



Demonstration of LARA: A Learning and Reading Assistant

Elham Akhlaghi¹, Branislav Bédi², Matthias Butterweck³, Cathy Chua³, Johanna Gerlach⁴
Hanieh Habibi⁴, Junta Ikeda³, Manny Rayner⁴, Sabina Sestigiani⁵, Ghil'ad Zuckermann⁶

¹Ferdowsi University of Mashhad, Iran; ²The Árni Magnússon Institute for Icelandic Studies, Iceland; ³Independent scholar; ⁴FTI/TIM, University of Geneva, Switzerland; ⁵Swinburne University, Australia; ⁶University of Adelaide, Australia

elham.akhlaghi@mail.um.ac.ir, branislav.bedi@arnastofnun.is, matthias@butterweck.de, cathy@pioneerbooks.com.au, Johanna.Gerlach@unige.ch, hanieh.habibi@unige.ch, ikedaj_91@hotmail.com, Emmanuel.Rayner@unige.ch, ssestigiani@swin.edu.au, ghilad.zuckermann@adelaide.edu.au

Abstract

We propose to demo LARA (Learning and Reading Assistant), a set of tools currently being developed in the context of a collaborative open project for building and using online computer-assisted language learning (CALL) content. LARA offers a range of options for semi-automatically transforming text into a hypertext version designed to give support to non-native readers. The demo is intended to accompany our full paper about LARA; here we focus on describing some of the content we will present.

Index Terms: CALL, reading, hypertext, open source

1. Introduction

LARA (Learning and Reading Assistant; <https://www.unige.ch/callector/text-content/>) is a collaborative open project, initiated during the second half of 2018, whose goal is to create resources via crowdsourcing techniques [1] that can help learners read texts in foreign/archaic languages. It does this by providing tools that make it easy to transform plain text documents into hypertext versions that provide non-native readers with various kinds of support. These including audio recordings of words, sentences and longer passages, translations, grammar information, and a personalised hyperlinked concordance based on the student's own reading history.

LARA is formally presented in a full paper by the same authors, submitted to this conference [2]. Here, we restrict ourselves to describing some of the pieces of LARA content we propose to show. Our demonstration of LARA's generic capabilities will particularly focus on the following: 1) rapid construction of hyperlinked text content using the LARA tools, and 2) continual updating of the hyperlinked concordance driven by the learner's reading progress.

2. Content

The content we propose to show is summarised in Table 1. In the rest of this section, we describe it in more detail.

2.1. *Tína fer í frí* (Branislav Bédi)

We use *Tína fer í frí*, an Icelandic translation of a Danish children's book by Esther Skriver, to introduce LARA. Figure 1 presents a screenshot. The text is on the left hand side of the screen. The student can hover the mouse over any word to hear an audio recording and see a popup translation. They have just clicked on the word *grátandi* ("crying"), which has brought

Table 1: LARA content we propose to demonstrate. "Lng" = language; "#Seg" = number of segments; "#Tok" = number of surface word tokens; "#Typ" = number of lemma types; "Link" = link to online LARA resource.

Text	Lng	#Seg	#Tok	#Typ	Link
Tína	IS	207	2743	466	
Inferno	IT	135	840	395	
Hy. Ikita Neko	JP	88	967	225	
Arash	FA	34	329	128	
Revivalistics	HE	69	129	124	
Alice	EN	1478	26868	2007	
Nibelungenlied	DE	11937	81853	2826	

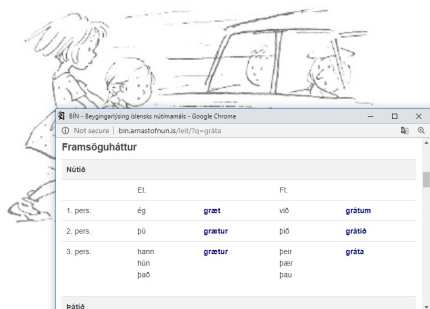
up the concordance page, on the right, for *að gráta*, "to cry". Note that the page contains examples featuring three different inflected forms of *að gráta*, highlighted in red. The student then clicked on Grammar information, which opened the popup window visible on the lower left. This accesses an online Icelandic grammar resource and fetches the conjugation page for *að gráta*.

The LARA version of *Tína fer í frí* is currently being used as a resource in an Icelandic-Practical-Diploma course held at the University of Iceland.

2.2. Inferno (Sabina Sestigiani)

Inferno consists of three long extracts from Canti I, III and V of Dante's *Inferno*, the most famous work of poetry in the Italian language. The text is presented so that students can listen to audio at three levels: individual words, single lines, and sections consisting of multiple lines. Similarly, they can access translations both at the word and at the line level. The line versions are taken from Longfellow's 1867 translation.

Inferno is currently being used in an upper-intermediate Italian class at Swinburne University, Australia. Although the fourteenth-century Italian would normally be beyond the capabilities of any but the best students in this class, the LARA version makes it generally accessible. Students can begin by using the word- and line-level support to familiarise themselves with the text, then listen to the long audio recordings to appreciate the beauty of the language.



"Ég vil fara í rauðu rútunni
heim til mömmu,"
segir Rósa grátandi
og reynir að losa sig.

að gráta

Grammar information

← "Já," segir Bói, "hún fór að gráta af því að hún fékk ekki að fara í fri. Þá var henni leyft að fara." Rósa er ánægð. "Ég er svo stór að ég get farið í fri ein. Ekki með pabba og mömmu. Ég er nefnilega 4 ára."

← En áður en Tina kemst að stóra, gula tjaldinu heyrir hún barn gráta. "Mamma, mamma," hrópar barnið og grætur hástöfum.

← "Ég vil fara í rauðu rútunni heim til mömmu," segir Rósa grátandi og reynir að losa sig.

← Nú fyrst tekur Rósa eftir því að það er Tina sem heldur henni og þá hættir hún að gráta. "Hvar er Bói?" spyr Tina aftur.

← Tina er nærri farin að gráta.

Figure 1: Example of LARA content: a passage from *Tína fer í fri* showing concordance and grammar information.

2.3. Hyakumankai Ikita Neko (Junta Ikeda)

Hyakumankai Ikita Neko (100 万回生きたねこ; "The Cat Who Lived a Million Times") illustrates LARA's ability to handle non-European scripts. The text is a well-known children's book by Yōko Sano. It is an excellent first reader for Japanese students with one or two years of experience; the cyclical structure helps reinforce vocabulary and grammar (the selfish cat hero is repeatedly reincarnated until he finally escapes the Wheel of Karma at the end), and LARA lets unconfident students who might otherwise give up too soon get a foothold in the text.

This resource also shows that LARA itself is easy to learn. Junta Ikeda, who had no previous contact with the project, constructed it in three or four evenings mostly using the online documentation [3] with a couple of brief Skype conversations to discuss issues concerning Japanese tagging.

2.4. Arash (Elham Akhlaghi and Hanieh Habibi)

Arash is one of a series of short Farsi LARA texts constructed for the Ferdowsi University of Mashhad's course on Farsi as a foreign language; it shows that LARA is also capable of handling right-to-left scripts. The resources, used by Russian- and Arabic-speaking students, are similar in nature to the Icelandic resource described in §2.1.

2.5. Revivalistics (Ghil'ad Zuckermann)

Although the original purpose of LARA was to help learners read L2 texts, it also seems useful for presenting mixtures of first-language (L1) and L2 text of the kind found in linguistics papers, language textbooks and similar. This example illustrates, using a four page extract from a forthcoming linguistics book in which the phonetic aspects of language are salient. The text was annotated to indicate that the English text should be left unchanged and passages in the various other languages should be processed by LARA; audio was then recorded for each such phrase. The upshot is that the text could easily be transformed into a version where the reader is able to listen to any non-English phrase by hovering the mouse over it.

2.6. Alice (Cathy Chua and Manny Rayner)

This item represents a currently ongoing attempt to produce a LARA version of a full-length book, Carroll's *Alice in Wonderland*. The text has been fully annotated, using automatic tagging with a tagger built from resources provided by the Python NLTK package followed by manually cleaning; audio is currently being recorded, and is about 60% done. There are so far no translations. Given the current rate of progress, we expect the project to be complete by the time of the SLATE workshop.

2.7. Nibelungenlied (Matt Butterweck)

The last item, and also the most ambitious one, is another attempt to transform a full-length book into LARA form. *Das Nibelungenlied*, sometimes called "the German Iliad", is a heroic poem of about 12,000 lines written in Middle High German. It was tagged using the TreeTagger package for Middle High German (TreeTagger is integrated into LARA) and manually cleaned. Segment translations in High German were added from a publicly available source. There is embedded audio for one representative section.

This resource is also interesting because it shows that users with coding skills can extend LARA to include new features. Specifically, Matt Butterweck, who is carrying out the project, has added the possibility of dynamically highlighting text as the audio is played, advancing the highlighted line so that it is in sync with the audio. This feature may soon be included in other LARA resources.

3. References

- [1] V. Lyding, L. Nicolas, B. Bédi, and K. Fort, "Introducing the European NETWORK for COmbining Language LEarning and Crowdsourcing Techniques (enetCollect)," *Future-proof CALL: language learning as exploration and encounters—short papers from EURO-CALL 2018*, p. 176, 2018.
- [2] E. Akhlaghi, B. Bédi, M. Butterweck, C. Chua, J. Gerlach, H. Habibi, J. Ikeda, M. Rayner, S. Sestigiani, and G. Zuckermann, "Overview of LARA: A Learning and Reading Assistant," in *Proc. SLATE 2019*, Graz, Austria, 2019.
- [3] M. Rayner, H. Habibi, and M. Butterweck, *Constructing LARA content*, <https://www.issco.unige.ch/en/research/projects/callector/LARADoc/build/html/index.html>, 2019, online documentation.