

Reading Assistance through LARA, the Learning And Reading Assistant

Elham Akhlaghi¹, Ingibjörg Iða Auðunardóttir², Branislav Bédi³, Hakeem Beedar⁴,
Harald Berthelsen⁵, Cathy Chua⁶, Catia Cucchiari⁷, Brynjarr Eyjólfsson²,
Nedelina Ivanova⁸, Christèle Maizonniaux⁹, Neasa Ní Chiaráin⁵,
Manny Rayner¹⁰, John Sloan^{5,10}, Sigurður Vigfússon⁸, Ghil'ad Zuckermann⁴

¹Ferdowsi University of Mashhad, Iran; ²University of Iceland, Iceland;

³The Árni Magnússon Institute for Icelandic Studies, Iceland;

⁴University of Adelaide, Australia; ⁵Trinity College, Dublin, Ireland; ⁶Independent scholar;

⁷CLST, Radboud University Nijmegen, The Netherlands;

⁸The Communication Centre for the Deaf and Hard of Hearing, Iceland;

⁹Flinders University, Adelaide, Australia; ¹⁰FTI/TIM, University of Geneva, Switzerland;

elhamakhlaghi80@gmail.com, iia2@hi.is, branislav.bedi@arnastofnun.is,

hbeedar@hotmail.com.au, berthelh@tcd.ie, cathyc@pioneerbooks.com.au,

c.cucchiari@let.ru.nl, jcy1@hi.is, nedelina@shh.is, christele.maizonniaux@flinders.edu.au,

Neasa.NiChiarain@tcd.ie, Emmanuel.Rayner@unige.ch, sloanjo@tcd.ie,

siggivig@gmail.com, ghilad.zuckermann@adelaide.edu.au,

Abstract

We present an overview of LARA, the Learning And Reading Assistant, an open source platform for easy creation and use of multimedia annotated texts designed to support the improvement of reading skills. The paper is divided into three parts. In the first, we give a brief summary of LARA's processing. In the second, we describe some generic functionality specially relevant for reading assistance: support for phonetically annotated texts, support for image-based texts, and integrated production of text-to-speech (TTS) generated audio. In the third, we outline some of the larger projects so far carried out with LARA, involving development of content for learning second and foreign (L2) languages such as Icelandic, Farsi, Irish, Old Norse and the Australian Aboriginal language Barnjarla, where the issues involved overlap with those that arise when trying to help students improve first-language (L1) reading skills. All software and almost all content is freely available.

Keywords: CALL, multimodality, reading, open source, evaluation

1. Introduction and overview

LARA (<https://www.unige.ch/collector/lara/>) is an open source learning-by-reading platform under development by an international consortium since 2018. Starting at the University of Geneva, the user base has grown quickly and now includes groups in over a dozen countries. LARA supports easy construction of annotated multimodal texts using open source tools which can either be invoked from the command-line or, more commonly, through an online portal. These texts typically include various features for reading assistance such as integrated audio, translations, and an automatically generated concordance. A screenshot showing a page from a LARA text is shown in Figure 1.

The basic idea of adding multimedia annotations to texts in order to help non-L1 language learners is natural, and has been implemented in some form in many other platforms; prominent examples include LingQ¹, Learning With Texts², the Perseus Digital Library's Scaife viewer³ and Clilstore⁴. What primarily distinguishes LARA from these is the project's open source nature, where new features are added in a bottom-up process driven by the demands of a diverse community involved in many different kinds of language-related projects.

As noted, LARA has originally been developed to help people learn and read non-L1 languages. However, the boundary between non-L1 and L1 turns out to be less clear than we had initially expected. For example, looking ahead to §4.1., a major hurdle for beginner learners of Irish is the opaque writing system, which makes it unusually difficult to acquire a good understanding of the letter/sound rules. The problems these people face are not dissimilar to those experienced by people with L1 reading difficulties. In addition, recent extensions to the LARA functionality are moving in a direction that could make it more directly useful as a tool for providing assistance to people with reading difficulties.

This paper is intended to provide a self-contained overview of the reading assistance facilities in LARA that may be relevant to the reading difficulties community. We start by presenting a brief summary of the core LARA functionality (§2.), then describe recently added functionality particularly relevant in the present context (§3.). In §4. we describe some LARA projects where the issues would appear to overlap with those arising when helping people with reading difficulties. The final section summarises and looks ahead.

2. Core LARA functionality

The core of LARA is a set of tools that make it easy to convert text into the multimodal annotated form illustrated in Figure 1. The conversion process consists of the following

¹<https://www.lingq.com/>

²<https://sourceforge.net/projects/lwt/>

³<https://scaife.perseus.org/>

⁴<http://multidict.net/clilstore/>

Fairceallach Fhinn Mhic Chumhail ← 1

2

4

Lá dá raibh bhí Fionn agus na Fianna ag fiach. Chaitheadar an lá ó mhoch na maidne go dul faoi na gréine agus ar a gceasadh dóibh abhaile d'iompaigh siad siar féachaint an t-*Once upon a time, Fionn and na Fianna were out hunting, ... ina stumpa beag ramhar, chucu. Bhí síúl maith acu á dheanamh agus dá theabhas é an stiú a bhí acu ní rabhadar ag imeacht aon ní ón bhFairceallach. Bhí sé ag coimeád leo i gcónaí; faoi mar a bhíodís ar an ardán, bhíodh an Fairceallach san ísleán, agus nuair a bhíodís san ísleán bhíodh an Fairceallach san ardán. Choinnigh sé ina ndiaidh mar sin chun go dtángadar go dtí a mbunúit féin. Shuofodar síos chun bia agus cuireadh féasta maith bia rompu agus Fionn ag freastal orthu ag gearradh bia agus feola dóibh fad a bhíodar ag ithe. Tháinig an Fairceallach isteach agus shuigh sé síos sa chéinne. Chonaic Fionn é agus thug sé chun boird é agus chaitheadar a sásamh.*



bí ← 3

(Verb)

- ← Lá dá raibh bhí Fionn agus na Fianna ag fiach.
- ← Chaitheadar an lá ó mhoch na maidne go dul faoi na gréine agus ar a gceasadh dóibh abhaile d'iompaigh siad siar féachaint an raibh Oisín ina ndiaidh.
- ← Chonaic siad Fairceallach, fear a bhí ina stumpa beag ramhar, chucu.
- ← Bhí síúl maith acu á dheanamh agus dá fheabhas é an síúl a bhí acu ní rabhadar ag imeacht aon ní ón bhFairceallach.
- ← Bhí sé ag coimeád leo i gcónaí; faoi mar a bhíodís ar an ardán, bhíodh an Fairceallach san ísleán, agus nuair a bhíodís san ísleán bhíodh an Fairceallach san ardán.
- ← Shuofodar síos chun bia agus cuireadh féasta maith bia rompu agus Fionn ag freastal orthu ag gearradh bia agus feola dóibh fad a bhíodar ag ithe.
- ← An lá dár gcionn d'ímíodar rompu ar an slí chéanna agus an Fairceallach ina ndiaidh arís mar a bhí sé an lá roimhe sin.
- ← Nuair a chasadar abhaile tráthnóna bhí sé sa tsíúl céanna i gcónaí agus é á leanúint.
- ← D'fhan sé mar sin ina dtéannta nó go raibh lá agus bliain tugtha aige.
- ← Dúirt sé ansin: "A Fhinn Mhic Chumhail", ar seisean, "dá cuireadh dinnéir agam á thabhairt duit féin agus d'Fhianna Éireann".
- ← "Ó", ar seisean, "ní aon fhear i seacht gcatha na nGnáth-Fhéinne nach gcuirfidh mé scian agus forc agus pláta ar a aghaidh amach!"
- ← Bhí an Fairceallach rompu amach ag taispeáint an bhóthair dóibh.
- ← Leanadar é go díreach agus bhí sé chomh fada rompu an lá sin is a bhí sé ina ndiaidh an chuid eile den bhliain.
- ← Agus iad ag teacht go dtí an doras, bhí bean óg agus a droim le hursain an dorais aici.
- ← Bhí a dá lámh faoina hascaill aici agus ní raibh aon éadach óna dá uilinn amach uirthi.
- ← Bhí cúl breá gruaige uirthi ar dhath an óir síos go caol a droma.
- ← Bhí dhá shúil ghlasa ghorma aici agus muinte uirthi chomh geal agus a chonaiceadar ar aon bhean riamh.

Figure 1: Example of Irish LARA content, *Fairceallach Fhinn Mhic Cumhail*, ('Fionn's burly friend'), reproduced from (Zuckerman et al., 2021). A 'play all' audio button function is included at the top of the page to enable the listener to hear the entire story in one go (1). The text and images are in the pane on the left hand side. Clicking on a word displays information about it in the right hand pane. Here, the user has clicked on *bhí* = "to be (past tense)" (2), showing an automatically generated concordance; the lemma *bí*; and every variation of *bí* that is in this text (3). Hovering the mouse over a word plays audio and shows a popup translation at word-level. Clicking on a loudspeaker plays audio for the entire sentence as well as showing a popup translation (4). 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 examples page.⁶

steps:

Segment: Segment the text. For European languages, this means splitting up lists of words into sentence-length segments using a sentence tokeniser. The result is then in general manually post-edited.

Tag: Tag the text, to mark each word with its lemma form and (optionally) part of speech. This is needed in order to build the lemma-oriented concordance. When a tagger/lemmatiser is available, this is first used to perform the tagging automatically, after which the result is again manually post-edited. Several tagger/lemmatisers are now integrated into LARA, covering over 20 languages.

Identify resources needed: Process the text to create a set of resource files which specify other annotation data that needs to be added. The most important are i) associations of words and segments with audio files, ii) associations of words and segments with translations, and iii) potential occurrences of multiword expressions (MWEs) taken from an MWE lexicon for the language in question.

Instantiate resources: Upload the resource files to tools which support easy entry of the missing information. Audio files are created through a user-friendly online recording tool. Translations are entered through a

spreadsheet-like interface. Candidate MWEs are confirmed or rejected through another interactive tool.

Create pages: Combine all the information to create the multimedia pages.

These operation can either be carried out directly using command-line tools, or can be invoked through the LARA portal, a free online service that provides a user-friendly wizard-style interface. Full details and examples can be found in the online documentation⁷.

3. Functionality

We describe three pieces of LARA functionality introduced over the last year that are potentially relevant to helping people with reading difficulties: phonetic texts, annotated images and picture lexica, and integrated TTS.

3.1. Reading assistance through phonetic texts

LARA documents were originally conceived as texts with a hierarchical structure consisting of pages, segments and words, where the words are associated with lemmas. In order to address the needs of students who are uncertain

⁷<https://www.issco.unige.ch/en/research/projects/collector/LARADoc/build/html/index.html>

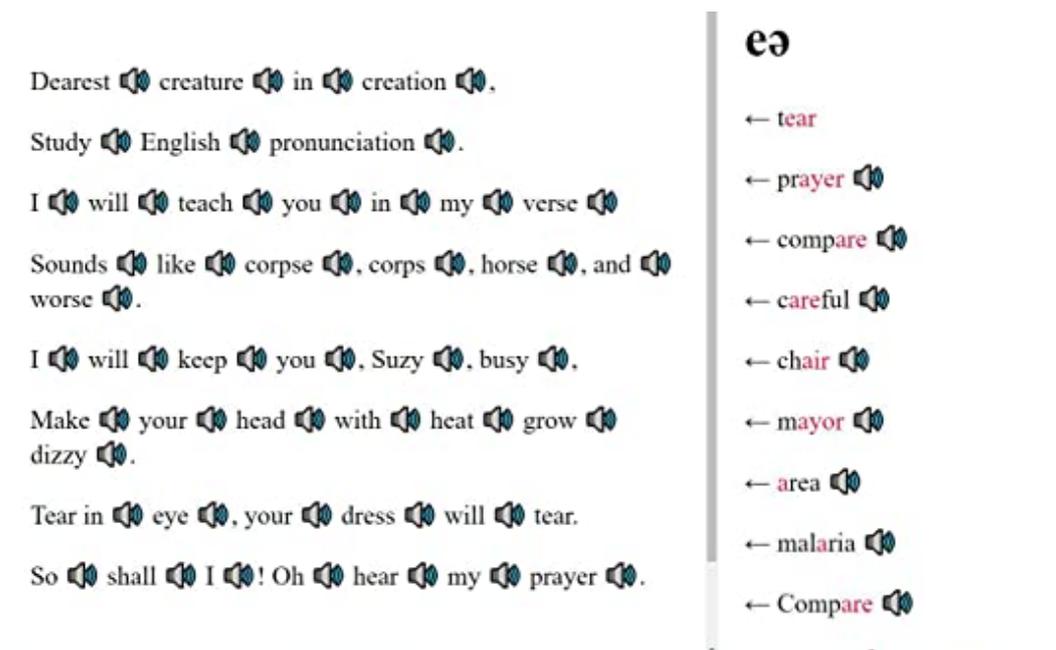


Figure 2: “Phonetic” LARA version of Charivarius’s 1920 pronunciation poem *The Chaos* (screenshot of first page). Clicking on a letter group, here “ear” in “tear” at the end of the penultimate line, highlights the group, plays the sound, and shows a phonetic concordance on the right; clicking on a speaker icon plays audio for whole of the preceding word. A link to the text is posted on the LARA examples page.

about letter/sound correspondences, we have recently extended the framework to allow the option of creating texts annotated at the phonetic level. A “phonetic” LARA text is hierarchically divided into pages, words and letter-groups, where each letter-group is associated with a phonetic value. The same notation is used for both types of text, and nearly all of the processing associated with normal (word-oriented) LARA texts carries over to phonetic texts; in particular, a compiled phonetic text contains a phonetic concordance, giving examples of contexts where each phonetic value occurs. A playful example illustrating the “phonetic text” functionality is shown in Figure 2.

It would be extremely laborious to construct phonetic LARA texts by hand, and there is a script that converts a normal text into the corresponding phonetic version. This post-processes the internalised text to convert each word into a corresponding phonetic version, while keeping formatting unchanged. For languages which are written completely phonetically, this only requires the annotator to supply the list of phonetically meaningful letter groups defining the orthography of the language. An example for Barnarla (cf. §4.2.) is *Mangiri Yarda*; this also uses the annotated image functionality described in §3.2..

For languages where online phonetic lexica exist, phonetic versions of most words can be read off the lexicon; free phonetic lexica for many languages are for example available from the IPA-dict project.⁸ The challenge is to align the letters with the phonetic symbols. At the moment, the conversion script helps the annotator compile an aligned phonetic lexicon, where typical entries are as illustrated in

⁸<https://github.com/open-dict-data/ipa-dict>

Figure 4. The script creates new entries automatically using a simple dynamic programming method which maximises the number of alignments already seen in the lexicon (this idea is partly inspired by the one from (Jiampojarn and Kondrak, 2010)), after which a human annotator cleans up the result. Further details are given in the online documentation.⁹

Once a reasonable number of examples of aligned words have been collected, error rates become low and the cleaning-up process is quick. We present the results of a preliminary evaluation for English and French to support this claim. In English, we began by constructing an initial aligned word lexicon for a few small texts, the largest of which was “The Chaos” (cf. Figure 2). This produced a total of 990 aligned words, which included 264 unique primitive grapheme-sequence/phoneme-sequence correspondences; phonetic transcriptions were taken from the UK English IPA-dict resource. We then ran the alignment guessing script on the text of an English translation of Saint-Exupéry’s *Le Petit Prince*. This contained 1833 unique words, of which 309 were already in the aligned-word lexicon. Of the remaining 1524, 78 were not in the IPA-dict lexicon, most of them either because they were heteronyms (“close”, “live”, “wind”) or proper names (“africa”, “antoine”, “siberia”).

Editing the aligned lexicon took an expert annotator, one of the authors, about three hours. Comparing the edited and raw versions, we found that the script had correctly aligned 1410/1524 of the new words (92.5%) and 9041/9580 of the

⁹https://www.issco.unige.ch/en/research/projects/collector/LARADoc/build/html/phonetic_texts.html

(a)

```
<annotated_image>  
  
chair man glass ||  
table ||  
glass woman chair ||  
</annotated_image>
```

(b)

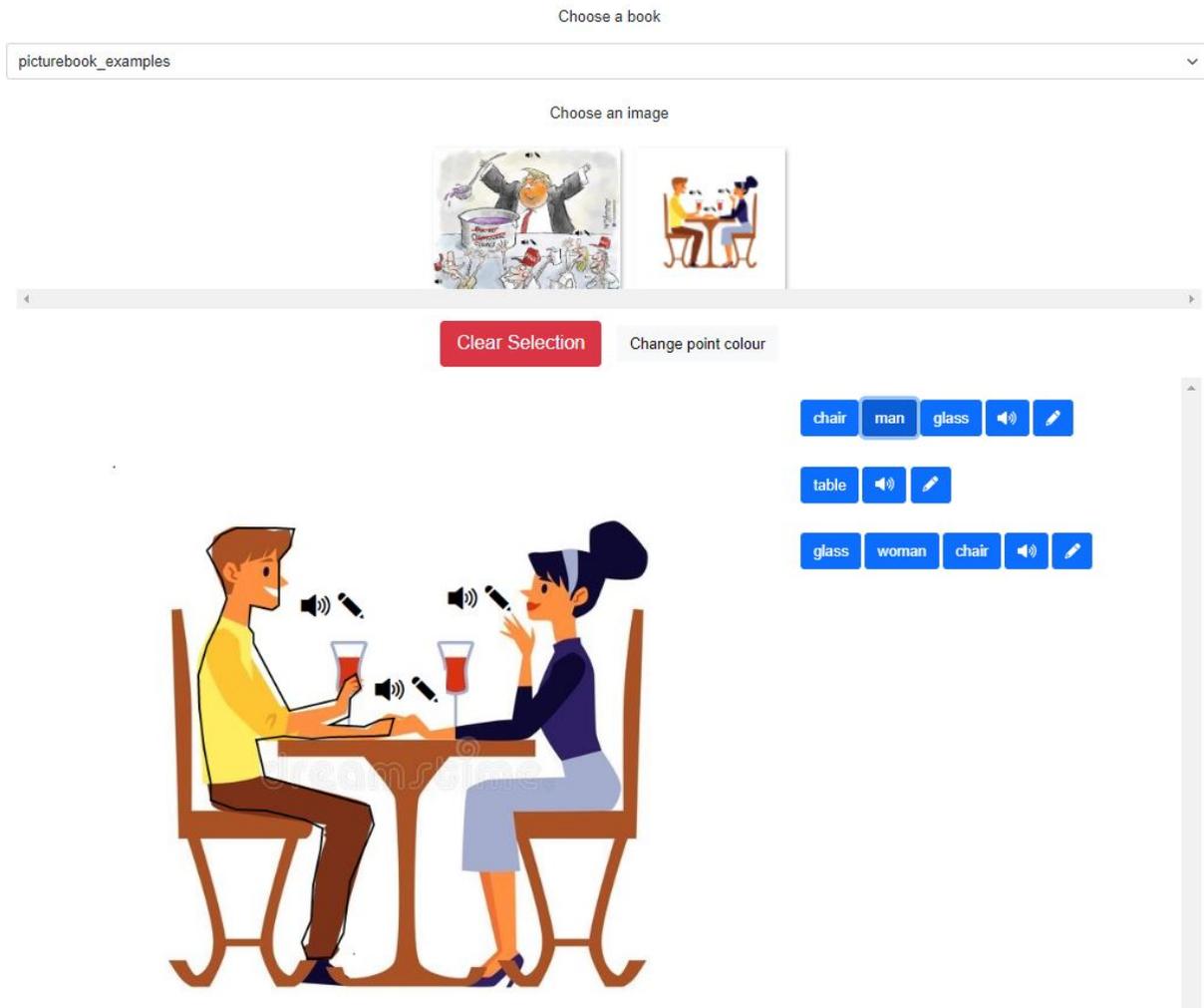


Figure 3: Toy example of a piece of image-based text based on a simple cartoon, taken from (Bédi et al., 2022). The LARA source (a) is given above. The screenshot below (b) shows the tool used to create the word locations file. The top control allows the annotator to choose the text to annotate, after which the slider with the series of thumbnails allows them to choose a page by its image. The bottom left pane presents the selected image, and the bottom right pane the associated words. The annotator can draw a polygon on the left and save it to a word, or select a word on the right to show the current polygon. Here, the annotator has just selected the word “man” on the right, showing the polygon for the picture of the man on the left. The speaker and pencil icons optionally associate audio or text with a whole line. A link to the LARA document is posted on the LARA examples page under “Two cartoons”.

graphemes (94.4%). Looking only at the subset for which IPA-dict entries were available, the figures were 1410/1446 (97.5%) for words and 9041/9102 (99.3%) for characters. Audio for the phonetic content was recorded by one of the authors, a native speaker of English.

The French experiment was similar, though the initial version of the aligned lexicon was based on a smaller sample of

language. This time, we used the original French edition of *Le Petit Prince* as the evaluation text. This contained 2583 unique words, of which 559 were already in the aligned-word lexicon. Of the remaining 2024, 32 were not in the IPA-dict lexicon. Again, editing the guessed aligned lexicon took about three hours. The script correctly aligned 1876/2024 of the new words (92.7%) and 13722/14191 of

the graphemes (96.7%). For the subset where IPA-dict entries were available, the figures were 1876/1992 (94.2%) and 13722/13966 (98.3%) for graphemes. This time, we experimented with a different strategy for creating the phonetic audio, and generated it using one of the French voices on the IPA-reader site.¹⁰

Links to the English and French versions are posted on the LARA examples page under the titles “The Little Prince” and “Le petit prince”.

"admirateur"
"a d m i r a t e u r"
"a d m i ʁ a t œ ʁ"
"ainsi"
"ain s i"
"ɛ̃ s i "
"alors"
"a l o r s"
"a l ɔ ʁ "

Figure 4: Examples of entries from French aligned pronunciation lexicon. Several letters can map into one (beginning of "ainsi"), and letters can map into the empty string (end of "alors").

3.2. Annotated images and picture lexica

Another extension to the original text-based LARA document structure concerns images. LARA has always supported inclusion of images using the HTML `` tag, but these were represented internally as atomic constituents without internal structure. Layout was added using other HTML tags. Although this model works well for many documents, it ignores the fact that a written text is not just a collection of strings but also a visual object. For some kinds of documents, for example picture-books and posters, the visual content can be as important as the words.

In order to address these issues, which are particularly relevant to helping students with reading difficulties, we have recently added new functionality to allow “annotated images” as components of a text. A component of this kind is delimited using the `<annotated_image>` tag. It must contain exactly one `` tag and some text, where the text is interpreted as being associated with locations in the image. During the processing phase which identifies resources needed for a text, the images and associated words are extracted, after which they are uploaded to an online graphical tool where the annotator can draw the outline for the location of each word in an image. Figure 3 illustrates. The graphical correlate of a word can, but does not need to be a graphical representation of the word; it can equally well be a part of the image associated with the word. So for example the word “apple” could be associated in the image with the handwritten text *A P P L E*, but it could also be associated with an area of the image showing an apple. In the final LARA text, the locations in the image marked as associated with words react to clicking or hovering actions. Another piece of image-related functionality is the provision of support adding a “picture lexicon” to a text, which

associates some lemmas with graphical images. This has already been used for the Barngarla project (cf. §4.2.). The initial text “Welcome to Country with picture lexicon” linked from the examples page was warmly approved by the Barngarla elders guiding the language revival process.

3.3. Reading assistance through integrated TTS

Integration of TTS was primarily implemented to support the Irish group (cf. §4.1.), who have from the start used it exclusively to create Irish language audio. Initially, other groups were sceptical about creating LARA content that used TTS audio, believing that the quality would be insufficient compared to human-recorded audio.

Two collaborative evaluation exercises have however demonstrated that, for many languages, TTS works much better than was generally expected. The first of these exercises was carried out during Q1 2021 and involved the Australian, Icelandic, Iranian, Irish, Dutch, Polish, Slovak and Swiss groups. About twenty LARA documents, in various languages, were produced in both TTS and human audio form and compared by 130 evaluators using an anonymous web form. One expects TTS audio to be much quicker to produce, but of lower quality: the goal was to obtain quantitative and qualitative data exploring the issues. The results were presented at EUROCALL 2021 (Akhlaghi et al., 2021). To our surprise, TTS audio was in fact rated equal to or better than human audio in three of the ten languages. A follow-on study was carried out in Q1 2022 and will be presented at the LREC 2022 conference. Since it used data taken from a uniform text, different translations of Saint-Exupéry’s *Le petit prince*, comparisons between languages were more obviously meaningful, and the number of evaluators was approximately doubled. The results were similar to those of the first study. Although the quality of TTS varied widely between languages, the best TTS voices were of a quality comparable with non-professional human voices and again were in some cases preferred.

4. Example projects

LARA was originally designed for creating annotated texts that would improve learners’ reading and listening skills in L2 languages. After three years of experience in using the tool, it turns out that the dividing line between L2 and L1 is less clear than we had realised, and that the issues appear to overlap to a considerable extent. We briefly describe some substantial projects exemplifying this observation. In §4.1. we consider the paradoxical case of Irish, where a country’s official first language is simultaneously an endangered language. §4.2. describes use of LARA with Barngarla, an Australian Aboriginal language which for several decades was considered dead, but which is now being revived by ethnic Barngarla people. In §4.3. we look at texts designed to help Deaf Icelandic children improve their reading skills in Icelandic, and in §4.4. at Old Norse, the archaic form of Icelandic taught as an obligatory subject in Icelandic schools. §4.5. reviews a project carried out in Iran, where a series of Farsi readers have been converted into LARA form.

¹⁰<http://ipa-reader.xyz/>; “Celine” voice.

4.1. Online resources for reading assistance in Irish

Irish is in the possibly unique position of being both the official first language of a sovereign state and also an endangered language. It is a community language only in relatively small regions (Gaeltacht regions) in the West of Ireland, with daily speaker numbers of about 20,000, or less than 0.5% of the Irish population (CSO, 2016). At the same time, it is an obligatory subject in schools, with 700,000 learners in the education system in the Republic of Ireland (Ní Chiaráin and Ní Chasaide, 2020).

Teaching and learning Irish presents multiple challenges, and learning to read Irish is one of them. The first language of most learners will be English, a Germanic language whose structure diverges substantially from that of the Celtic language Irish; the basic word-order of Irish is different (VSO as opposed to English SVO), and it is highly inflected, with up to 42 inflected forms for verbs. A striking feature of the sound system is the contrast of palatalised and velarised consonants, with a very large inventory, relative to English. This feature is partially obscured, and complicated for learners, by the notoriously opaque writing system, and the link of the sounds to written forms is often poorly understood. The initial sounds of lexical items ‘mutate’ in certain grammatical contexts, so that e.g., in a word like *bord* ‘table’ it may be [b], [w], [v] or [m]. Although there is an agreed standardised written form, there is no single spoken standard, but rather three major dialects. Teachers are typically second language learners themselves, and their own confidence in the language can be problematic. They often feel overburdened with the major responsibility placed on them in the revitalisation and maintenance initiative, but report inadequate resources and training to fulfil it (Dunne, 2019).

In this context, it turns out that LARA has much to offer in terms of reading assistance. Using the synthetic voices developed for the main Irish dialects by the AB AIR project (AB AIR, 2020) and integrated into LARA, it is easy to link text to audio in any of the three dialects, bringing a native speaker model directly into the classroom; the lemma-based concordance similarly allows the learner to access the dictionary form of any word with a single click. Starting with pilot LARA adaptations of traditional Irish folk-stories, the team at Trinity College Dublin have created a substantial set of Irish reading material in LARA form, posted on the *An Scéalai* (“The Storyteller”) CALL platform.¹¹ User feedback has been extremely positive. In a recent survey, for example, 92% of 494 adult respondents reported that using *An Scéalai* had a positive impact on their language learning journey. 90% of same stated they would like to continue using the platform into the future. Many users commented that LARA made complex texts accessible - learners felt they engaged more deeply and spent more time on ‘difficult’ reading materials than they would otherwise have done if presented to them in a more traditional format.

4.2. Reading assistance in Barngarla, a revived Australian Aboriginal language

Barngarla is an Australian Aboriginal language belonging to the Thura-Yura language group, a subgroup of the large Pama-Nyungan language family. Typically for a Pama-Nyungan language, Barngarla has a phonemic inventory featuring three vowels ([a], [i], [u]) and retroflex consonants, an ergative grammar with many cases, and a complex pronominal system. Unusual features include a number system with singular, dual, plural and superplural and matrilineal and patrilineal distinction in the dual.

During the twentieth century, Barngarla was intentionally eradicated under Australian ‘stolen generation’ policies, the last original native speaker dying in 1960. Language reclamation efforts were launched in 2011 (Zuckermann, 2020). Since then, a series of language reclamation workshops have been held in which about 120 Barngarla people have participated. The primary resource used has been a dictionary, including a brief grammar, written by the German Lutheran missionary Clamor Wilhelm Schürmann (Schürmann, 1844; Clendon, 2015).

Other published resources for Barngarla, non-existent ten years ago, are now emerging. The most visible example to date is *Barngarlidhi Manoo* (Zuckermann and the Barngarla, 2019), a Barngarla alphabet book/primer compiled by Ghil’ad Zuckermann in collaboration with the nascent Barngarla revivalistic community. A first step in evaluating the possible relevance of LARA to Barngarla was to convert this book into LARA form (Butterweck et al., 2019). The LARA reading assistance functionality is primarily used to attach audio recordings to Barngarla language: words and phrases marked in red can be played by hovering the mouse over them.

A second resource was produced as part of the “Fifty Words Project”¹², which collects together fifty basic words such as “fire”, “water”, “sun” and “moon” for several dozen Aboriginal languages. The Barngarla version, recorded by ethnic Barngarla language custodian Jenna Richards from Galinyala (= Port Lincoln), is available on the Fifty Words page. A third Barngarla text, *Mangiri Yarda* (“Healthy Country”) (Zuckermann and Richards, 2021) has been designed as a teaching resource. In contrast to *Barngarlidhi Manoo*, which is almost exclusively focused on vocabulary, *Mangiri Yarda* introduces some grammar.

Links to all of these texts are posted on the LARA examples page.

4.3. Helping Deaf Icelanders improve their reading skills

Although Icelandic is the primary language of Iceland, Deaf children usually grow up learning a signed language as their first language. In practice, written Icelandic is not perceived as an L1 for these children, so tools that can help them make progress in reading are potentially very useful. It turned out to be quite easy to extend LARA so that it can support this kind of scenario: basically, all that was necessary was to arrange things so that recorded signed video can systematically be used as an alternative to recorded au-

¹¹<https://abair.ie/scealai/#/landing>

¹²<https://50words.online/>

dio. Thus a LARA text of this type is written in Icelandic, but words and sentences are associated with Icelandic Sign Language (ÍTM) signed videos. The signed video for a word is accessed by clicking on the word; the signed video for a sentence is accessed by clicking on a camera icon inserted at the end of the sentence. (In ‘video mode’, the camera icon replaces the usual loudspeaker icon).

Videos are recorded using the same third-party recording tool as is used for recording audio content; the tool had already been adapted for this purpose in a previous project (Ahmed et al., 2016). The workflow for recording is modality-independent. The LARA portal creates the recording script from the text and uploads it to the recording tool; the voice talent/signer records the audio/video from the script; at the end, the portal downloads the recorded multimedia and inserts it into the LARA document.

A link to an initial example of a LARA document of this kind, a children’s story about 2.7K words long, is posted on the LARA examples page. The student who created the signed content turned to two members of staff at the Center for feedback. One is a native ÍTM signer and the other has worked as an sign language interpreter for over two decades. There were many things to consider, as ÍTM is not a standardised language, even to the extent that the basic word order is unclear: research (Brynjólfssdóttir, 2012) shows that subjects accept both SOV and SVO word orders. The central issue was the question of whether the signed translation of the text should be true to the Icelandic original or re-expressed in ÍTM. One argument is that, as a tool to learn written Icelandic, the translation should be faithful to the source so that ÍTM signs corresponding to the Icelandic words appearing there. The argument in the opposite direction is that a free re-interpretation is better suited to helping Deaf children understand the signed content. In the end an interpreting strategy was preferred for three reasons. Comprehension of the signed text is crucial for Deaf children; the interpreting strategy seemed to be a better fit to the content of a children’s book; and in LARA learners can click on a word in the Icelandic text to see the ÍTM sign, if the corresponding sign did not appear in the freely translated segments.

4.4. Assistance in reading Old Norse epic poems

Old Norse, the language spoken in what is now Scandinavia from the 7th to the 15th century, is an important part of Icelandic culture. The linguistic evolution of Icelandic has proceeded more slowly than that of the mainland languages (Danish, Norwegian and Swedish), and it is close enough to Old Norse that Old Norse literary works are still more or less comprehensible; a reasonable comparison point for anglophones might be Chaucerian English. Old Norse language and literature is an obligatory subject in Icelandic secondary schools. It is however clear that many students find it challenging. They are particularly challenged by the Poetic Edda, a classic poem-cycle first written down in the late 13th century, which occupies a central place in the Old Norse canon. The dense, allusive language is much harder to understand than that in prose works, and the less motivated students often experience it as close to incomprehensible.

Particularly as a tagged version of the Poetic Edda already existed, the group at the Árni Magnússon Institute for Icelandic Studies felt that this combination of circumstances made it a good target for conversion into LARA form. Three of the best-known poems from the cycle — the *Völuspá*, *Hávamál* and *Lokasenna* — have now been completed, and several more are in preparation. The *Völuspá* project is presented in (Bédi et al., 2020); as described there, initial feedback from Icelandic users has been very positive. All three of the Eddaic poems so far released have also been used as the basis of reading groups on the popular Goodreads review site.¹³ They attracted a small but enthusiastic audience, with a total of 185 posts for the three groups.

4.5. Online resources for reading assistance in Farsi

The Ferdowsi University of Mashhad (FUM; third highest ranked university in Iran) has used LARA since shortly after the inception of the LARA project. FUM began by developing short LARA texts in Farsi, for use in a Farsi course for Arabic-speaking students at FUM. Early results are reported in (Akhlaghi et al., 2019). This pilot exercise was successful enough that Iranian funding was granted to convert an five-volume series of Farsi textbooks, developed at FUM by Professor Ehsan Ghabool, into LARA form. The project was completed during Q1 2021, and the result is now being used at FUM’s International Center for Teaching Persian to Non-Persian People¹⁴.

5. Summary and further directions

We have presented a brief overview of the LARA community and platform, focusing on issues that overlap with those relevant to supporting people with reading difficulties and illustrating with some practical use cases. Work in several of these areas is under active development. We are particularly interested in exploring the possibilities opened up by the new “phonetic text” and “annotated image” functionalities, and welcome suggestions from the reading difficulties community about ways to repurpose the LARA technology to this new domain.

6. Bibliographical References

- ABAIR, (2020). *ABAIR: An Sintéiseoir Gaeilge - The Irish Language Synthesiser AB AIR*. <http://www.abair.ie>. As of 8 September 2020.
- Ahmed, F., Bouillon, P., Destefano, C., Gerlach, J., Hooper, A., Rayner, M., Strasly, I., Tsourakis, N., and Weiss, C. (2016). Rapid construction of a web-enabled medical speech to sign language translator using recorded video. In *Proceedings of FETLT 2016*, Seville, Spain.

¹³<https://www.goodreads.com/topic/show/21221144-v-lusp-reading-group-verses-1-6>; <https://www.goodreads.com/topic/show/22046862-h-vam-1-reading-group>; <https://www.goodreads.com/topic/show/22089678-lokasenna-reading-group>

¹⁴2100 students; <https://ctpl.um.ac.ir/index.php?lang=en>

- 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 *Proc. SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education*, pages 99–103.
- Akhlaghi, E., Bączkowska, A., Berthelsen, H., Bédi, B., Chua, C., Cuchiarini, C., Habibi, H., Horváthová, I., Hvalsøe, P., Lotz, R., Maizonniaux, C., Ní Chiaráin, N., Rayner, M., Tsourakis, N., and Yao, C. (2021). Assessing the quality of TTS audio in the LARA learning-by-reading platform. In N. Zoghiani, et al., editors, *CALL and professionalisation: short papers from EUROCALL 2021*, pages 1–5.
- Bédi, B., Bernharðsson, H., Chua, C., Guðmarsdóttir, B. B., Habibi, H., and Rayner, M. (2020). Constructing an interactive old Norse text with lara. *CALL for widening participation: short papers from EUROCALL*, pages 27–35.
- Brynjólfssdóttir, E. G. (2012). *Hvað gerðir þú við peningana sem frúin í Hamborg gaf þér? Myndun hv-spurninga í íslenska táknmálinu*. Ph.D. thesis.
- Butterweck, M., Chua, C., Habibi, H., Rayner, M., and Zuckermann, G. (2019). Easy construction of multimedia online language textbooks and linguistics papers with LARA. In *Proc. 12th annual International Conference of Education, Research and Innovation*, Seville, Spain.
- Bédi, B., Beedar, H., Chiera, B., Ivanova, N., Maizonniaux, C., Chiaráin, N. N., Rayner, M., Sloan, J., and Zuckermann, G. (2022). Using lara to create image-based and phonetically annotated multimodal texts for endangered languages. In *Proceedings of the Workshop on Computational Methods for Endangered Languages*.
- Clendon, M. (2015). *Clamor Schürmann's Barngarla grammar: A commentary on the first section of A vocabulary of the Parnkalla language*. University of Adelaide Press.
- CSO, (2016). *Census 2016 Profile 10 – Education, Skills and the Irish Language*. <https://www.cso.ie/en/csolatestnews/presspages/2017>. As of 11 October 2020.
- Dunne, C. M. (2019). Primary teachers' experiences in preparing to teach Irish: Views on promoting the language and language proficiency. *Studies in Self-Access Learning*, 10(1):21–43.
- Jiampojarn, S. and Kondrak, G. (2010). Letter-phoneme alignment: An exploration. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden*, pages 780–788.
- Ní Chiaráin, N. and Ní Chasaide, A. (2020). The potential of text-to-speech synthesis in computer-assisted language learning: A minority language perspective. In Alberto Andujar, editor, *Recent Tools for Computer- and Mobile-Assisted Foreign Language Learning*, chapter 7, pages 149–169. IGI Global, Hershey, PA.
- Schürmann, C. W. (1844). *A Vocabulary of the Parnkalla Language. Spoken by the natives inhabiting the western shore of Spencer's Gulf. To which is prefixed a collection of grammatical rules, hitherto ascertained*.
- Zuckerman, G., Vigfússon, S., Rayner, M., Chiaráin, N. N., Ivanova, N., Habibi, H., and Bédi, B. (2021). Lara in the service of revivalistics and documentary linguistics: Community engagement and endangered languages. In *Proceedings of the Workshop on Computational Methods for Endangered Languages*, volume 1, pages 13–23.
- Zuckermann, G. and Richards, E. (2021). *Mangiri Yarda (Healthy Country: Barngarla Wellbeing and Nature)*. Revivalistics Press.
- Zuckermann, G. and the Barngarla. (2019). *Barngarlidhi Manoo: "Speaking Barngarla Together"*. Barngarla Language Advisory Committee (BLAC).
- Zuckermann, G. (2020). *Revivalistics: From the Genesis of Israeli to Language Reclamation in Australia and Beyond*. New York: Oxford University Press.