

# Using LARA for learning Icelandic

Branislav Bédi<sup>1</sup>, Cathy Chua<sup>2</sup>, Hanieh Habibi<sup>3</sup>, Ruth Martinez-Lopez<sup>4</sup>, and Manny Rayner<sup>5</sup>

---

## Abstract

This paper presents a brief overview of LARA (Learning and Reading Assistant), an open source online tool that has been under development since Summer 2018. The central goal is to support reading comprehension while learning an L2 language. Content as of mid-July 2019 consists of a corpus of seventeen texts in ten languages and with a crowdsourced model is continuing to expand. The focus here is on the development of Icelandic content and its use during pilot testing amongst adult learners in classes of Icelandic as an L2 language. Preliminary feedback from users is reported. This was mostly positive whilst suggesting ways in which the tool can be improved.

**Keywords:** crowdsourcing, open source, pilot testing, reading.

---

## 1. Introduction

LARA is a free online tool which makes it easy to transform plain texts into a hyperlinked multimedia form designed to support non-native readers: it thus connects content providers, such as teachers, with readers/learners. When accessing LARA content, the user sees a split screen with the text on one side and various options on the other. The unique feature of the interface is the personalised concordance created for each reader-user. That is, the reader's first point of reference for questions such as 'what is that word?' is their own history which will show if and how often it has been seen before, as well as the context in which it appeared. Other functionalities include dictionary translations, voice recordings of words and passages, links to online grammar resources and vocabulary lists ordered both alphabetically and by frequency of occurrence in the text. All of these can be accessed by pointing and clicking. Figure 1 illustrates using our initial Icelandic resource, *Tína fer í frí* (Skriver, 1981), a children's book containing about 2700 words. To put this in context, the total LARA content repository, as of mid-July 2019, consists of seventeen texts in ten languages, with a total of about 130,000 words. A detailed overview of LARA appears in Akhlaghi et al. (2019).

Organisationally, LARA is a part of the open source CALLector project<sup>6</sup>, which in turn is closely linked to the enetCollect<sup>7</sup> COST network. Ethical issues are central to both LARA and CALLector. The users, whether teacher/content developer or learner/content user are in control of their data, and the platform/tools are designed to be self-sufficient through best practice of crowd-sourced free models. In terms of software engineering, the key medium-term goals are to make the content development process simple and user-friendly, with only basic computer skills required, and embed the platform in a dedicated social network which links together content constructors and content users. Although initial LARA content is already being trialled by classroom teachers in Iceland and elsewhere, these goals have still only been partially achieved. In particular, the social network level is still not available; the tools still need to be downloaded and installed on a local machine; and for some languages, currently including Icelandic, the process of annotating the text still requires laborious manual tagging of surface words by head-word. (Automatic tagging is supported for many other languages). Work on all these issues is progressing. A first version of a LARA web portal is currently being tested, which will obviate the need to install software; taggers for other languages, in particular, an open source Icelandic tagger, are being integrated; and we expect the social network to go live towards the end of 2019.

LARA's pedagogical model builds on Palincsar and Brown's (1984) reading comprehension, Oxford's (1990) reading strategy, and on the early example of Johns's (2002) Data-Driven Learning (DDL). Based

---

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

<sup>2</sup> Independent scholar, Geneva, Switzerland; cathy@pioneerbooks.com.au

<sup>3</sup> FTI/TIM University of Geneva, Geneva, Switzerland; Hanieh.Habibi@unige.ch

<sup>4</sup> Seville University, Seville, Spain; ruthmart@gmail.com

<sup>5</sup> University of Geneva, Switzerland, Emmanuel.Rayner@unige.ch

<sup>6</sup> <https://www.unige.ch/callector/>

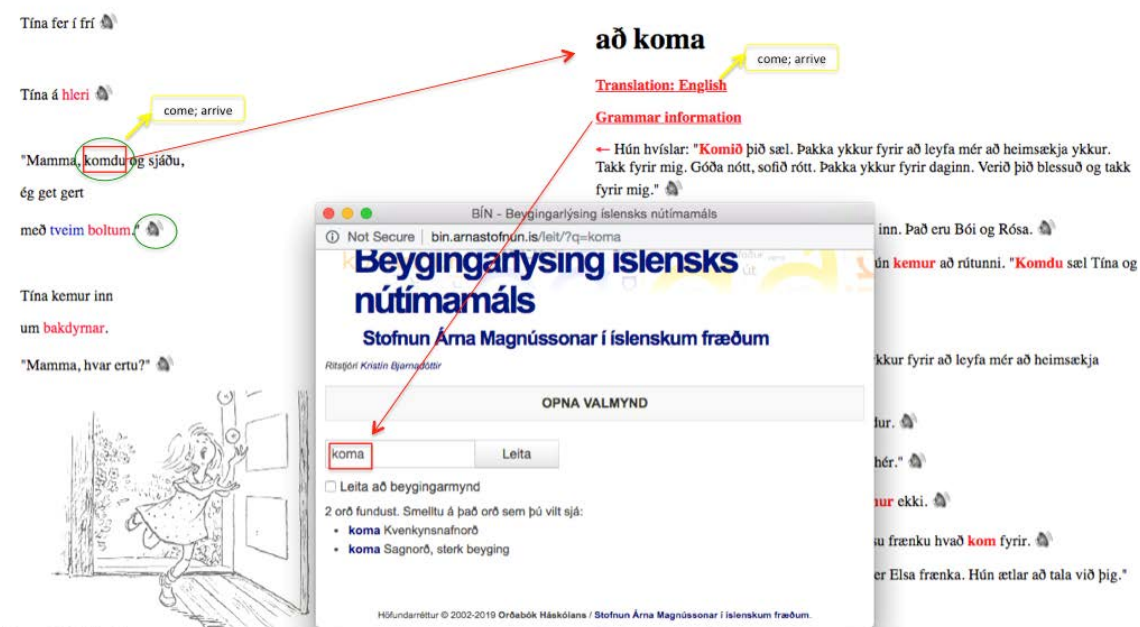
<sup>7</sup> <http://enetcollect.eurac.edu>

on these sources we can view reading comprehension as a product of three main factors: a) providing learners with considerate texts that are suitable to their level, b) making the content compatible with the reader's knowledge and c) supporting active strategies which readers may employ to enhance understanding and retention, and to facilitate comprehension. DDL from the start has been compatible with the use of computers in L2 education. In particular it provides a framework for allowing a mechanical learning assistant to make available a textual corpus and associated language expertise, with the goal of eventually supporting scaffolding and learner autonomy (Boulton and Tyne, 2013; Corino and Onesti 2019).

The use of corpora in L2 education is still relatively rare. Some of the most known online corpus analysis tools with open access are the NoSketchEngine (Rychlý, 2007), SkELL (Baisa and Suchomel, 2014), AntConc (Anthony, 2019), and LexTutor (Cobb, 2019). The commonality of these tools are crowdsourced corpora consisting of online texts in various languages; information about frequency of words; supporting language learning by finding meanings of words, phrases, idioms and other language expressions in context.

Next we describe the methodology of pilot testing of the Icelandic content and present some preliminary results.

Figure 1. LARA functionality for Icelandic text. Featured are translation tags, grammar information, audio icons, contextual word reference in the corpus.

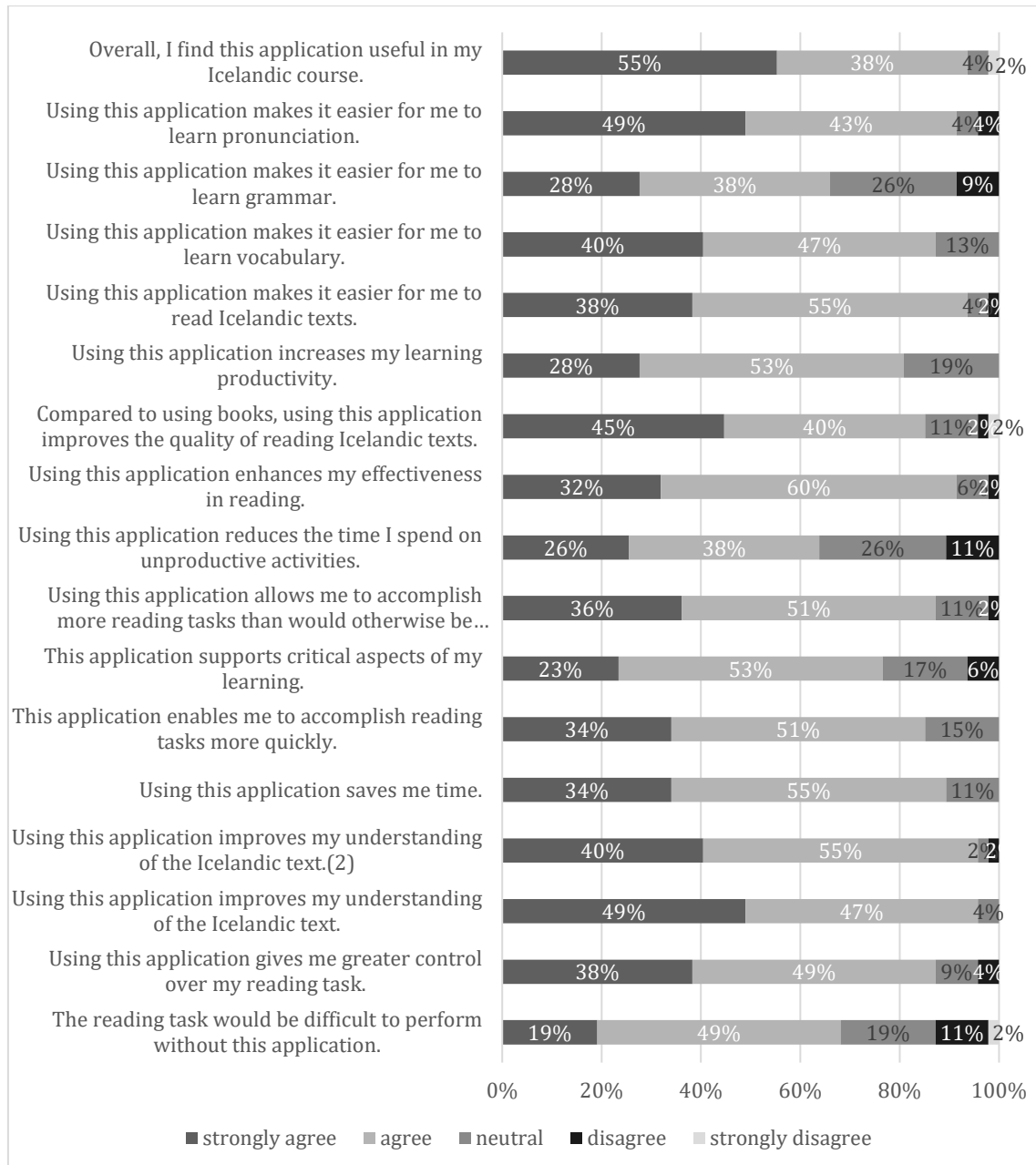


## 2. Pilot testing of LARA for learning Icelandic

The pilot test was organised around an anonymous questionnaire in March, 2019. There were forty-seven participants consisting of students of the Icelandic Practical Diploma at the University of Iceland (twenty-one female, twenty-three male, one other, two prefer not to say; thirty-one beginners, eight intermediate; seventeen countries;). After a short introduction to the tool, they spent sixty minutes using it, and were then given twenty minutes to respond to the questionnaire. It had thirty-one questions about participants' background and perceived usefulness with Likert-scale answers (Davis, 1989), and open-ended questions about assessment of digital tools (Nesbitt, 2013). The answers about the users' general experience and learning effect are summarised in Figure 2. Most of the learners either strongly agreed or agreed with positive statements about LARA. Compared with the traditional way of reading a book, most users said that the tool's functionalities increased their efficiency in reading with understanding, learning vocabulary and (to some extent) learning grammar. Nearly all students said that the integrated audio helped improve pronunciation.

Results from open-ended questions suggest that over 90% of the learners liked the application. Problems experienced included poor internet connection, missing or inaccurate word translations and lack of support for mobile devices. Currently Chrome is the only browser permitting full functionality. About 70% suggested improvements, e.g. including instructions before the task, on-screen instructions to help navigate, options for both female male voices, adding a bookmark, and options for changing font/background colour. About the same proportion of learners liked the design. About 81% of learners said that the application met their needs; 11% suggested improvements like adding more explanation about grammar, sentence structure, vocabulary and phrases. 8% said it didn't meet their needs.

Figure 2. Learners' perceptions regarding usefulness.



### 3. Discussion of results and future work

These first impressions encourage the view that the tool will appeal to learners/readers, assisting in the

development of reading and associated skills including vocabulary, grammar, listening comprehension and pronunciation. The personalised concordance has not yet come into play for Icelandic learners due to the tiny corpus so far available. However, the crowdsourcing approach enables content providers, in this case teachers of Icelandic as L2, to add more texts and as this occurs, the personalised concordance will take its place in the learning process. Similar pilot studies were done in Australia and Iran, with Italian and Farsi respectively, and came to similar conclusions (Akhlaghi et al., 2019). The pilot test suggested several ways to improve the tool. Some important improvements have already been added, with others in the pipeline.

#### 4. Acknowledgements

This article is based upon work from COST Action enetCollect (CA16105) supported by COST (European Cooperation in Science and Technology). Work at the University of Geneva was funded by the Swiss National Science Foundation under grant IZCOZ0\_177065.

#### References

- Akhlaghi, E., Bédi, B., Butterweck, M., Chua, C.; Gerlach, J., Habibi, H., Ikeda, J., Rayner, M., Sestigiani, S. & Zuckermann, G. (2019). LARA: A Learning and Reading Assistant. in *Proc. SLATE* 2019, Graz, Austria.
- Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. <https://www.laurenceanthony.net/software>
- Boulton, A. & Tyne, H. (2013). Corpus linguistics and data-driven learning : a critical overview. In *Bulletin VALS-ASLA* 97, 97-118. <https://core.ac.uk/download/pdf/20661944.pdf>
- Baisa, V. & Suchomel, V. (2014). SkELL – Web Interface for English Language Learning. In *Eighth Workshop on Recent Advances in Slavonic Natural Language Processing*, 63-70. Brno: Tribun EU. Retrieved <https://nlp.fi.muni.cz/raslan/2014/12.pdf>
- Corino, E. & Onesti, C. (2019). Data-Driven Learning : A Scaffolding Methodology for CLIL and LSP Teaching and Learning. In *Teaching and Learning* 4(7). Frontiers of Education. <https://www.frontiersin.org/articles/10.3389/educ.2019.00007/full>
- Cobb, T. (2019) *Complete Lexical Tutor* v.8.3 [computer program]. <https://www.lextutor.ca>.
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. In *MIS Quarterly* 13(3), 310-340. MISRC University of Minnesota. <http://www.jstor.org/stable/249008>
- Johns, T. (2002). Data-driven Learning : The Perpetual Challenge. In *Teaching and Learning by Doing Corpus Analysis*, 105-117. Brill Rodopi. [https://doi.org/10.1163/9789004334236\\_010](https://doi.org/10.1163/9789004334236_010)
- Nesbitt, D. (2013). Student evaluation of CALL tools during the design process. In *Computer Assisted Language Learning* 26(4), 371-387. Routledge. <https://doi.org/10.1080/09588221.2012.680471>
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117-175. [http://dx.doi.org/10.1207/s1532690xci0102\\_1](http://dx.doi.org/10.1207/s1532690xci0102_1)
- Rychlý, P. (2007). Manatee/Bonito – A Modular Corpus manager. In *1<sup>st</sup> Workshop on Recent Advances in Slavonic Natural Language Processing*. Brno: Masaryk University, 2007, 65-70. <https://nlp.fi.muni.cz/trac/noske>
- Skriver, E. (1981). *Tína fer í frí. Námsgagnastofnun*. Reykjavík: Prentúsið.