Citizen-linguists and Decolonial Lexicography: Co-creative Dictionary-building in Grassroots Digital Language Documentation

Anna Luisa Daigneault
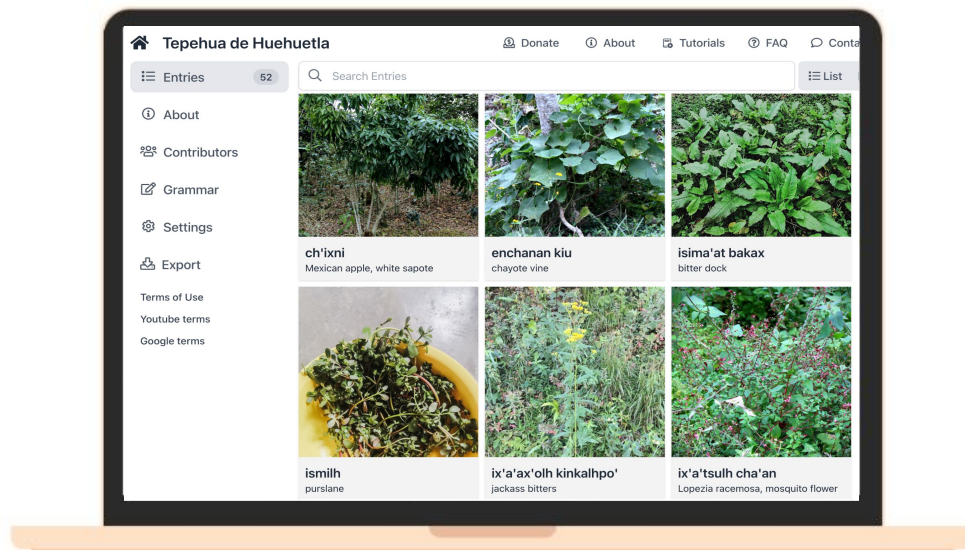annaluisa@livingtongues.org

Gregory D. S. Anderson
gdsa@livingtongues.org

ComputEL-8
March 5, 2025

LIVINGTONGUES
INSTITUTE FOR ENDANGERED LANGUAGES

Université
de Montréal

Living Tongues: A Grassroots Approach to Global Issues

Living Tongues Institute is a globally focused non-profit that supports grassroot efforts in language maintenance and engages in scientific linguistic research.

The **Living Dictionaries** web app allows for the free development of digital lexicographic resources with or without linguists.

Background

- State-run language programs often serve as vectors of total assimilation to dominant languages and the abandonment of heritage ones (Skutnabb-Kangas, 2000; 2023), grassroots digital projects can serve as a counterbalance and bring visibility to lesser- known languages.

- Access to high-quality digital resources is essential for language communities in the modern age, as information is increasingly consumed and disseminated digitally, specifically through mobile platforms.

- Assisting communities in developing such accessible resources is a tangible contribution by linguists in response to the colonialist underpinnings of linguistics.

Towards a decolonial lexicography

- Linguistics is rightly critiqued as rooted in colonialist projects.
  (Errington, 2001; Makoni, 2013; Zimmermann and Kellermeier- Rehbein (eds.), 2015; Hudley et al., (eds.) 2024).

- It is a moral imperative of the 21st century to address colonialist legacies and thus for linguists to make possible decentralized and decolonial approaches to language documentation, language resource development and linguistic analysis.

- It is not enough to simply acknowledge and critique the colonialist underpinnings of the field as these scholars largely have done, which, if not tied to action, remains empty, self-defeating rhetoric.

- Documenting languages is not only important to the scientific field of linguistics, but also to speech communities who are urgently looking for tools to combat language loss.

Towards a decolonial lexicography

- Relying on institutional actors – state, academic, juridical – to act in the interests of linguistic minority communities and to enforce linguistic human rights (not just on paper) has proven to be largely ineffective to date, except in very few contexts where governments have successfully helped revitalize languages that are typically the sole or main minority Indigenous language of the nation (e.g., in Wales, Ireland, New Zealand, Hawai'i).

- As such, grassroots efforts that combine technology, collaboration and community stewardship provide a meaningful path to combat language loss.

- This presentation shows the global applicability of the ***Living Dictionaries*** platform, which was engineered for straightforward co-creation of lexicographic materials between grassroots collaborators, to create accessible digital resources for all underrepresented and endangered languages.

The Living Dictionaries team

Clockwise from top left: Diego Córdova Nieto (México), Anna Luisa Daigneault (Canada), Jacob Bowdoin (Taiwan / USA) and Gregory D. S. Anderson (USA)

How the Living Dictionary platform works

💲 Donate ⓘ About 📖 Tutorials ❓ FAQ 💬 Cont

☰ Entries **714**

🔍 Search Entries ☰ List

Entries: 1-20 / 714 (2ms) **Reset local caches (admin)**

ⓘ About

👥 Contributors

✏️ Grammar

⤵ Import

⚙️ Settings

⤓ Export

Terms of Use

Youtube terms

Google terms

📢 Listen **"Xánteyy', shĕg ts'udêel," stsųų nįh.**
English: Grandma says, "let's hurry to get there". — Kinship and relationships — 🎥 ☁Photo

📢 Listen **Ậą, ałts'ęy gha k'éxdíhtiił.**
English: Yes, I can count to five. — Quantities and Quantifiers — 🎥 ☁Photo

📢 Listen **Ậą, ch'ehtlé' éedlah.**
English: Yes, she has pants. — Clothing and adornment — 🎥 ☁Photo

📢 Listen **Ậą, ch'ekent'ăag' k'et nat-tl'êdz éł ch'etnih'ées.**
English: Yes, I can use beads on the moccasin top. — Clothing and adornment — 🎥 ☁Photo

📢 Listen **Ậą, ch'óndéyh dhíhdlah.**
English: Yes, I have medicine. — Health, well-being and sickness — 🎥 ☁Photo

📢 Listen **Ậą, chit-tsez eltsiin wuk'éetáa ch'ĕl'tsîig dhihdlah.**
English: Yes, I have different patterns. — 🎥 ☁Photo

📢 Listen **Ậą, dąą'eek eedlah.**
English: Yes, she has a fur coat. — Clothing and adornment — 🎥 ☁Photo

☰ Entries **1,553**

🔍 Search Entries

☰ List   ⊞ Table   🖼 Gallery   🖨 Print

ⓘ About

🅰 Grammar

👥 Contributors

🕐 History

⚙ Settings

📥 Import

📤 Export

Terms of Use

Youtube terms

Google terms

🎤 **àkwàn** [àkʷǎ]
n, English: bridge
Water, bodies of water and the water cycle
🎥 ☁ Photo

🎤 **àlàgó** [àlàgó]
n, English: skirt
Clothing and adornment
🎥 ☁ Photo

🎤 **alẹ́** [alɛ́]
n, English: base, earth, field, ground, land, plot
🎥 ☁ Photo

🎤 **álẹ́** [álɛ́]
n, English: night
Universe and the natural world
🎥 ☁ Photo

🎤 **alẹ̀ orie** [alɛ̀ orie]
adj, English: plain
Universe and the natural world   Earth, geology and landscape
🎥 ☁ Photo

🎤 **álẹ́rungbẹ̀** [álɛ́rũgbɛ̀]
n, English: orchard
Universe and the natural world   Plants, trees and other vegetation
🎥 ☁ Photo

🎤 **àlùbásà** [àlùbásà]
n, English: onion
Food, cooking and foodways   Plants, trees and other vegetation
🎥

« ‹ 1 2 3 4 **5** 6 7 8 9 10 11 12 › »   **+ Add Entry**

Filters   **View Entries**

TYPO TOLERANCE

0   1   2   3   4   5

Search Part of Speech

☐ noun (865)
☐ preposition (14)
☐ adverb (27)
☐ adjective (190)
☐ verb (412)

⌄ Show More

Search Semantic Domain

☐ Universe and the natural world (60)
☐ Humans, Social Relations and Organization (110)
☐ Language and Communication (9)
☐ Daily life (9)

捐赠　关于　教程　FAQ　联系我们　语言

词条　296

搜索词条

名单　表格　展览表　打印

词条: 1-20 / 296 (2ms)

Reset local caches (admin)

过滤　　　查看词条

词条

关于

语法

贡献者

历史

设置

进口

出口

使用条款

YouTube 条款

谷歌条款

听　一 /id1/
名, 英语: one, 普通话语: 一, 粤语: 一　　数字　　照片

听　七 /cid1/
名, 英语: seven, 普通话语: 七, 粤语: 七　　数字　　照片

听　七十 /cid1sib5/
名, 英语: seventy, 普通话语: 七十, 粤语: 七十　　数字　　照片

听　七百 /cid1bag1/
名, 英语: seven hundred, 普通话语: 七百, 粤语: 七百　　数字　　照片

听　丈人婆 /chong44njin11po11/
名, 英语: mother-in-law, 普通话语: 岳母, 粤语: 岳母　　人类，社会关系和组织　亲属和关系　照片

听　丈人老 /chong44njin11lao31/
名, 英语: father-in-law, 普通话语: 岳父, 粤语: 岳父　　人类，社会关系和组织　亲属和关系　照片

听　丈夫；老公 /cong53fu44/ /lau31gung44/
名, 英语: husband, 普通话语: 丈夫, 粤语: 丈夫　　人类，社会关系和组织　亲属和关系　照片

TYPO TOLERANCE

0　1　2　3　4　5

词性

☐ 名词 (291)

☐ 副词 (4)

☐ 动词 (4)

搜索 语义域

☐ 宇宙和自然世界 (3)

☐ 人类，社会关系和组织 (45)

☐ 语言和交流 (4)

☐ 日常生活 (13)

☐ 数量和量词 (4)

⌄ 显示更多

扬声器

☐ Ying (277)

☐ 没有 图片 (296)

☐ 有 音频 (277)

☐ 没有 音频 (19)

1　2　3　4　5　6　7　8　9　10　11　12　13　14　15　›　»　✚ 添加词条

bir saɽu aɽaʔ

Devanagari Script

बिर साडु अड़ाः

Phonetic

[bir saɽu aɽaʔ]

Hindi: Translation

अरबी

Semantic Domain

Plants, trees and other vegetation
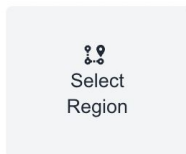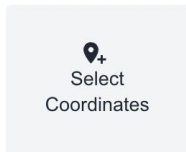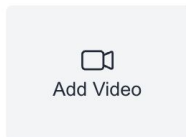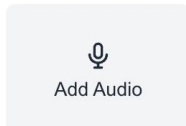
Health, well-being and sickness

Scientific Name
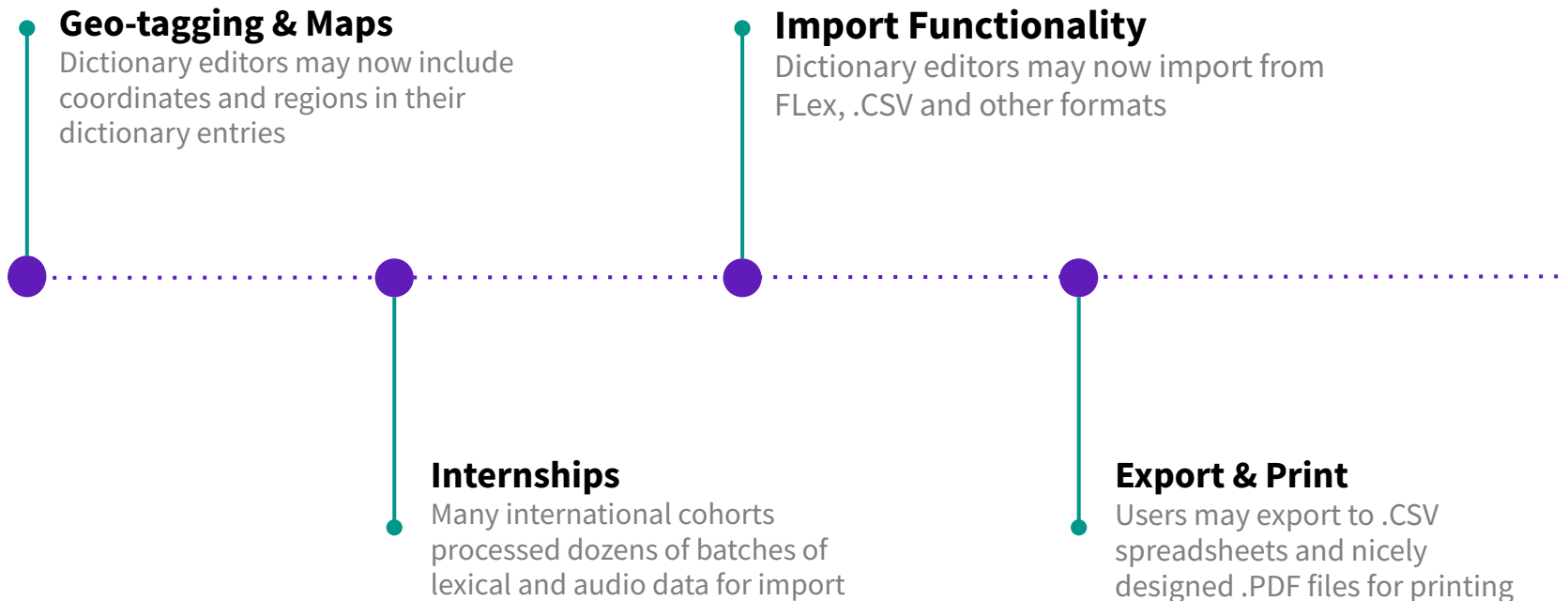
Colocasia esculenta (L.) Schott

Notes

The leaf of **bir saru aɽaʔ** is a nutritious leafy green that resembles spinach. While the leaves can be poisonous when eaten raw (*berel*), cooked leaves can be a versatile and nutritious addition to one's diet.

🎤
Add Audio

📹
Add Video

📍
Select
Coordinates

Select
Region

A sample entry from the
Birhor Living Dictionary

https://livingdictionaries.app/birhor

Living Dictionaries: Some Recent Innovations (2023-2024)

Geo-tagging & Maps
Dictionary editors may now include coordinates and regions in their dictionary entries

Import Functionality
Dictionary editors may now import from FLex, .CSV and other formats

Internships
Many international cohorts processed dozens of batches of lexical and audio data for import

Export & Print
Users may export to .CSV spreadsheets and nicely designed .PDF files for printing

Geo-Tagging in Entries

← Back   Delete 🗑

Dominica, Wai'tubuli, Oüàitoucoubouli

🎤
Add Audio

English: Translation

"Tail is Her Body" (one of the Windward Islands, among the Lesser Antilles)

☁
Upload Photo

Semantic Domain

Location and Directions   Place names

🎥
Add Video

Phonetic

✏

French: Translation

✏

Spanish: Translation

✏

Internship Cohorts
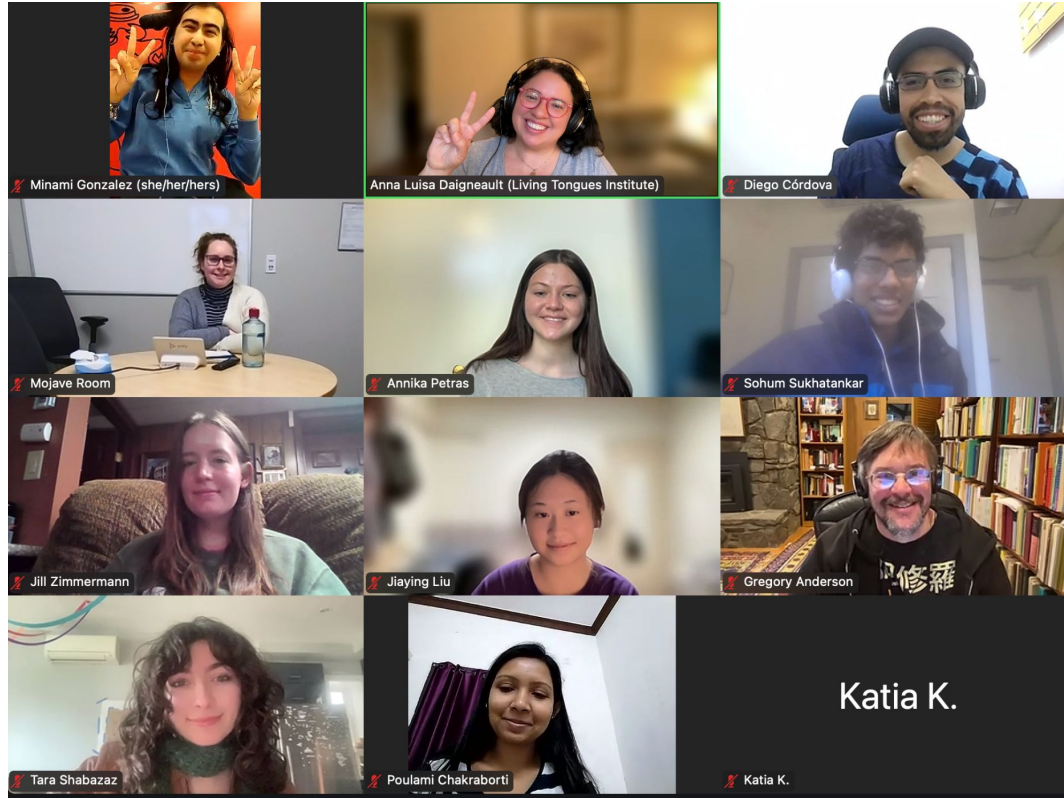
	lexeme	localOrthography	en_gloss	cmn_gloss	yue_gloss
1	**lexeme**	**localOrthography**	**en_gloss**	**cmn_gloss**	**yue_gloss**
2	*word/phrase in Hakka*	*alternative spelling or script (optional)*	English	Mandarin	Cantonese
3	常常	song11song11	often	常常	常常
4	经常	gin44shong11	always	经常; 总是	经常; 總是
5	准备	zhun31pi53	ready	准备好	準備好
6	停	tin11	cease, stop	停止，停止	停止，停止
7	结束	ged1chug1	finish (vb)	结束	結束
8	结束；收庄	ged1chug1; shiu33zong33	end (temporal) (noun)	结束;收尾	結束;收尾
9	持续	ci11sug5	last, endure	持续	持續
10	开始	hoi44ci31	begin, beginning	开始	開始
11	延迟	ien11ci11	retard, delay	延缓，延迟	延緩，延遲
12	匆忙	cung44mong11	hasten, hurry	匆忙	匆忙
13	慢	man53	slow	慢的	慢的
14	快	kai53	fast, quick	快	快
15	立即；马上；即刻	lib5zid1; ma44song53; zid1ked1	immediately	立即地	立即地
16	今下	gim44ha53	now	现在; 如今	現在; 如今
17	迟	cii11	late	晚的; 迟的	晚的; 遲的
18	朝 （副词）	zau44	early	早早地	早早地
19	朝 （名词）	zau44	breakfast	早餐	早餐
20	朝晨頭	zau44 sin11tiu11	morning	上午，早上	上午，早上
21	老；旧	lau31; kiu53	old	老的; 旧的	老的; 舊的
22	年轻;后生	njen11kiang44; hiu53sen44	young	年轻的	年輕的; 後生的
23	年龄	njen11lin11	age	年龄	年齡
24	新	sin44	new	新的	新的
25	时间	si11gan44	time	时间	時間
26	三次	sam44cii53	three times	三次	三次
27	第三	ti53sam44	third	第三	第三

PDF Export for Printing

💲 Donate   ⓘ About   🗐 Tutorials   ❓ FAQ   💬 Contact Us   🔤 Language

☰ Entries   **1,819**

🔍 Search Entries

☰ List   ▦ Table   🖼 Gallery   🖨 Print

ⓘ About

Entries: 1819 (2ms)

👥 Contributors

🖨 **Print**   Max entries [52]   Columns [2]   Headword size (pt) [17]

✏️ Grammar

Font size (pt) [14]   Images: [50] % ☑ Translation ☑ Phonetic

📥 Import

☑ Part of Speech   ☑ Plural form   ☑ Notes   ☐ Show Labels   ☐ Show QR Code

⚙️ Settings

bamitu [bâmítú] *n.* nape of neck, nuque bamtɨtutu
Lexeme collected on: 23/Feb/2005
Tones: HL H] H
IPA transcription of plural form: bamtɨtutu

📤 Export

banakwàŋà [bánákwàŋà] *n.* bread, pain
Lexeme collected on: 02/Dec/2004
Tones: H H L L

Terms of Use

bannà [bánnà] *n.* enemy, ennemi bannàsù
Lexeme collected on: 23/Feb/2005
Tones: H] L
IPA transcription of plural form: banna-su

Youtube terms

Google terms

batàwan [bátàwán] *n.* baby sling, écharpe (pour porter bébé) bachuwansu

bufì [búfì] *n.* request (n), demande, requête
Lexeme collected on: 25/Jan/2005
Tones: H] L

bukəmòla [búkəmͻlá] *n.* spring, source bukumolu
Lexeme collected on: 08/Dec/2004
Tones: H H L H
IPA transcription of plural form: bukumͻlu

bula mògh [búlá mͻɣ] *n.* nine -9, neuf -9
Lexeme collected on: 15/Nov/2004
Tones: H H LH

buma [búmâ] *n.* baobab tree, baobab bumasù
Lexeme collected on: 23/Feb/2005
Tones: H HL

« ‹ 1 2 3 4 5 6 7 › »

Filters

PART OF SPEECH
☐ noun (1068)
☐ verb (728)
☐ adjective (6)
☐ pronoun (6)
☐ numeral (6)

SPEAKER
☐ Ntang Julius (1)

☐ Has Audio (1)
☐ Has Speaker (1)
☐ Has Noun Class (538)
☐ Has Plural form (756)
☐ Has Part of Speech (1809)

Search by 🔍 algolia

Ndokobaï, et al., 2023. Bum Living Dictionary. Living Tongues Institu

Living Dictionaries: 2025 Roadmap

Offline Mode
We will test this functionality with dictionary teams in low-bandwidth regions of the world

Imports from IDS
We are working on 100+ batches of data from *The Intercontinental Dictionary Series*

Multimedia Files
Editors will soon be able to add multiple audios, images and other files to their entries

Summer Internship (July 7-31, 2025)
Providing an opportunity for students and emerging linguists to work on Living Dictionaries

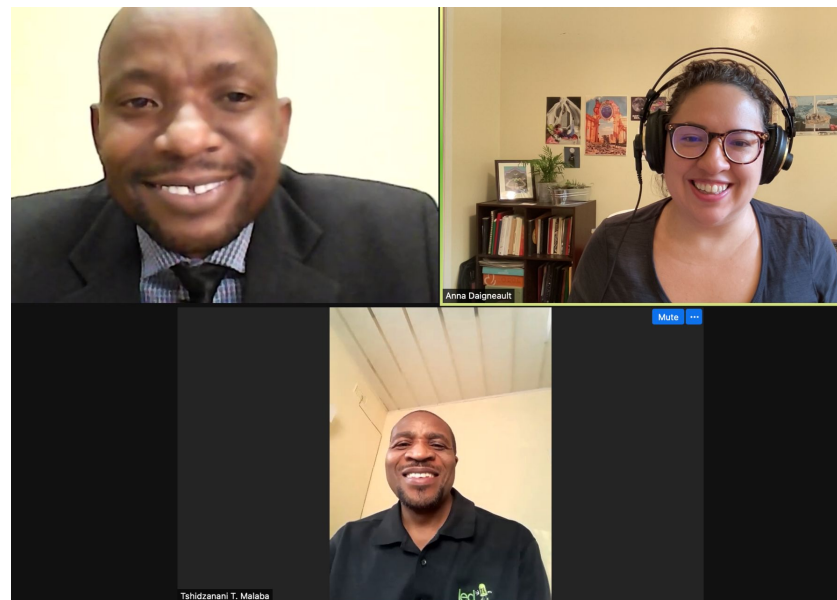# Remote and in-person collaboration

Garifuna Living Dictionary team in California, USA

Remote and in-person collaboration

Jewish Languages Consortium (for endangered Jewish languages)

Kalanga Living Dictionary team (Zimbabwe - USA)

Building the Teŋukan Living Dictionary

01

Planning and Consultation

Living Tongues linguists and the lead Teŋukan community researcher will have a series of Zoom calls (in English) discussing the overall project goals and schedule of work. Local travel concerns are addressed.

02

Digital Training in Bamako, Mali

The local community researchers will travel to the city of Bamako and receive remote digital training in French from the linguists and lead digital curator, via Zoom.

03

Data Collection & Programming

The Teŋukan team undertakes lexical data collection, multimedia recording, editing, glossing and curation in Mali. Associated GPS coordinates will be collected as well. Developers code new mapping features, and test Offline Mode.

04

Assessment

Through systematic review with Teŋukan speakers, the team will review all the data for consistency, and the web developers will address any issues involved with geo-mapping and Offline Mode.

05

Import and Deployment

The lead digital curator imports batches of data that are ready for public viewing. Web developers roll out new digital features such as the Teŋukan Living Dictionary Map. The project is shared with audiences in Mali and in the diaspora.

Citizen Science and Digital Lexicography

- It is not an overstatement to assert that language documentation is crucial to conserving humanity's intangible heritage on Earth.

- *Living Dictionaries* offer a central point where different communities of stakeholders can contribute equally and respectfully.

- The platform was created to make building linguistic resources from the ground up easy and accessible to anyone who knows how to operate a smartphone or tablet.

- Equitable resource development is at the heart of this work.

- As platform administrators, we have sought to lessen the burden of colonial and capitalistic frameworks (such as bureaucracy and subscription models) so there are as few limitations as possible for newcomers to lexicography.
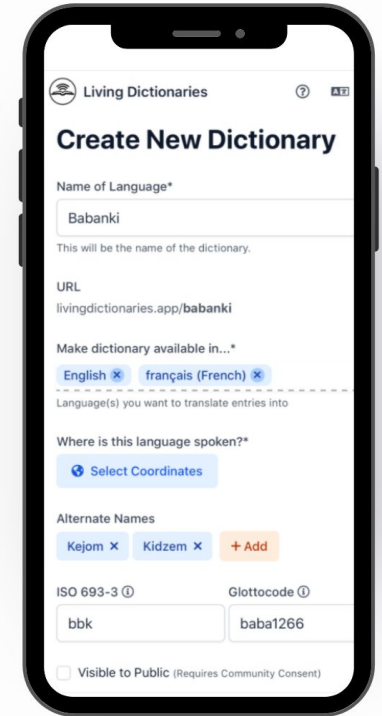
Citizen Science and Digital Lexicography

- *Citizen-linguists* play a huge role in the process of building these digital dictionaries. Citizen-linguist is here understood as a person who is actively engaged in their speech community, believes in safeguarding their native (or heritage) language and works towards transmitting it to future generations.

- Citizen-linguists are motivated to create accessible language resources that are tailored to their community's needs.

- They are people who fulfill the multi-faceted roles of documentarian, language activist and digital content creator, whether they have formal training in linguistics or not.

- Some are educators, some are students; some are people with advanced training in other academic fields who see the value in protecting their language, whether they are fluent speakers or not.

Citizen Science and Digital Lexicography

- Citizen-linguists take on the difficult challenge of recording and sharing their language, which may be under-studied by scholars and almost assuredly under-valued by their own community and the wider public.

- Many citizen-linguists are leaders in other areas of community life and undertake language work as volunteers in their spare time, because they understand that the act of preserving their language is connected to their cultural group's well-being, identity, and survival.

- Like other citizen scientists who study and celebrate local phenomena (e.g., flora or fauna, etc.), citizen- linguists are grassroots actors who may bridge divides between diverse groups of people and help bring local knowledge forward, while also feeling a sense of personal pride and connection to the language in question.

- Citizen-linguists are often excellent (co-) creators of language materials because they see the long-term value of the work they are doing and know that teamwork can benefit large, detailed undertakings like language documentation.

Werikyana Citizen Science and Digital Lexicography

The creation of the first-ever Werikyana Living Dictionary in Brazil (photos by Spike Gildea used with permission)

☰ Entries   **3,299**

🔍 Search Entries   ☰ List

ⓘ About

🔤 Grammar

👥 Contributors

🕘 History

⚙ Settings

⇥ Import

⇤ Export

Terms of Use

Youtube terms

Google terms

Entries: 121-140 / 292 (2ms)   Reset local caches (admin)

kuwa
Tiriyó: koi

kuwama
Tiriyó: pakoro

kuwanamare
Tiriyó: kumuimë

Kuwaruru
Tiriyó: Kurunpapikë

kuwatxi
Tiriyó: kunyarekena

kuwayakpitxoro
Tiriyó: Tonoro

kuyuwi
Tiriyó: kuwi

ma'maya
Tiriyó: mapaya

« ‹ 1 2 3 4 5 6 **7** 8 9 10 11 12 13 14 › »

+ Add Entry

Summary

- The website offers the insights and systematicity of linguistic science in a user-friendly context, free of cost to citizen-linguists, with no institutional or other administrative roadblocks preventing access to this online tool.

- It highlights and preserves essential ecological, social, and linguistic knowledge that lie at the foundation of cultural survival.

- All intellectual property rights associated with dictionary data remain with the language community from which the data originates. Our code is also accessible.

- The platform supports citizen science by providing STEM opportunities to community activists to document their own languages and help them gain access to technical guidance from our team of professional linguists, anthropologists and web developers.

Thanks for watching!

Anna Luisa Daigneault
annaluisa@livingtongues.org

Gregory D. S. Anderson
gdsa@livingtongues.org

References

Bialystok, E., Craik, F. I. M., and Freedman, M. 2007. Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, 45(2):459–464. https://n.neurology.org/content/81/22/1938

Errington, J. 2001. Colonial linguistics. *Annual Review of Anthropology*, 30(1):19–39.

Hudley, A. H. C., Mallinson, C., and Bucholtz, M., eds. 2024. *Decolonizing linguistics*. Oxford University Press.

Makoni, S. B. 2013. An integrationist perspective on colonial linguistics. *Language Sciences*, 35:87– 96.

Olko, J., Lubiewska, K., Maryniak, J., Haimovich, G., de la Cruz, E., Cuahutle Bautista, B., Dexter- Sobkowiak, E., and Iglesias Tepec, H. 2022. The positive relationship between Indigenous language use and community-based well-being in four Nahua ethnic groups in Mexico. *Cultural Diversity and Ethnic Minority Psychology*, 28(1):132–143. https://doi.org/10.1037/cdp0000479.

Olko, J., and Andrason, A., eds. 2023. Introduction: Heritage languages and the well-being of speakers. In *Heritage languages and the well-being of speakers*, pages 5–16. Linguapax. https://www.linguapax.org/uploads/2024/01/Lingu apax-2023-baixa.pdf.

Perani, D., and Abutalebi, J. 2015. Bilingualism, dementia, cognitive and neural reserve. *Current Opinion in Neurology*, 28(6):618–625. https://journals.lww.com/co- neurology/Abstract/2015/12000/Bilingualism,_dem entia,_cognitive_and_neural.12.aspx.

Skutnabb-Kangas, T. 2000. *Linguistic genocide in education or worldwide diversity and human rights*. Mahwah, NJ/London: Lawrence Erlbaum Associates.

Skutnabb-Kangas, T. 2023. Preventing the implementation of linguistic human rights in education. In T. Skutnabb-Kangas and R. Phillipson, eds., *The handbook of linguistic human rights*, pages 109–126. Hoboken, NJ: Wiley- Blackwell.

van Beek, S. 2016. Intersections: Indigenous language, health and wellness. *First Peoples' Cultural Council*.

Zimmermann, K., and Kellermeier-Rehbein, B., eds. 2015. *Colonialism and missionary linguistics*. Colonial and Postcolonial Linguistics Vol. 5. Walter de Gruyter GmbH & Co KG.

Zuckerman, G. 2020. Our ancestors are happy. Language revival and mental health. In G. Zuckerman, ed., *Revivalistics: From the genesis of Israeli to language reclamation in Australia and beyond*, pages 9–36. Oxford University Press. https://doi.org/10.1093/oso/9780199812776.003.00 09.