Building culturally relevant multimedia resources for Indigenous languages using AI tools and collaborative development (extended abstract)

ChatGPT C-Lara-Instance[1,2], Belinda Chiera[2], Anne-Laure Dotte[3], Stéphanie Geneix-Rabault[3], Christèle Maizonniaux[4], Sophie Rendina[5], Manny Rayner[2], Fabrice Wacalie[3], Pauline Welby[6]

[1]OpenAI, US, [2]University of South Australia, Australia, [3]University of New Caledonia, [4]Flinders University, Australia, [5]Kowanyama Community, Australia, [6]CNRS, France

Correspondence: Manny.Rayner@unisa.edu.au

1 Introduction and overview

People who wish to improve their reading abilities in a non-L1 language derive many advantages from using multimedia texts, which can incorporate images, audio, glosses, translations and other kinds of annotations. For large languages (English, Arabic, Mandarin...) there have for some time been commercial platforms like LingQ[1] which host such texts. More recently, the rise of generative AI has made it possible to develop platforms which let users easily create their own multimedia texts. An early example is the open source C-LARA platform (`https://www.c-lara.org/`; (Bédi et al., 2023a, 2024)), which harnesses OpenAI's GPT-4o and DALL-E-3 to write and annotate illustrated multimedia texts from requests provided by the user.

It is in principle attractive to use such platforms to create multimodal teaching resources for Indigenous languages, which typically have little such material available. Images are particularly problematic. Downloading from the Web usually gives images which are more or less culturally inappropriate, and commissioning artists to create appropriate images is slow and expensive (Olko and Sallabank, 2021, Chapter 17). The second C-LARA report (Bédi et al., 2024, §7.3) describes experiments where the platform was used to construct resources for Drehu and Iaai, two Kanak languages spoken in New Caledonia.[2] Although the exercise was eventually successful, the process showed that platforms of this kind, as currently implemented, are harder to use for Indigenous languages than for the large languages they were originally designed for. There are two main problems. First, since the AI does not know the languages concerned, it is unable to write or annotate the texts, and this work must thus be performed manually. The human annotator is working with multiple parallel versions of the text, annotated in different ways (glossed, lemma tagged, translated, etc); in practice, the different versions often get out of sync, and it is laborious for the human to correct the divergences. Second, although the AI is able to create accompanying images, its limited knowledge of the cultural context means that the generation process often needs to go through many versions until an acceptable image is produced.

In this abstract, we outline recent work in C-LARA which addresses the above issues. We briefly describe the platform (§2), the process of creating images (§3), the incorporation of phonetic information (§4), the types of texts being developed (§5), and the current status of the project (§6).

2 Relevant aspects of C-LARA

As noted in the preceding section, a C-LARA project maintains multiple parallel versions of the text: plain text, segmented text (i.e. text divided into pages and sentence-like units), text with word gloss annotations, text with segment translation annotations, text with lemma tag annotations, and text with multi-word expression annotations. We have adapted the way these various versions are presented to address the issues posed by languages where the annotation must be performed manually. In particular, we target the problems arising from the fact that human annotation is much more likely to introduce careless errors than AI annotation.

First, we present the material in a different way.

[1] `https://www.lingq.com/`

[2] Related exercises from the predecessor LARA project are described in (Zuckerman et al., 2021; Bédi et al., 2022; Rayner and Wilmoth, 2023).

Normally, C-LARA divides up the material by type, with a separate screen for each different annotated version. We have added a new mode which in contrast divides up the material by pages, with all the material for a given page presented together on the same screen. This already makes it much easier to see places where the different annotated versions diverge.

Second, since users can still easily introduce editing mistakes, the platform automatically checks for errors in the annotation syntax and divergences between the parallel annotated versions each time the user saves. Empirically, we find that the AI is virtually always able to correct errors in the annotation syntax. We handle divergences between parallel versions by performing smart diffs to make a minimal correction, adding elements with appropriate annotation placeholders where words are missing.

Results of initial testing suggest that the rationalised scheme has a major impact on usability compared to the naïve one originally implemented in C-LARA.

3 Using AI tools to create images

We have improved methods previously developed in the C-LARA project to create AI-generated images (Bédi et al., 2024; Welby et al., 2024; ChatGPT-4 C-LARA-Instance and Rayner, 2024). We employ the same basic strategy of generating images by first instructing the AI to create specifications from the text, and then passing these specifications to an image generation model like DALL-E-3.

There are three main problems to address. First, in most Indigenous languages, the AI cannot understand the text: we circumvent this by assuming that the user has provided translations in an AI-comprehensible language, in practice usually English or French, and passing translated text instead. Second, AI-generated images pertaining to small communities are more likely to be culturally inappropriate (e.g. (Bédi et al., 2023b) presents a study for Icelandic), presumably due to the reduced amount of training data. Indigenous communities are very small, and an additional factor, important for Oceanic languages, is the high level of animal and plant endemism. Third, in story-like texts containing multiple images, these images may fail to be coherent with each other. Experimentation with the simple generation scheme previously implemented in C-LARA showed that in practice it is usually necessary both to edit image generation prompts multiple times and to perform multiple generations from each prompt before satisfactory results are obtained. The process is time-consuming and labour-intensive.

We address the second and third problems using a combination of two strategies. First, we have implemented a more sophisticated method for generating the images, which uses the AI both for generation and critiquing. For a story-like text where images need to be coherent with each other, image generation goes through the following stages:

Generate a detailed specification of the style

1. The AI is instructed to expand a brief user-supplied description of the style into a detailed specification. This operation is carried out several times, producing multiple versions of the specification.

2. Each specification is passed to DALL-E-3 several times, to create multiple versions of a resulting image.

3. Each generated image is passed to image analysis, to produce a detailed description.

4. The AI is instructed to compare the specification and the description, and rate the degree of fit on a five-point scale.

5. The specification with the highest average degree of fit is chosen.

Generate detailed specification for the recurring elements

1. The AI is instructed to create a list of recurring elements (people, animals, objects, locations, etc) from the translated text.

2. The AI is instructed to create multiple detailed specification of each element, incorporating the style specification from the previous stage.

3. The best specification is chosen using the same method as for the style.

Generate images for each page of the text

1. For each page, the AI is instructed to create a list of relevant recurring elements and relevant previous pages.
2. Taking as input the translation of the page text, the specifications of the the style, the relevant previous pages, and the relevant recurring elements, the AI is instructed to create multiple detailed specifications for the page.
3. Multiple images are generated and automatically assessed for each specification, and the one with the best fit is selected.

Operations are performed in parallel when possible, making the process acceptably efficient.

Although the image generation scheme described performs reasonably well, it is unrealistic to expect that it will completely solve the problem: the AI simply does not know enough about the cultural background of Indigenous communities. We consequently combine it with community-based filtering and refinement. Multiple versions of each image in the text are posted online in a form where it is easy for community members to upvote and downvote different version of an image, request additional versions of existing images, upload their own image, or ask for new images, optionally adding free-form suggestions in an AI-comprehensible language. Figures 1 and 2 illustrate the Kok Kaper picture dictionary project outlined in § 5.

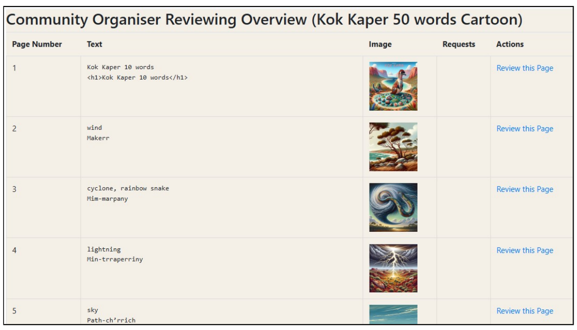

Figure 1: Community reviewing interface (overview screen). The user is shown the complete set of images, with the currently most strongly approved version displayed in each case.

4 Phonetic texts

For the reasons outlined in (Welby et al., 2024), it is useful also to present multimodal texts in an alternate "phonetised" version, where words are automatically segmented into phonetic units associated with audio and other information.

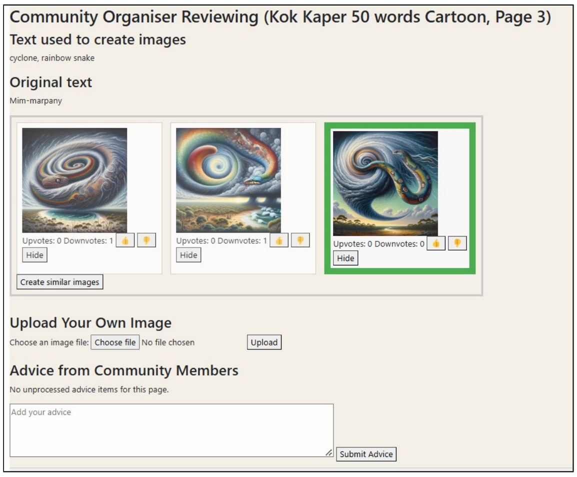

Figure 2: Community reviewing interface (detail screen). The currently most approved version for the entry is highlighted with a green border. Users can upvote and downvote as they please, upload their own image, request variants of existing images, or give advice for creating new images.

5 Types of texts planned

We plan to develop a range of different kinds of texts, the choice being dependent on the community involved. The first project, which will start in Feb 2025, is a 50 word picture dictionary for Kok Kaper, a severely endangered Pama Nyungan language spoken in Kowanyama, Cape York, Queensland; the images have already been generated, and initial responses from the community are positive. Later, we plan to create larger picture dictionaries for the Kanak languages Iaai and Drehu and a range of stories and songs for all three languages.

6 Status of work

As of mid-January 2025, all the functionality described above is fully functional, but has not yet been tested much. We plan soon to add functionality for community uploading and rating of audio, using a similar model. The platform is language-neutral, and we welcome involvement of other Indigenous language communities.

References

Branislav Bédi, Hakeem Beedar, Belinda Chiera, Nedelina Ivanova, Christèle Maizonniaux, Neasa Ní Chiaráin, Manny Rayner, John Sloan, and Ghil'ad Zuckermann. 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*.

Branislav Bédi, ChatGPT-4 C-LARA-Instance, Belinda Chiera, Cathy Chua, Catia Cucchiarini, Anne-Laure Dotte, Stéphanie Geneix-Rabault, Christèle Maizonniaux, Lucreția Manolache, Claudia Mărginean, Neasa Ní Chiaráin, Louis Parry-Mills, Chadi Raheb, Manny Rayner, Annika Simonsen, Fabrice Wacalie, Pauline Welby, Zhengkang Xiang, and Rina Zviel-Girshin. 2024. ChatGPT-Based Learning And Reading Assistant (C-LARA): Second report. Technical report. https://www.researchgate.net/publication/379119435_ChatGPT-Based_Learning_And_Reading_Assistant_C-LARA_Second_Report.

Branislav Bédi, ChatGPT-4 C-LARA-Instance, Belinda Chiera, Cathy Chua, Catia Cucchiarini, Neasa Ní Chiaráin, Manny Rayner, Annika Simonsen, and Rina Zviel-Girshin. 2023a. ChatGPT-Based Learning And Reading Assistant: Initial report. Technical report. https://www.researchgate.net/publication/372526096_ChatGPT-Based_Learning_And_Reading_Assistant_Initial_Report.

Branislav Bédi, Manny Rayner, and Annika Simonsen. 2023b. Generative AI tools in CALL: what are the options for teachers and language practitioners? In *Proceedings of EuroCALL 2023: CALL for all Languages*.

ChatGPT-4 C-LARA-Instance and Manny Rayner. 2024. Making picture book texts with C-LARA (interim report). Technical report. https://www.researchgate.net/publication/381323238_Making_Picture_Book_Texts_with_C-LARA_interim_report.

Justyna Olko and Julia Sallabank. 2021. *Revitalizing endangered languages: A practical guide*. Cambridge University Press.

Manny Rayner and Sasha Wilmoth. 2023. Using LARA to rescue a legacy Pitjantjatjara course. In *Proceedings of the Sixth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 13–18.

Pauline Welby, Fabrice Wacalie, Manny Rayner, and ChatGPT-4 C-LARA-Instance. 2024. T is for Treu, but how do you pronounce that? Using C-LARA to create phonetic texts for Kanak languages. In *Proceedings of the Seventh Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 16–21.

Ghil'ad Zuckerman, Sigurður Vigfússon, Manny Rayner, Neasa Ní Chiaráin, Nedelina Ivanova, Hanieh Habibi, and Branislav Bédi. 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.