for Icelandic. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouéšny (Eds), *CALL and complexity – short papers from EUROCALL 2019* (pp. 33-38). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.38.982>
- Krashen, S. D. (1982). Principles and practice in second language acquisition [First Internet Edition July 2009]. Pergamo

Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

© 2020 by Editors (collective work)
© 2020 by Authors (individual work)

CALL for widening participation: short papers from EUROCALL 2020
Edited by Karen-Margrete Frederiksen, Sanne Larsen, Linda Bradley, and Sylvie Thouéšny

Publication date: 2020/12/14

Rights: the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence.** Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2020.48.9782490057818>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net

Cover theme by © 2020 Marie Flensburg (frw831@hum.ku.dk), based on illustration from freepik.com
Cover layout by © 2020 Raphaël Savina (raphael@savina.net)

ISBN13: 978-2-490057-81-8 (Ebook, PDF, colour)

British Library Cataloguing-in-Publication Data.

A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2020.