LARA in the Service of Revivalistics and Documentary Linguistics: Community Engagement and Endangered Languages[*]

Ghil'ad Zuckermann
The University of Adelaide
Australia
`Ghilad.Zuckermann@adelaide.edu`

Sigurður Vigfússon
The Communication Centre for the Deaf
and Hard of Hearing, Iceland
`siggivig@gmail.com`

Manny Rayner
FTI/TIM, University of Geneva
Switzerland
`Emmanuel.Rayner@unige.ch`

Neasa Ní Chiaráin
Trinity College, Dublin
Ireland
`Neasa.NiChiarain@tcd.ie`

Nedelina Ivanova
The Communication Centre for the Deaf
and Hard of Hearing, Iceland
`nedelina@shh.is`

Hanieh Habibi
FTI/TIM, University of Geneva
Switzerland
`Hanieh.Habibi@unige.ch`

Branislav Bédi
The Árni Magnússon Institute
for Icelandic Studies, Iceland
`branislav.bedi@arnastofnun.is`

Abstract

We argue that LARA, a new web platform that supports easy conversion of text into an online multimedia form designed to support non-native readers, is a good match to the task of creating high-quality resources useful for languages in the revivalistics spectrum. We illustrate with initial case studies in three widely different endangered/revival languages: Irish (Gaelic); Icelandic Sign Language (ÍTM); and Barngarla, a reclaimed Australian Aboriginal language. The exposition is presented from a language community perspective. Links are given to examples of LARA resources constructed for each language.

1 Introduction

When talking about languages with small, shrinking or non-existent speaker bases, we can adopt a positive or a negative attitude. If we say "endangered" or "dead" languages, the terms predisposes us towards a negative, deficit point of view: perhaps the most important thing is to try to document the language as well as we can while information about it is still available. In this paper, we will in contrast accentuate the positive. All languages, except those where 100% of the children in the relevant group grow up speaking the language, are

under threat. Given a particular language we care about, what can we do to improve its prospects?

To underline our commitment to this positive perspective, we will situate the discussion within the emerging trandisciplinary field of Revivalistics (Zuckermann, 2020), a neologism which combines the notions of "revival" and "linguistics". This in no way should be read as playing down the importance of language documentation, which for obvious reasons is an essential component of the enterprise. Rather, we want to view documentary linguistics through a revivalistic lens, documenting the language so that the material we produce can directly help community members who are trying to strengthen their linguistic abilities but may not be linguistically sophisticated. In this paper, we argue that LARA (Learning And Reading Assistant; (Akhlaghi et al., 2019)), an open source web platform that supports easy conversion of text into multimodal form, offers functionality that fits surprisingly well with the goals of revivalistics, and makes available a plethora of possibilities for rapid creation of useful online teaching materials.

In the rest of the paper, we start in §2 by giving a brief introduction to revivalistics, as here conceptualised, and to LARA. In §§3–5, we present case studies showing how we have used LARA to construct resources for three widely different

---

[*] Authors in reverse alphabetical order.

endangered/Sleeping Beauty ("dead") languages: Irish (Gaelic); Icelandic Sign Language (ÍTM); and Barngarla, a reclaimed Australian Aboriginal language. The final section concludes and outlines ongoing work.

2 Background: Revivalistics and LARA

2.1 Revivalistics

Revivalistics is defined by Zuckermann (2020) as "a new global, trans-disciplinary field of enquiry surrounding language reclamation, revitalization and reinvigoration". Zuckermann considers these terms as different points on a "revival cline or spectrum". Here, **Reclamation** is the revival of a no-longer spoken language ("Sleeping/Dreaming Beauty"), the best-known case being Hebrew; **Revitalization** is the revival of a severely endangered language, for example Adnyamathanha of the Flinders Ranges in Australia; and **Reinvigoration** is the revival of an endangered language that still has children speaking it, for example the Celtic languages Irish and Welsh. Zuckermann argues at length that it is helpful to adopt a broad perspective, both when considering the above as instances of a single set of issues, and when considering revivalistics as an inherently trans-disciplinary field of inquiry, which by its nature involves not only linguistics and language technology but also anthropology, sociology, politics, law, mental health and other disciplines. To help a language that is under threat, it is necessary to consider why it is under threat, what the consequences are for the (current and potential) speakers of the language, what their motivation is, if any, for wanting to strengthen the language, and what in practice can be done.

In this paper, we will be most immediately concerned with language and language technology, but the other aspects are also implicitly present.

2.2 LARA

LARA[1] (Learning and Reading Assistant) is a collaborative open source[2] 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 online.

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 listening skills. In particular, LARA texts include translations and human-recorded audio (video, in the case of sign languages) attached to words and sentences, and a personalised concordance constructed from the learner's reading history. An important point is that the concordance is organised by lemma, rather than by surface form; this requires the LARA text to be marked up so that each word is annotated with its associated lemma, a process which for many languages can be performed semi-automatically with an integrated third-party tagger/lemmatizer doing most of the work (Akhlaghi et al., 2020).

From the user perspective, the consequence of the above is that 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 (look like, in the case of sign languages), and where have I seen some form of the word before. Figure 1 shows an example for an Irish text. Students can test their knowledge of a text using several kinds of automatically generated flashcards, with a new set of flashcards created on each run (Bédi et al., 2020).

Related platforms, from which the project has adapted some ideas, include Learning With Text[3] and Clilstore[4]. The LARA tools are accessed 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; on top, there is a web layer implemented in PHP. Comprehensive online documentation is available (Rayner et al., 2020).

In the following sections, we describe how LARA is being used to create resources for the three languages which are our main focus in this paper.

---

[1] https://www.unige.ch/callector/lara/
[2] https://sourceforge.net/projects/callector-lara/

[3] https://sourceforge.net/projects/lwt/
[4] http://multidict.net/clilstore/

Fairceallach Fhinn Mhic Chumhaill (1)

bí (3)
(Verb)

Lá dá raibh bhí Fionn agus na Fianna ag fiach. (2) Chaitheadar an lá ó mhoch na maidne go dul faoi na gréine agus ar a gcasadh dóibh abhaile d'iompaigh siad siar féachaint an r... O... Once upon a time, Fionn and na Fianna were out hunting. ... ina stumpa beag ramhar, chucu. Bhí siúl maith acu á dhéanamh agus dá fheabhas é an siúl a bhí acu ní rabhadar ag imeacht aon ní ón bhFairceallach. Bhí sé ag coimeád leo i gcónaí; faoi mar a bhídís ar an ardán, bhíodh an Fairceallach san ísleán, agus nuair a bhídís san ísleán bhíodh an Fairceallach san ardán. Choinnigh sé ina ndiaidh mar sin chun go dtángadar go dtí a mbunáit féin. Shuídodar 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 chúinne. Chonaic Fionn é agus thug sé chun boird é agus chaitheadar a sásamh.

← 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 gcasadh dóibh abhaile d'iompaigh siad siar féachaint an raibh Oisín ina ndiaidh.
← Chonaic siad Fairceallach, fear a bhí ina stumpa beag ramhar, chucu.
← Bhí siúl maith acu á dhéanamh agus dá fheabhas é an siúl a bhí acu ní rabhadar ag imeacht aon ní ón bhFairceallach.
← Bhí sé ag coimeád leo i gcónaí; faoi mar a bhídís ar an ardán, bhíodh an Fairceallach san ísleán, agus nuair a bhídís san ísleán bhíodh an Fairceallach san ardán.
← Shuídodar 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'imíodar 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 tsiúl céanna i gcónaí agus é á leanúint.
← D'fhan sé mar sin ina dteannta nó go raibh lá agus bliain tugtha aige.
← Dúirt sé ansin: "A Fhinn Mhic Chumhaill", ar seisean, "tá cuireadh dinnéir agam á thabhairt duit féin agus d'Fhianna Éireann".
← "Ó", ar seisean, "níl aon fhear i seacht gcatha na nGnáth-Fhéinne nach gcuirfidh mé scian agus forc agus pláta ar a aghaidh amach! (5)
← 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 uillinn amach uirthi.
← Bhí cúl breá gruaige uirthi ar dhath an óir síos go caol a droma.
← Bhí dhá shúil ghlasa ghorma aici agus muince uirthi chomh geal agus a chonaiceadar ar aon bhean riamh.

Figure 1: Example of Irish LARA content, *Fairceallach Fhinn Mhic Cumhaill*, ('Fionn's burly friend'). 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 content page.

3 Irish (Gaelic)

3.1 The context

Irish belongs to the Goidelic branch of the Celtic languages. It is community language spoken in relatively small regions (Gaeltacht regions), primarily in the West of Ireland, with daily speaker numbers of about 20,586, or 0.43% of the Irish population (CSO, 2016). Note, however, that there are no monolingual communities, and even in the Gaeltacht, English is increasingly dominant. Outside these rather small and scattered communities, Irish is spoken as a first language in individual households, mostly in urban areas.

As a minority and endangered language, Irish is perhaps unique in being the recognised first official language of the State, and since its foundation, the revival of Irish has been a national policy. It has always been implicitly understood that the teaching of Irish on a national scale is the central plank of language maintenance. Irish is an obligatory subject in schools, and there is a thus a large cohorts of learners (c. 700,000 learners in the education system in the Republic of Ireland (Ní Chiaráin and Ní Chasaide, 2020)). Irish language teaching in this context poses numerous challenges. Some are discussed here, as a prelude to discussion of how LARA can help address them.

3.2 Challenges in the teaching of Irish

A major challenge is that the teachers are typically second language learners themselves, and their own command of the language (or their confidence in it) can be problematic. Teachers 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). One area of particular difficulty is pronunciation (see more below) – as learners typically lack native speaker models, they tend to acquire spoken forms very far from those of Gaeltacht speakers.

Other problems arise with language-teaching resources and teaching approaches: materials are often outdated for the modern child — particularly as Irish lives in the shadow of English, for which endless exciting resources are at people's fingertips. Inevitably, there are issues of student attitude and motivation, related to the modality and

the content of language teaching materials. There is a need for materials adapted to local context, materials that are exciting and varied for the young digitally-savvy child, and materials that encourage 'reading for fun'.

Social media has gone a long way to connect the Irish-language community generally in recent years, and this in turn has benefitted the morale of Irish language teachers. Nonetheless, they often feel isolated and left to their own strategies and feel the lack of a shared pool of high-quality resources: there is a great need for a community of users and contributors that might support them and the learner.

3.3 Linguistic Challenges

Quite apart from the sociolinguistic context of Irish learning, there are many linguistic features in the language that are very different from English, and therefore challenging to the L1 English-speaking learner. Firstly, although there is an agreed standardised written form, there is no single spoken standard, but rather three major dialects.

As mentioned, the pronunciation of Irish is very different from English. 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 rather opaque writing system, and the link of the sounds to written forms is often poorly understood. There are numerous other features. The basic word order is VSO. 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]. And the language is highly inflected (e.g., verbal forms can have as many as 42 inflected forms).

3.4 LARA and Irish

A major feature of LARA for Irish is the integration of synthetic voices for spoken output. This is made possible due to the recent development of synthetic voices for the main Irish dialects by the ABAIR project (they are available at www.abair.ie). The Irish co-author of this paper is directly involved with this initiative (ABAIR, 2020).

The integration of native-quality synthetic speech, rather than relying on pre-recorded speech is particularly interesting for the Irish context for many reasons. Firstly, it allows for rapid development of content by teachers, with native-speaker speech output, that essentially bring the native speaker model into the classroom – available to learners at all times. It thus provides not just one exemplar, but a potential environment where speech is integral to numerous reading activities with the consequence that children are constantly exposed to the native spoken language and developing native speaker pronunciation as a norm. For native-speaker children in the Gaeltacht, the fact that the choice of dialects is available means that the local teacher can add the appropriate speech output to assist with reading acquisition. For learners both within and outside of the Gaeltacht, the prevalent use of speech output in reading materials means that the linkage of pronunciation and written forms is being reinforced while reading, and this should greatly assist learners to gain an intuitive grasp of the phonic regularities of the written language – crucial to literacy acquisition.

The flexibility afforded by the inclusion of the synthetic voices offers the possibility for creative teachers to expand the diversity of materials, appropriate to the cohort they are teaching, and varied to maintain their interest in an ongoing way. The content of teaching materials can be adapted to the local context and to the topics that are currently of interest to the learners, and content can even be personalised to suit the wide diversity of individual learners' interests and levels. It is hoped that this widening of the palate of offering for the learner may increase the motivation of learners of different ages, levels and learning styles – key to a positive attitude and better learning outcomes.

The other features of LARA are also valuable in making more complex content accessible because of the level of scaffolding provided by the dictionaries, etc. so that learners can browse materials of interest, which might be beyond their language level. The linkage provided to the lemma forms and the access to all forms of a given lemma provides constant reinforcement, helping to relate words that may look and sound like very different items – helping to consolidate the learner's implicit knowledge of grammatical forms. While the explicit teaching of grammar and vocabulary has tended to be out of favour in recent teaching approaches, to the detriment of students' literacy accuracy, LARA is a tool that can be used to help direct learner's attention to grammatical features,

in a way that may not be felt as onerous by the learner.

The emphasis in LARA on identifying the more frequent words is another helpful feature, as it allows vocabulary acquisition to be maximally useful for the learner, directing their attention to the core vocabulary of the language, and giving a sense of mastery.

From the teachers' perspective, the Irish version of LARA, with its synthetic speech facility also enables the creation of a community of learners (classrooms / people with shared interests). LARA, along with other educational speech-based resources, is a move in the right direction to empowering the community of Irish language teachers by connecting them to powerful new resources with which to engage their students. In the future it also promises a mechanism for sharing high quality resources, to the benefit of the broader community of Irish language teachers. It also has the potential to engage the young digital generation in a way that traditional materials are now less effective for.

4 ÍTM/Icelandic Sign Language

4.1 Overview, history and social context

Icelandic Sign Language (*íslenskt táknmál*; ÍTM) is a natural language and the first language of the Deaf and their children in Iceland. ÍTM is an indigenous minority language (Stefánsdóttir et al., 2019). The Deaf community in Iceland defines itself as a linguistic minority, i.e. not on the basis of biological or medical deafness but from the language they consider their mother tongue (Stefánsdóttir, 2005; Sverrisdóttir, 2007). ÍTM was acknowledged as the first language of the Deaf, hard of hearing and deaf-blind people with the establishment of Act No. 61/2011 on the Status of the Icelandic Language and Icelandic Sign Language. According to the 2015 annual report from the Icelandic Sign Language Council (Stefánsdóttir et al., 2015) and to Stefansdottir, Kristinsson and Hreinsdottir (2019), ÍTM is an endangered language and the language community is still experiencing discrimination. In their 2019 article, Stefansdottir et al stress that ÍTM is still associated with disability and impairment.

ÍTM is the first language of 250–300 Icelanders (Report of the committee on the judicial status of Icelandic and Icelandic Sign Language, 2010:86; (Brynjólfsdóttir et al., 2012; Þorvaldsdóttir and Stefánsdóttir, 2015; Stefánsdóttir et al., 2019). According to (Þorvaldsdóttir and Stefánsdóttir, 2015), The Communication Centre for the Deaf and Hard of Hearing (the Center) in Iceland provided interpreter service to 178 Deaf signers according to the latest data from 2013. They point out that not all Deaf signers use the interpreter service from the Center, so the number of native signers may be higher. In addition, there are about 50 Deaf immigrants in Iceland who use ÍTM (Stefánsdóttir et al., 2019). The authors estimate that there are around 1000–1500 hearing L2 users of ÍTM.

Atypically, hereditary deafness barely exists in Iceland. It is to be found only in three families, two going only one generation back. On the other hand all Deaf immigrant families who have Deaf children report hereditary deafness several generations back (the Center, Ivanova). This has given rise to a unique situation where 75% of the children under 9 who use ÍTM are second generation Deaf of Deaf immigrants, and all of them are born in Iceland (Koulidobrova and Ivanova, 2020). They grow up unimodal bilingual with the heritage sign language and ÍTM; they learn written Icelandic when they start school, and it is their L3 or L4 depending on whether they also learn the written language of the heritage sign language. Koulidobrova and Ivanova (2020) report that 25% of the signing children under the age of 9 have access to sound by hearing aids and hearing parents. They grow up bimodal bilingual, with spoken Icelandic and ÍTM. Taking all facts into account, an unavoidable impact is that ÍTM changes because the youngest generation of native signers is second generation Deaf with Deaf immigrant parents, and there is only one Deaf family with Deaf children with the other parent being of Icelandic origin. According to the typology of §2.1, the fact that hardly any child signers of ÍTM have parents who are native ÍTM signers places it firmly in the "revitalization" category.

The first school for the Deaf in Iceland was established in 1867 by the Rev. Páll Pálsson (Þorvaldsdóttir and Stefánsdóttir, 2015). Before that time 24 deaf children were sent to Denmark of whom 16 came back (Kristjánsson). Until 1922 when the Act on education of deaf and mute children[5] came into effect, parents of deaf children could choose whether their Deaf child would be

---

[5]Lög um kennslu heyrnar- og málleysingja nr. 24/1922

educated in Iceland or in Denmark. After 1922 all Deaf children were required to be educated in Iceland from the age of eight and their education was compulsory. Even though some children were educated in Denmark before 1922, there is insufficient scientific evidence for the claim that ÍTM is derived from Danish Sign Language. Research (Aldersson and McEntee-Atalianis, 2007; Sverrisdóttir and Þorvaldsdóttir, 2016) shows similarities in the lexicon between ÍTM and DTS. Those are due to borrowing but not of a genetic relation (Sverrisdóttir and Þorvaldsdóttir, 2016). As Thorvaldsdottir and Stefansdottir point out (2015), ÍTM has been in contact with other spoken and sign languages which have influenced its development. Today, there are 13 foreign sign languages used in Iceland.

ÍTM does not have dialects or genderlects, but there is significant age-graded variation (Þorvaldsdóttir and Stefánsdóttir, 2015). The lack of dialects may be explained by the fact that both the preschool and the compulsory school for the Deaf have always been in Reykjavík, as are the Association for the Deaf and the interpreter services provided by the Center. More research is needed on ethnolects in ÍTM today resulting from the Deaf immigrants. ÍTM is taught as an L2 in different type of courses at the Center and in the Sign Language Studies BA program at the University of Iceland. There is a lack of teaching materials and tools for ÍTM as an L1.

4.2 Research

The grammar of ÍTM has been researched less than many other sign languages. Even if research on ÍTM has been ongoing for 30 years, only certain parts of the language have been studied (Sverrisdóttir, 2000; Aldersson and McEntee-Atalianis, 2007; Thorvaldsdóttir, 2007, 2008; Þorvaldsdóttir, 2011; Brynjólfsdóttir, 2012; Beck, 2013; Brynjólfsdóttir et al., 2012). Research at the Center has been mainly of a practical nature; much of the information that has been gathered has not been published in journals. The Center for Sign Language Research was established in 2011 with the aim of enhancing theoretical grammar research and cooperation between the Center and the Institute of Linguistics at the University of Iceland.

There is currently no proper corpus for ÍTM, but such a resource would greatly help the survival of the language. Most of existing video recordings are in practice difficult to use; since as they are not annotated, it is hard to search for examples in them. Some first steps towards building up an annotated corpus for ÍTM at the Center were taken with two short-term projects in 2014 and 2017 funded by the Student Innovation Fund (the Center, p.c. Thorvaldsdottir). All data and results are kept at the Center and in the light of the information above it is very important that they are used for ongoing work in order to preserve and document the language, follow its historical development, and secure the survival of ÍTM.

4.3 LARA and ÍTM

There are several different scenarios involving language learning and ÍTM that can potentially make use of LARA: (1) Deaf people learning written Icelandic; (2) Hearing people learning ÍTM as L2; (3) Deaf immigrants learning ÍTM as L2; (4) Deaf children learning ÍTM as L1; (5) Hearing children learning ÍTM as L1 (Coda). The first case is the most straightforward one, and formed a logical starting point. As noted above, written Icelandic is in most cases an L3 or even an L4 for Deaf children in Iceland, so tools that can help them make progress in reading are potentially very useful.

It turned out to 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 audio. Thus a LARA text of this type is written in Icelandic, but words and sentences are associated with 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 docu-

ment 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 (Brynjólfsdóttir, 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.

5 Barngarla

5.1 The context

Barngarla is a dreaming, sleeping beauty tongue belonging to the Thura-Yura language group, which also includes Adnyamathanha, Kuyani, Nukunu, Ngadjuri, Wirangu, Nawoo, Narangga, and Kaurna. The name *Thura-Yura* derives from the fact that the word for 'man, person' in these languages is either *thura* or *yura* —- consider Barngarla *yoora*. The Thura-Yura language group is part of the Pama-Nyungan language family, which includes 306 out of 400 Aboriginal languages in Australia, and whose name is a merism derived from the two end-points of the range: the Pama languages of northeast Australia (where the word for 'man' is *pama*) and the Nyungan languages of southwest Australia (where the word for 'man' is *nyunga*). According to (Bouckaert et al., 2018), the Pama-Nyungan language family arose just under 6,000 years ago around Burketown, Queensland.

Typically for a Pama-Nyungan language, Barn-

garla 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 (*warraidya*, 'emu'; *warraidyalbili* 'two emus'; *warraidyarri* 'emus'; *warraidyailyarranha*, 'a lot of emus') and matrilineal and patrilineal distinction in the dual. For example, the matrilineal ergative first person dual pronoun *ngadlaga* ('we two') would be used by a mother and her child, or by a man and his sister's child, while the patrilineal form *ngarrrinyi* would be used by a father and his child, or by a woman with her brother's child.

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 on September 14 2011 in a meeting between Zuckermann and representatives of the Barngarla people (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).

5.2 Using LARA for Barngarla

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 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. Links to the freely available LARA text are posted on the LARA examples page[6] in two versions, one recorded by Zuckermann and one by ethnic Barngarla language custodians.

A second resource was produced as part of the

"Fifty Words Project"[7], 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 both on the Fifty Words and the LARA examples pages.

Two more LARA resources for Barngarla are as of early 2021 nearly complete. We describe them in the next section.

6 Summary and further directions

We have described work where the LARA platform has been used to construct annotated multimodal resources for three endangered languages, outlining the relevant background and the our reasons for believing that LARA has strong potential for assisting in various kinds of language documentation and "revivalistics" efforts, and describing some early examples of LARA resources for these languages. LARA documents can be produced quickly and easily (Akhlaghi et al., 2020),[8] and more ambitious efforts are in an advanced stage of preparation. We briefly outline ongoing work here.

In the case of Irish, the immediate intention is to seek to engage the community of learners, and then, increasingly the community of teachers, involving them in the development of LARA resources for their own teaching contexts. The current provision is for eight short stories, spanning the three major dialects. These are based on episodes from the mythology sagas of the band of warriors Na Fianna, adapted from historic collections (see https://www.abair.tcd.ie/scealai/#/resources). These stories, prepared in Summer 2020, are currently being used with a third level group of various backgrounds and levels approximating B2 level, who are pursuing an advanced level module in 'Irish Language and Culture'. Although the texts are fairly difficult, the LARA features allow individuals to take what they need from it. It facilitates teaching to a mixed-level group, and supports nonlinear progression of language learning. The platform will be evaluated by this learner group, and learner feedback sought concerning how engaging and user-friendly they found the multimodal text presentation; how useful they found the specific LARA features; and whether they found it encouraged them to be more active in their learning.

As a second step, the teacher community will be targeted. The aim is not only to get a wider user group and to elicit feedback but also to involve groups of teachers in the creation of materials that are targeted at their own learner cohorts and contexts. As with other learning platforms being developed, the long term goal is for a community of teachers contributing to a growing bank of shared resources, catering for learners at every level – a need already identified by teachers (cf. §3).

With regards to Irish teachers adapting their own materials it should be noted that the integration of TTS has, to date, been done in an ad hoc way, using the command-line interface. We hope to streamline the process so that in future only a very small investment of time would be required for a teacher to adapt their resources into a LARA format. The part-of-speech tagging and synthesis steps are unsupervised, therefore once the content is fixed a short story could take as little as 10 minutes to convert to LARA format. Teachers may wish to proof-listen to the TTS output and, if necessary, correct the transcriptions, which could take some time. If they wished, teachers could also add their own sentence-level translations, another task where the time commitment depends on the length of the text. A more principled solution to integrate TTS capability into LARA is currently being investigated and we hope will be ready to roll out in 2021.

Looking to the future, the ABAIR initiative is aiming to document, and produce virtual speakers, i.e., synthetic voices for even the most endangered dialects. (Although in §3 we refer to the three major dialects of Irish, there are also subdialects which are severely endangered as well as dialects no longer spoken.) It is even envisaged that it may be possible to 'resurrect' recently deceased dialects such as the Irish of Co. Clare, no longer spoken, but for which recordings do exist. In the shorter term, LARA provides a tool that can be purposed to permit access to Clare texts and recordings and should help bring this dialect alive – in a way that could engage young people from that region and beyond. In these ways, the case

---

[7]https://50words.online/

[8]Special considerations extraneous to LARA may considerably increase the workload. For the ÍSL text described here, by far the most laborious task was creating the sign language videos. For the Barngarla texts, well over half the effort was spent adding the complex HTML formatting needed to reproduce the layout used in the paper version.

TÍNA	BAKDYR	KOMA-INN	MAMMA	HVAR 🎥
Tina	back-door	come in, classifier	Mom + start Q non-manual	where, Q non-manual

("Tina comes in through the back door. 'Where are you, Mom?' she asks")

Figure 2: ÍTM sentence from 'pure ÍTM' version of *Tína* story, lines as follows: (1) sequence of ID-glosses, (2) annotations, (3) translation. Only the first line is shown. Hovering over an ID-gloss shows a popup with the corresponding annotation. Clicking on an ID-gloss plays a video for the sign and shows the concordance. Clicking on a camera icon plays signed video for the whole sentence.

(54a) Ngarrinyelbudninge ninna Parnkalliti

Ngarrinyarlboodninga		nhina	Barngarlidhi. 🔊	
ngarrinyarlboo-	*dninga*	*nhina*	*Barngarla-*	*dha* 🔊
us (all)-	with	you	Barngarla-	PRES

'Through us or with us you are Parnkalla [you have learned the language from us].'

Figure 3: Example of Barngarla sentence from LARA version of (Schürmann, 1844), lines as follows: (1) sentence in original Schürmann orthography, (2) sentence in modern Barngarla orthography, (3) sentence with words decomposed into morphemes, (4) glosses for morphemes, (5) translation. All five lines are shown on the screen. Hovering over a word or morpheme plays recorded audio and shows a popup with a list of all possible associated glosses in the text; clicking on a word or morpheme shows a concordance; clicking on a loudspeaker icon plays audio for the whole sentence.

use here is not that distant from the Barngarla use.

For ÍTM, the next goal is to develop methods for creating LARA documents for true ÍTM texts, as opposed to Icelandic texts with ÍTM annotations (cf. §4). The immediate problem is that sign languages lack a written form. In practice, a signed utterance is represented as a sequence of manual (hand) sign representations, typically accompanied by a parallel line showing non-manual (non-hand) signs, most often facial expressions. Lexical signs are conventionally written in uppercase as "ID-glosses" (Johnston, 2008). Signs which are not lexicalized, usually "classifier constructions"[9], are written in lowercase as short general descriptions of what the sign stands for.

The approach used to represent signed documents in LARA is to make the "text" the sequence of ID-glosses and classifier descriptions, and attach information to these elements appearing as popups. An ÍTM document of this kind is currently under construction. Figure 2 illustrates.

As of early 2021, two new LARA texts for Barngarla are nearly complete. The first of these is a LARA version of a language documentation and revivalistic project (Zuckermann, 2021), which presents the full set of 708 Barngarla sentences extracted from (Schürmann, 1844). Each sentence is shown using the interlinear form illus-

trated in Figure 3. An advantage of LARA is that the text can easily be produced in two forms: one is designed for the professional linguist, the other tailored for ethnic Barngarla readers.

A second Barngarla text, *Mangiri* (Zuckermann and the Barngarla, 2021) has been designed as a teaching resource. In contrast to *Barngarlidhi Manoo* (cf. §5), which is almost exclusively focused on vocabulary, *Mangiri* introduces some nontrivial grammar using annotated example sentences in the same interlinear form as (Zuckermann, 2021). Many of these examples are taken from translations of English and Hebrew songs based on Biblical verses.

In a broader perspective, this thread of work can be seen as an interesting process of turning historical injustice against itself. A book, written in 1844 in order to help a German Lutheran missionary bring the "Christian light" to Aboriginal people at the expense of their own spirituality, is used to do the opposite thing, and reunite the Barngarla people with their own heritage; Christian texts are being translated to provide attractive linguistic examples in reclaimed Barngarla; and technology used during the colonial and Stolen Generation periods to oppress indigenous people is used in the form of LARA to help them recover their own autonomy, spirituality and well-being. These thoughts encourage us to proceed further down the path we have begun to explore.

---

[9]"Classifiers" can most easily be thought of as short semi-lexicalized playlets performed with the hands. They are common in all sign languages.

References

ABAIR: An Sintéiseoir Gaeilge – The Irish Language Synthesiser ABAIR. 2020. http://www.abair.ie. As of 8 September 2020.

Farhia Ahmed, Pierrette Bouillon, Chelle Destefano, Johanna Gerlach, Angela Hooper, Manny Rayner, Irene Strasly, Nikos Tsourakis, and Catherine Weiss. 2016. Rapid construction of a web-enabled medical speech to sign language translator using recorded video. In *Proceedings of FETLT 2016*, Seville, Spain.

Elham Akhlaghi, Branislav Bédi, Fatih Bektaş, Harald Berthelsen, Matthias Butterweck, Cathy Chua, Catia Cucchiarini, Gülşen Eryiğit, Johanna Gerlach, Hanieh Habibi, Neasa Ní Chiaráin, Manny Rayner, Steinþór Steingrímsson, and Helmer Strik. 2020. Constructing multimodal language learner texts using LARA: Experiences with nine languages. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 323–331.

Elham Akhlaghi, Branislav Bédi, Matthias Butterweck, Cathy Chua, Johanna Gerlach, Hanieh Habibi, Junta Ikeda, Manny Rayner, Sabina Sestigiani, and Ghil'ad Zuckermann. 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.

Russell R. Aldersson and Lisa McEntee-Atalianis. 2007. A lexical comparison of Icelandic sign language and Danish sign language. *Birkbeck Studies in Applied Linguistics*, 2:123–158.

Þórhalla Guðmundsdóttir Beck. 2013. *Í landi myndanna. Um merkingu og uppruna lýsandi orða í táknmáli*. Ph.D. thesis.

Branislav Bédi, Matt Butterweck, Cathy Chua, Johanna Gerlach, Birgitta Björg, Hanieh Habibi Guðmarsdóttir, Bjartur Örn Jónsson, Manny Rayner, and Sigurður Vigfússon. 2020. LARA: An extensible open source platform for learning languages by reading. In *Proc. EUROCALL 2020*.

Remco R. Bouckaert, Claire Bowern, and Quentin D Atkinson. 2018. The origin and expansion of Pama–Nyungan languages across Australia. *Nature ecology & evolution*, 2(4):741–749.

Elísa Guðrún Brynjólfsdóttir. 2012. *Hvað gerðir þú við peningana sem frúin í Hamborg gaf þér? Myndun hv-spurninga í íslenska táknmálinu*. Ph.D. thesis.

Elísa Guðrún Brynjólfsdóttir, Jóhannes Gísli Jónsson, Kristín Lena Thorvaldsdóttir, and Rannveig Sverrisdóttir. 2012. Málfræði íslenska táknmálsins. [The grammar of Icelandic Sign Language]. *Íslenskt mál og almenn málfræði*, 34:9–52.

Matt Butterweck, Cathy Chua, Hanieh Habibi, Manny Rayner, and Ghil'ad Zuckermann. 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.

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.

Claire M. Dunne. 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.

Trevor Johnston. 2008. Corpus linguistics and signed languages: no lemmata, no corpus. In *3rd Workshop on the Representation and Processing of Sign Languages*.

Elena Koulidobrova and Nedelina Ivanova. 2020. Acquisition of phonology in child Icelandic Sign Language: Unique findings. *Proceedings of the Linguistic Society of America*, 5(1):164–179.

Stephen Krashen. 1982. Principles and practice in second language acquisition.

Ó.P. Kristjánsson. Málleysingjakennsla hér á landi. [the teaching of the mute here in the country]. *Menntamál*, 18(1):7–17.

Neasa Ní Chiaráin and Ailbhe Ní Chasaide. 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.

Manny Rayner, Hanieh Habibi, Cathy Chua, and Matt Butterweck. 2020. *Constructing LARA content*. https://www.issco.unige.ch/en/research/projects/callector/LARADoc/build/html/index.html. Online documentation.

Clamor Wilhelm Schürmann. 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*.

V. Stefánsdóttir, A.P Kristinsson, H.D. Eiríksdóttir, H.A. Haraldsdóttir, and R. Sverrisdóttir. 2015. Annual report of the Icelandic Sign Language council. Technical report. In Icelandic. Available at https://www.islenskan.is/images/Skyrsla-MIT-2015.pdf.

Valgerður Stefánsdóttir. 2005. *Málsamfélag heyrnarlausra: Um samskipti á milli táknmálstalandi og íslenskutalandi fólks*. Ph.D. thesis.

Valgerður Stefánsdóttir, Ari Páll Kristinsson, and Júlía G Hreinsdóttir. 2019. Legal recognition of Icelandic Sign Language: Meeting Deaf people's expectations? *The Legal Recognition of Sign Languages: Advocacy and Outcomes Around the World*, page 238.

Rannveig Sverrisdóttir. 2000. Signing simultaneous events: The expression of simultaneity in children's and adults' narratives in Icelandic Sign Language. Master's thesis, University of Copenhagen.

Rannveig Sverrisdóttir. 2007. Hann var bæði mál- og heyrnarlaus: um viðhorf til táknmála [he was both mute and deaf: On attitudes towards sign languages]. *Ritið*, 7(1):83–105.

Rannveig Sverrisdóttir and Kristín Lena Þorvaldsdóttir. 2016. Why is the sky blue? on colour signs in Icelandic Sign Language. *Semantic Fields in Sign Languages Colour, Kinship and Quantification. Berlin, Boston: De Gruyter*, pages 209–250.

Guðný Björk Thorvaldsdóttir. 2007. The use of space in Icelandic Sign Language. Master's thesis, The University of Dublin.

Guðný Björk Thorvaldsdóttir. 2008. Tillaga um nýja íslenska táknmálsorðabók á málvísindalegum forsendum [proposal for a new Icelandic Sign Language dictionary on linguistic terms]. Master's thesis, University of Iceland. Urlhttps://skemman.is/handle/1946/3452.

Ghil'ad Zuckermann. 2020. *Revivalistics: From the Genesis of Israeli to Language Reclamation in Australia and Beyond*. New York: Oxford University Press.

Ghil'ad Zuckermann. 2021. *Barngarla sentences*. Barngarla Language Advisory Committee (BLAC). In preparation.

Ghil'ad Zuckermann and the Barngarla. 2019. *Barngarlidhi Manoo: "Speaking Barngarla Together"*. Barngarla Language Advisory Committee (BLAC).

Ghil'ad Zuckermann and the Barngarla. 2021. *Mangiri: "Barngarla Wellbeing"*. Barngarla Language Advisory Committee (BLAC). In preparation.

Kristín Lena Þorvaldsdóttir. 2011. *Sagnir í íslenska táknmálinu. Formleg einkenni og málfræðilegar formdeildir*. Ph.D. thesis.

Kristín Lena Þorvaldsdóttir and Valgerður Stefánsdóttir. 2015. Icelandic Sign Language. *Sign languages of the world: a comparative handbook*, page 409.