

Community Dictionaries

Everyday Pictures, Rapid Human–AI Development

ChatGPT C-LARA-Instance¹ and Manny Rayner²

¹AI collaborator, C-LARA Project ²C-LARA Project, Adelaide, Australia

13 October 2026

Abstract

This paper is an invitation to try out a work in progress: Community Dictionaries, a working, mobile-friendly application for building language resources from a group’s own pictures, voices and experiences. During roughly two and a half weeks, repeated use and discussion have led to several substantial redesigns, with a strong AI writing the code, a human providing strategic direction, and a second human contributing initial content. We describe what users can do now, how the collaboration works, and the next experiment: whether we can keep improving the app quickly as more groups bring their own needs and feedback.

1. Start with something worth talking about

A photograph of a familiar cat on a familiar chair carries associations that a generic vocabulary illustration does not. Community Dictionaries starts from such images. One person can supply a picture, another its words and voice; the result belongs to a collection the group can build and use together.

Our initial material is almost entirely a personal collection assembled by Cathy Chua, the human author’s partner: photographs from home, walks and earlier occasions. She chose what interested her, rather than illustrating a prescribed word list. Rayner first created a Swedish dictionary by adding words and recordings; it now contains about 150 illustrated sentences with linked vocabulary. During development we added automatic language conversion to create French, German, Italian, Icelandic and Japanese versions. Conversion retains the photographs, uses both source text and images as input to AI translation, and generates new audio through text-to-speech (TTS). The app is open to everyone: browse these six versions without an account, or register to create your own dictionary:

<https://c-lara-2.c-lara.org/community-dictionaries/browse/>

2. What people can do now

Browse and practise. Tap a picture, listen to its sentence or words, reveal a translation when wanted, and follow links to other examples (Figures 1–2). Flashcards with multiple-choice answers, picture-clue crosswords and word-search puzzles reuse the dictionary’s material. Public browsing and these activities need no login or paid AI calls.

Make a dictionary together. Create an account and a dictionary, invite partners, and take or upload photographs. Add words, translations or explanations, and recordings directly through the browser. No programming or software installation is needed. Under the current hosted arrangement, creating dictionaries and adding human-produced material carry no platform charge. Phones suit capture, recording and browsing; laptops offer more room for editing.

Use AI where it helps. For supported languages, describe a picture by typing, speaking, or asking the AI to suggest a description. Alternatively, the owner can process a whole collection: approve one cost estimate and let the app create sentences, recordings and linked word pages automatically. The staged review option remains available. AI can also generate pictures in a chosen common style, or create another language version of a dictionary while retaining its pictures. Human contribution remains a complete route for languages with little or no useful AI support.

Pictures, sentences and words

These screenshots show parts of a French dictionary automatically converted from the Swedish original. The generated wording is retained; we discuss the details of this example in §4. The translation control in the first image has been opened.

FINLEY'S FRENCH · FRENCH · PUBLIC VIEW

Le chat est allongé sur un fauteuil.



Listen

▼ Translation

The cat is lying on an armchair.

AI-generated voice

Figure 1: A picture becomes something to say: *Le chat est allongé sur un fauteuil*, with the English explanation “The cat is lying on an armchair.” Listen plays the recording; Translation reveals the English explanation. Word controls continue below the pictured section. Photograph: Cathy Chua.

Sentences using this word

[Le chat est allongé sur un fauteuil.](#)

[Un chat dort dans un fauteuil.](#)

[Il y a un lit et deux fauteuils dans la chambre.](#)

Pictures for this word



Open picture →



Open picture →



Open picture →

Figure 2: The lower part of the word page for *fauteuil* (displayed English gloss: “armchair”). One word links to several sentences and pictures, including the scene above. The learner can move between a particular experience and other uses of the same word. Photographs: Cathy Chua.

3. An application designed to change

The immediate predecessor was C-LARA-2 [1], itself a rewrite of C-LARA, whose work included tailor-made pedagogical picture books [2]. Feedback from their user group prompted the new direction: Sophie Rendina and Branislav Bédi stressed simplicity and mobile access; Rendina particularly emphasised Indigenous communities' needs. Community Dictionaries rethinks the application around those requirements, while reusing C-LARA-2 infrastructure and language-processing tools.¹

C-LARA-2 was developed using successive versions of ChatGPT-5 and Codex. Positive media coverage and our first experiments with the newly available ChatGPT-6 Astra in Extra High mode suggested that a much faster development cycle might now be possible. We set out to test that possibility through a fresh implementation of the user-facing application. Our experience so far strongly supports the practical hope: a request for a major new feature or a substantial redesign has generally yielded a working, installable patch in less than an hour. This is a participant observation, not a controlled model comparison; installation, user testing and subsequent fixes take additional time.

The AI has written the code, tests and documentation. Rayner supplies direction, deploys versions, tests them and returns observations; Chua contributes images, testing and design ideas. The AI also proposes alternatives when a request would make the app unnecessarily complicated. Three redesigns illustrate the process:

- Pictures initially had single words and recordings. They gained multiple word links, then sentences connected to shared word pages.
- Chua wanted to supply pictures without doing the language work. Processing grew from individual pictures to batches, then to the automatic option with no intermediate review.
- Fine-grained withdrawal controls proved confusing. Iteration produced a simpler model: one control withdraws all a member's content from the dictionary and keeps it private, after confirmation; a single click restores it if desired. This easy reversibility is important for projects built from shared content.

In roughly two and a half weeks, use has repeatedly reshaped the app. Inherited software and human feedback have been essential to that speed, but in our assessment the most important factor has been Astra's greatly improved ability to implement and revise substantial pieces of software.

4. What the first version establishes

Rayner and Chua have used the app on laptops and a physical phone; Rayner reports successful AWS deployment of the automatic workflow. The release passed 654 regression tests with mocked AI calls. These checks and the public examples establish working functionality, not measured learning gains or independently validated linguistic accuracy.

Rayner's initial, non-exhaustive browsing suggests that over 95% of the AI-generated content is correct. This is an informal estimate; subtle problems remain, as the apparently simple example in Figure 1 illustrates. Finley is lying on a padded office chair without arms, so *armchair* is inaccurate. French usage is less clear-cut: dictionaries define *fauteuil* as having arms, but *fauteuil de bureau* is also used commercially for armless chairs.² We would prefer *chaise de bureau* and *office chair* here. The example shows why an ordinary translation pair such as *fauteuil/armchair* still needs checking against the particular image.

Automatic publication is labelled; fluent-speaker review remains valuable. Whether users return and learn is still open. Reducing expert intervention in maintenance is a further goal: Rayner still handles deployment and recovery.

¹Code and dated development records: <https://github.com/mannyrayner/C-LARA-2>.

²Compare the Académie française's definition with Concept Bureau's KIVO chair description.

5. Community control comes first

Dictionaries start private and unlisted. Owners can allow public browsing or advertise basic information for membership requests. Contributors can withdraw their material into a private collection and later restore it; this cannot recall downloads or erase backups. Optional AI sends selected material to external providers and needs an appropriate agreement.

For Indigenous-language work, hosting, access, reuse and external processing must be agreed with the community. App controls alone do not establish data sovereignty; community authority, emphasised in the CARE Principles, remains central.³ Earlier collaborative work on culturally relevant Indigenous-language resources provides context [3]. Whether the app meets a community's needs must be established with that community; our initial trials cannot answer this.

6. An invitation to help shape the next version

We invite learners, teachers, linguists and community groups to try the public dictionaries or make their own. Partners can share the picture-taking and language work. Tell us what confused you, what was missing, and whether you wanted to come back.

To get started, visit the address in §1, choose *Log in*, then *Create an account*. No invitation or prior contact is needed. You can also use the AI functionality with your own OpenAI API key, entered under *Profile*. Contact Rayner if you would like to discuss free pilot credit.

The dictionaries created in these trials, each with about 150 images and 250–400 word entries, have typically cost around US\$15–20 each in AI usage, by Rayner's rough estimate. This covers creating language content and audio from existing images; costs vary with the processing required.

For a pilot involving up to ten groups, we offer an initial US\$10 of AI credit per group, with further support up to US\$50 per group in total if we are allowed to inspect the initial dictionaries produced. We would not publish or pass specific content to third parties without separate permission, but would report agreed, non-identifying statistics such as entry counts. This is separate from authorised external AI processing.

Ordinary use does not enrol somebody in research. Before a study, we would agree what feedback and usage data can be collected and reported. Participants can choose named credit or anonymity, with authorship for substantial research contributions where appropriate and desired.

We are preparing this report for AiraXiv, the preprint platform with AI-supported review [4]. Its shorter turnaround than conventional publication supports our accelerated cycle of building, trying, reporting and revising. We aim to publish a rapid sequence of papers tracking the project. This is the first.

With willing participants, we hope our second paper will report initial empirical data in the next few weeks: actual use, difficulties, revisions, costs and human effort. Learning gains need a further study. Can the app remain simple and become more useful as people outside its initial development circle shape it?

To discuss free pilot credit: mannyrayner@c-lara.org.

³Global Indigenous Data Alliance, *CARE Principles*: <https://www.gida-global.org/careprinciples>.

Authorship and acknowledgements

ChatGPT C-LARA-Instance wrote and revised the entire text from discussions with Rayner and the project records, and carried out the implementation work described above. Rayner contributed conception, direction, testing, critique and operations, and is the human contact.

We thank **Cathy Chua** for the charming, personal picture collection that has provided almost all the initial test material, for testing, and for several important design ideas. We thank **Sophie Rendina** and **Branislav Bédi** for the discussions about users’ needs from which the idea grew, and Rendina especially for her guidance on community control. Finally, we thank **Finley**, the charismatic tabby who deigns to share his territory with Rayner and Chua, and who has set his indelible stamp on the original dictionaries.

References

- [1] Branislav Bédi, ChatGPT C-LARA-Instance, Belinda Chiera, Cathy Chua, Stéphanie Geneix-Rabault, Christèle Maizonniaux, Manny Rayner, Sophie Rendina, Sarah Wright, and Rina Zviell-Girshin (2026). *An AI-Based Platform That Can Explain Its Own Functioning: Why Your Project Should Be Fully Codex-Generated*. Preprint, July. Available on ResearchGate.
- [2] Branislav Bédi, Hakeem Beedar, Belinda Chiera, Cathy Chua, Stéphanie Geneix-Rabault, Vanessa Kreusch, Christèle Maizonniaux, Manny Rayner, Sophie Rendina, Emily Ryan-Cooper, Vladyslav Sukhyi, Ivana Vargova, Sarah Wright, Chunlin Yao, and Rina Zviell-Girshin (2025). *AI-Generated Tailor-Made Pedagogical Picture Books: How Close Are We?* *Education Sciences*, 15(12), article 1704. doi:10.3390/educsci15121704. ResearchGate copy.
- [3] ChatGPT C-LARA-Instance, Belinda Chiera, Anne-Laure Dotte, Stéphanie Geneix-Rabault, Christèle Maizonniaux, Sophie Rendina, Manny Rayner, Fabrice Wacalie, and Pauline Welby (2025). *Building culturally relevant multimedia resources for Indigenous languages using AI tools and collaborative development (extended abstract)*. Eighth Workshop on the Use of Computational Methods in the Study of Endangered Languages, Honolulu, Hawai‘i, March. Available on ResearchGate.
- [4] Junshu Pan, Panzhong Lu, Yixuan Weng, QiYao Sun, Fang Guo, Zijie Yang, Qiji Zhou, and Yue Zhang (2026). *Airaxiv: An AI-Driven Open-Access Platform for Human and AI Scientists*. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pp. 636–647. San Diego, California: Association for Computational Linguistics. doi:10.18653/v1/2026.acl-demo.63. Platform: <https://airaxiv.com/>.