

Living Dictionaries: An Electronic Lexicography Tool for Community Activists

Gregory D. S. Anderson¹, Anna Luisa Daigneault¹

¹Living Tongues Institute for Endangered Languages, 4676 Commercial St SE, #454.
Salem, OR 97302, USA

E-mail: gdsa@livingtongues.org, annaluisa@livingtongues.org

Abstract

Living Dictionaries are comprehensive, free online technological tools integrating audio, images and other multimedia that can assist endangered and other language communities, providing a simple way to create high-quality multilingual documentation records. The platform is a progressive web application functioning within any Internet browser on any computer or mobile device, Android or iOS. If needed, Living Dictionaries can be created, managed and edited using only smartphones or tablets, which can function as complete workstations for recording and entering linguistic data and other multimedia. Living Dictionaries may be public or private and may include written entries with translations and example sentences in multiple languages and scripts, audiovisual files, parts of speech and semantic domains, morphosyntactic linguistic analysis and be tagged with other metadata. The platform is free because for almost all minority language communities the costs related to producing high-quality linguistic materials can be insurmountable. A moral imperative of the 21st century is the decolonisation and democratisation of linguistic resources. Online dictionaries should reflect the user communities, tailored to suit their needs as well as curated by citizen-linguists. Community resources have greater uptake and engagement by communities if they take a primary role in developing them.

Keywords: dictionary; language technology; endangered languages; lexicography; web application

1. Introduction

Technology can be “disruptive” because it can forever change the way people operate in their daily lives. But what if technology could also “disrupt” language bias and privilege? What if access to certain language technologies could help challenge language hierarchies and give endangered languages a fighting chance of survival? With over 3,000 languages in danger of being lost before the end of the century, we know there is a need to act quickly. Living Dictionaries¹ address the urgent need to provide comprehensive, free online technological tools that can assist endangered language communities simultaneously in conservation efforts and revitalisation programs by providing a simple way to create high-quality language documentation records. The Living Dictionaries platform can accommodate everyone from seasoned field linguists to emerging language activists in developing countries. The platform is free to use, and

¹ Available worldwide online at <https://livingdictionaries.app/>

the intended target audience of this web app is inclusive, diverse and multilingual.

2. The Advantages of Creating Digital Dictionaries in the 21st Century

The advantages of online dictionaries have been well-known for some time. Dmitrova et al. (2009: 77) discussed such features as their wide accessibility, the possibility for them to be continuously updated as well as corrected and edited, or the creation of an online community of real-time users in multiple different locations, which can lead to real-time editing and updating of the dictionary. Lew and de Schryver (2014: 345) concur that “[o]nline dictionaries as well as dictionary apps can be updated as often as needed, and all users can instantly benefit from the improved content or features right from the moment these become available.” Dmitrova et al. (2009: 77) also commented on a key feature of online digital dictionaries: no restrictions on the size. Indeed, the old dictionary-making paradigm was dependent on printing restrictions, content limits, page layouts, alphabetisation and other ‘corporate’ concerns, where when updating dictionaries “the editors usually had to grapple with the dilemma of what to sacrifice in order to make space for the new items,” (Lew and de Schryver, 2014: 345). Today, these types of bottom-line concerns are largely irrelevant, and powerful search functionality and the relatively low cost of database storage has obviated the challenges of the past. As Lew and de Schryver (ibid.) put it: “[t]he digital revolution has changed that, and now items are in fact very rarely removed when digital dictionaries are updated.” Other innovative advantages of electronic dictionaries include: “the option to hear new words being pronounced, being able to copy over foreign scripts one would be hard pressed to type in, the interconnectivity with other resources (such as corpora), and the fact that one stays within the same (digital) medium, rather than having to move back and forth between the screen and a book on one’s desk” (Lew & de Schryver, 2014: 347). Furthermore, we now benefit from the possibility of integrating large numbers of photos and other audiovisual multimedia, the ability to accommodate sign as well as oral languages, and perhaps most importantly, the capacity to address the vast gap in digital resource availability that disproportionately impacts minority communities worldwide. A multimedia online dictionary platform such as Living Dictionaries accommodates the needs of twenty-first century users of such tools by using the latest technologies to produce tools that in the long run can become encyclopaedic in nature.

3. The Impacts of Colonisation on Under-Represented Languages

Colonialism has had a deep impact on most countries of the world. The legal and social status of minority and under-represented languages, as well as the resources that support them, are characterised by unequal distribution and injustice in almost every polity across the globe. The linguistic consequences of colonialism entail in some cases the nearly complete elimination of most of the original languages spoken on a conquered

territory, and the nearly complete domination of the colonial language, e.g., Russian in Siberia (Anderson, 2017), English in the US and Australia, and Spanish in most countries in Central and South America. In other cases, this means the enfranchisement of a group who acquired power within the colonial structure and have held it in the postcolonial period, and who in a similar neocolonial hegemonic manner promote their language as a national one over others also spoken in the country, (e.g., Setswana in Botswana, Burmese in Myanmar or Hindi in India) or regionally within a section of the country, e.g., Hausa in northern Nigeria. In some countries, constructed national languages have been vigorously promoted at the cost of others in the country, e.g., in Indonesia or Philippines, which have rebranded de-ethnicised versions of languages of the just mentioned type as national ones, whether a neocolonialist hegemonic language (Filipino) or a former urban/trade lingua franca (colloquial Malay > Bahasa Indonesia). In Melanesia, colonial-era contact languages were adopted as national ones and are promoted at the expense of others, leading to a decline in linguistic diversity over time. With very few exceptions, most nation-states favour a single language of one of these types over all others spoken in their territory. This institutionalised disenfranchisement has resulted in half of the world's languages presently undergoing an active shift towards dominant languages, and another 40% or so being threatened in such a way that this process will likely begin soon.

The main reason dominant language groups use to justify continued disenfranchisement of the minority languages of their countries is that it is too costly to support all languages. They also believe a subtractive language policy is the best means for ensuring a kind of national sense of self and to maintain territorial integrity. Both reasons are false. The latter belief is rooted in a continuation and naturalization of European Romantic/Herderian notions creating an ideal of one nation, one people, one language. With regards to the financial impacts of multilingualism, the actual costs of maintaining language diversity have been shown to be not nearly as high as imagined (Grin, 2003). The mindset regarding linguistic diversity thus needs to evolve: diverse languages need to be seen as resources that empower nations and not weaken them.

While for nations the financial cost of supporting multilingualism is not preventative in the way typically imagined, for almost all minority language communities the costs related to producing high-quality linguistic materials can be insurmountable. As activists in the field of endangered language documentation globally, we know this is to be true. Thus, we have created a state-of-the-art dictionary-builder that we have made available free of charge to all users. Through the Living Dictionaries platform, the Living Tongues Institute has approached solutions to the massive global language extinction crisis by attempting to obviate institutionalised barriers that prevent equal status and equitable treatment of all forms of linguistic communication. Training local people to conduct language documentation and revitalisation work and build dictionaries for their own communities is a core, long-term aspect of our approach.

A moral imperative of the 21st century is the decolonisation and democratisation of

linguistic resources, as colonised peoples have often been forcibly resettled, assimilated and disenfranchised from their own heritage. Indeed, it can be almost impossible for marginalised people in some parts of the world to even access documented knowledge about their languages. Prinsloo (2019: 218), citing CCURL 2014, succinctly summarises one of the realities facing many minority language communities as follows: “[u]nder-resourced languages suffer from a chronic lack of available resources (human-, financial-, time- and data-wise), and of the fragmentation of efforts in resource development. This often leads to small resources only usable for limited purposes [...] without much connection with other resources and initiatives.”

Now, through the accessibility of online digital media collections, scholars and activists have a great opportunity (and indeed a duty) to connect communities with the data they are entitled to. Under-represented languages need online resources to thrive in the digital era because people need to be able to easily store, reference and share content in their languages. To be sure, the Internet is a place where linguistic hierarchies in theory could be potentially upended, subverted and reinvented according to the needs of individuals and communities. Technologists and digital lexicographers must thus be publicly inclusive when it comes to minority languages and take a positive stance towards multilingualism. We advocate for an inclusive, citizen science approach to digital lexicography. Living Dictionaries address the obvious need to provide comprehensive, free access to robust technological resources. This platform provides an easy-to-use framework for systematically storing and sharing dictionary data in thousands of endangered languages, thus increasing their viability for survival in the long-term. This comes with significant implications: studies in North America and Australia show that language revitalisation leads to better mental health, better performance in schools, and expanded economic opportunity (Whalen et al., 2016).

4. Citizen Science: The Future of Lexicography

The very concept of a dictionary has changed in this new era. Lew and de Schryver observe (2014: 342),

“[a]s dictionaries moved from the bookshelves gradually onto [...] internet servers, and now mobile devices, they found themselves as it were in the same league as utility and productivity software, which in turn encouraged a more pragmatic and less ideological or dogmatic view of dictionaries. This trend was only strengthened as users themselves started getting involved in bottom-up dictionary-making.”

Online dictionaries can now reflect the user communities in a meaningful way, they can be tailored to suit their needs as well as curated by citizen-linguists who wish to build resources for their languages. No longer the exclusive domain of academic expert authorities and state-sanctioned language academies, digital dictionaries of the electronic era indeed belong to the realm of the collective intellectual property of language communities themselves. We strongly feel that for endangered and threatened

minority languages, the future of lexicography is crowd-sourced citizen science.² Community resources developed by community members are almost certain to have greater uptake and engagement by communities if they take a primary role in developing these resources themselves. Speaking about (South) Africa, Prinsloo (2019: 220) reminds us that “[w]hat is emphasised and encouraged today is the urge to compile dictionaries for African languages in Africa, by Africans, for Africans”, see also Prinsloo et al. (2017). This includes taking into consideration, among other things, that the complex grammatical structures of many African languages differ rather significantly from those of other major European and Asian languages (Van Wyk, 1995). During our online and in-person training workshops at the Living Tongues Institute for Endangered Languages, language activists who are facing rapid language loss have enthusiastically voiced their desire to create and maintain their own digital resources. We have created the Living Dictionaries platform with them in mind, optimising it for global remote collaboration, ease of use and accessibility on mobile devices, and we integrate community user feedback into the design and programming of the tool.

The Living Tongues Institute stands at the intersection of linguistics and activism, with the capacity to launch technological solutions that help aspiring language activists and scholars alike. Our team has adopted a vertically integrated approach to language documentation, in which local language consultants learn transferable digital and scientific research skills to eventually become research assistants, colleagues, and ambassadors for their languages. By facilitating in-person and online workshops during which we train local indigenous language activists to record and edit words and phrases in their native languages, we have developed a strong strategy that prioritises documentation as well as professional empowerment. 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, and it is also crucial to conserving humanity’s intangible heritage. It is up to our generation to use the tools of globalisation to empower those who have been disenfranchised. We consider this project a humanitarian mission that requires collaboration between scientists and local activists to make a difference. By pairing technology with our passion to document endangered languages, our platform is positioned to make a big impact on this field. The work we do is essential to help bolster the contemporary linguistic identity of the communities we serve and ensure a future for them. The materials and resources we create in collaboration with citizen-linguists will become the driving force that helps our descendants revitalise their languages in the future.

² Note that this does not mean that we advocate for the use of search engines to replace dictionaries, an alarm sounded among others by participants at Australex 2019, who fear it is becoming widely believed that dictionaries are no longer needed. Rather, we advocate for providing an easy to use, multimedia online digital dictionary tool that can create quality, multilingual (or monolingual) lexicographic resources for the widest possible range of languages worldwide.

5. Living Dictionaries: Set-up and Design Considerations

While much of our user community grapples with limited Internet connectivity and digital literacy, they regularly have access to smartphones and other mobile devices that can function as complete workstations for recording and entering linguistic data and other multimedia. Living Dictionaries are fully creatable, manageable and editable using mobile technology alone. The platform is a web-based application that functions within any Internet browser on any device, whether it is Android or iOS. The software works seamlessly across all mobile devices and tablets as well as desktop computers, and a service worker allows some features to be used offline in locations with limited Internet connectivity (more details on this below).

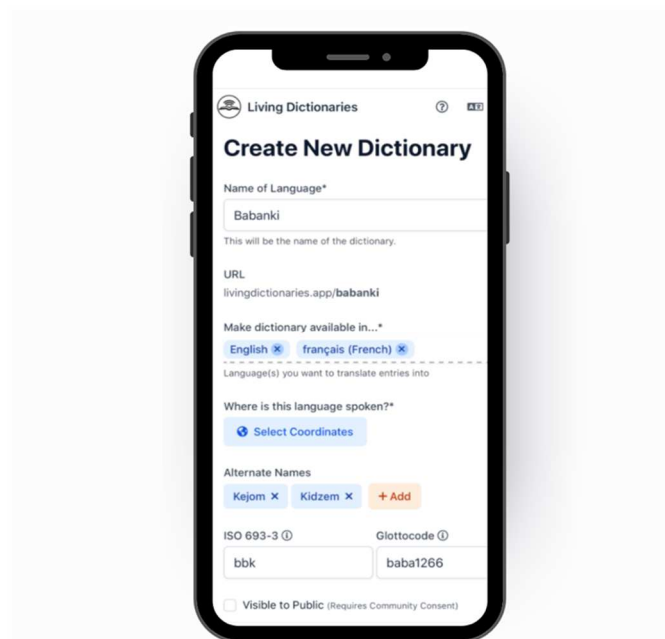


Figure 1: A mobile mock-up view of the creation of the Babanki Living Dictionary

Once a user registers for an account on the platform, they may create a new Living Dictionary right away, and become a manager of that dictionary. All of this, as well as the entry functions described below, can be done on mobile or desktop. Figure 1 shows the mobile view of the digital information required to create a new Living Dictionary for Babanki, a Grassfields Bantu language spoken by under 40,000 people in Cameroon (the depiction is based on how the process looks in a Chrome browser on an iPhone 6+). The dictionary creation process can take as little as a couple of minutes, or a bit longer if the dictionary manager needs to search online for the metadata relevant to their language project. We made the set-up process very user-friendly and fast so that activists can easily start their dictionary projects with as few bottlenecks as possible, and no institutional red tape. They do not have to go through the website administrators or through any type of approval process to get started.

Among the information requested to create a dictionary is the name of the language, a string of data which in turn automatically populates the ending of the URL of the new

dictionary. The name attributed to the dictionary itself can be modified by the manager at any time in the left sidebar “Settings” tab of their Living Dictionary. For example, communities may wish to modify the spelling or add an additional name in parentheses to the dictionary, to reflect contemporary ways of referring to the language. The URL, however, cannot be changed after it is established because it becomes hardcoded into the website.

Next, the dictionary manager is prompted to add glossing languages to the project. In the above example, since the Babanki language is spoken in Cameroon, English and French glossing languages are included here. This is done by choosing from a list of over 300 useful glossing languages that are worldwide in scope. We curated the list based on the dominant regional languages that users might need for their glosses. Then, geo-coordinates are requested under the prompt “Where is this language spoken?” to display the language on the Living Dictionaries homepage map. The manager may manually enter latitude and longitude coordinates or search our digital map (using an integrated MapBox plug-in) to drop a pin in the general area, or perhaps the exact village, where the language is spoken. This geo-location step is optional, and this data may be amended later by the platform administrators. We are currently working on the ability to drop multiple geo-pins as well as create polygons to better represent regions where languages are spoken, since many users have requested such options. User feedback and suggestions help drive our design process, and we value the input from dictionary managers on the platform.

After that, the dictionary manager may fill out “alternate names” for the language by typing them in one by one and hitting *enter* to lock them in. Many languages are known by multiple names in the linguistic literature and may also have various endonyms. We designed this naming aspect to be inclusive to all the possible naming conventions of the language, so there is no limit to how many alternate names one can list here in this step. They may also be typed in any script that is Unicode-compliant. All the “alternate names” will be used to tag the dictionary, which helps improve the search engine optimisation (SEO) of each Living Dictionary on the Internet, as well as assist people in searching for dictionaries on our homepage using any of the possible alternate names. The final steps in the Living Dictionary creation process include typing in the ISO 639-3 code and the Glottocode associated with the language. This also helps SEO, in case people are searching for online linguistic resources by one of those codes. Adding these codes is optional because 1) people may not know these codes or be aware that they even exist for their languages, and 2) some under-represented languages do not yet have these codes.

Lastly, the dictionary manager must decide whether the Living Dictionary will be “visible” to the public or not, by checkmarking a box indicating that they have community consent to put representations of this language online. The default setting for new dictionaries is “not visible to the public” which we consider to be a “private” mode. We designed it this way for various reasons: we want to be sure that the language

community has given their consent for the language being represented online, and we also want to give people the option of building their resources privately at their own pace before letting the rest of the world know that the Living Dictionary exists. It is important to note that a private Living Dictionary is not password-protected, but merely unlisted and not accessible to anyone who does not have the link. If made “visible” the Living Dictionary will be available for browsing on our public list of dictionaries on the platform’s homepage and will also be displayed on our map (if geo-coordinates are provided in the set-up process). The “visible to the public” option may be activated at any time using the “Settings” tab on the left sidebar. Many Living Dictionary managers populate their dictionaries privately with data, recordings and images and then switch the setting to “visible” when they are ready. At any time, whether the dictionary is set to private or public, a user may copy-paste the URL of the dictionary itself and share it with their friends, colleagues and relatives. Anyone who has been given the link can then view and browse it without having to type in a password or register for an account. Viewers cannot modify the Living Dictionary unless they are registered as a collaborator or manager of the project. Language communities own their own linguistic content on the platform. It is important to us that the intellectual property rights related to linguistic and cultural content remain in the hands of the native speakers and dictionary creators who work together to build the dictionaries on the platform. In terms of adding entries and multimedia to a Living Dictionary, this can be done on the platform by adding individual text entries and recording audio directly onto the platform. If the dictionary manager already has a large amount of text data in a .CSV, .PDF or .DOC file, they may request a batch import spreadsheet template from our team by using our “Batch Import Request” form found on the platform. It is also possible to merge two existing dictionaries once the data structure and any issues pertaining to orthography and duplicate content have been assessed by stakeholders and platform administrators.

Below is an individual lexeme entry page view from the San Sebastián del Monte Mixtec (Tò'on Ndà'vi) Living Dictionary, an indigenous language of Mexico. The possible fields to fill out in the data structure³ of the Living Dictionary are as follows: lexeme, English gloss, Spanish gloss, part of speech, phonetic transcription, semantic domain, morphology, interlinearisation, and an example sentence using the lexeme.

³ The data structure of Living Dictionary entries can be found here:
<https://gist.github.com/jwrunner/b8e658e3551f204225305d482f6743b2>

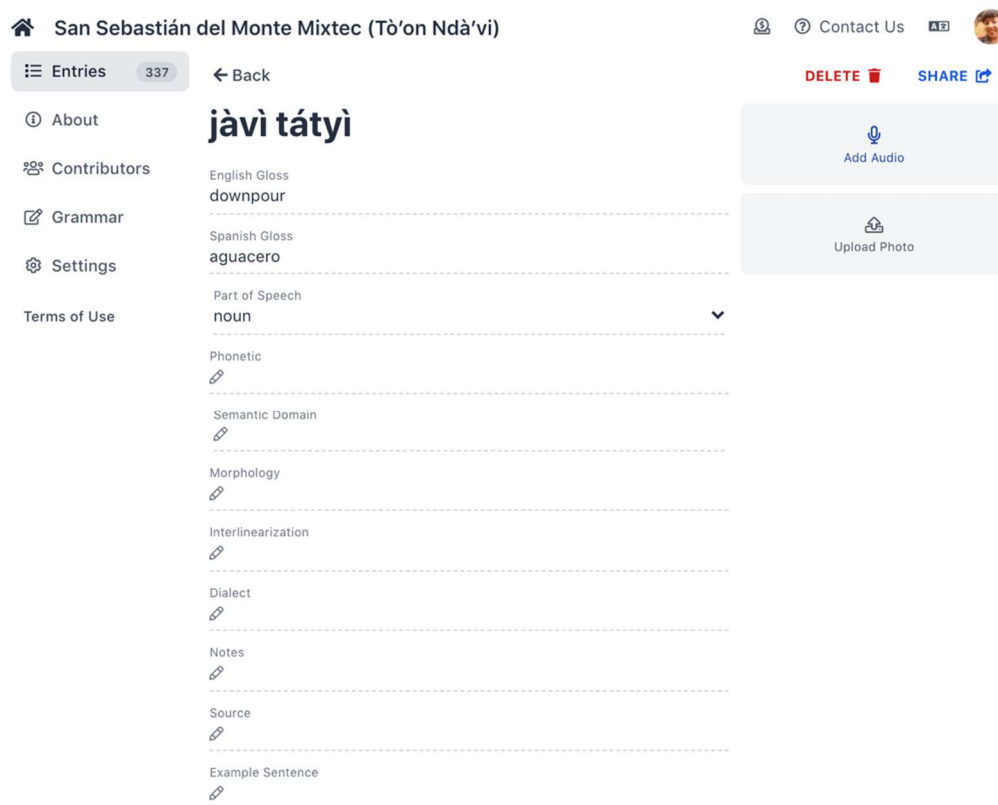


Figure 2: Lexeme entry page for the word for “downpour” in San Sebastián del Monte Mixtec

Source: <https://livingdictionaries.app/san-sebastian-del-monte-m/entries/list>

The Living Dictionaries platform is a “progressive web application” (PWA) that functions as a website and behaves like a mobile app on smartphones. PWAs do not require the user to download and install any software from the Internet. A Living Dictionary instead lives and caches data on the user’s device, and it also updates automatically from the Web. When launched from the user’s home screen, service workers enable a PWA to synchronise with the server and load text data instantly, regardless of the network state, so a user can be online or not. PWAs must be served from a secure origin and therefore live on HTTPS (and not http:). They are known to be secure, reliable and fast. Once a new digital dictionary has been created online, it can later be accessed and used offline, as well as modified. Text entries may be edited offline, and changes will automatically be uploaded to the cloud when the user is online. While one must currently be online to access and edit images and audio, plans are underway to make multimedia editing accessible offline in the future.

As Lew and de Schryver (2014: 342) aptly commented, “[m]odern dictionaries in the form of apps or online services are probably better seen as collections of structured data and code, rather than hardware.” This observation certainly applies to the Living Dictionary platform, which is programmed using HTML, CSS, Javascript and ReactJS with Svelte integration, and uses Google Firebase on the backend as a cloud-hosted

database. The language data, audio recordings and images are stored in the cloud. The code is currently stored on a private GitHub repository, with plans to make it open source in the coming years. The administrators have access to the backend from anywhere in the world. We partnered with the tech company Algolia to improve the platform's search engine capability on mobile and on desktop. The Algolia search integration allows users to search a Living Dictionary very efficiently, as well as use new filters that can search by categories such as part of speech, semantic domain, speaker name, or the presence of other kinds of tags. One can also use the powerful search bar (located in the centre right above the language data) to locate entries by lexeme, morpheme, part of speech, or semantic domain and other parameters. Search results are displayed alphabetically. It is important to note that users can easily search for any morphemes that are embedded inside lexemes. This is a very important search feature in polysynthetic languages such as Sora, where users may want to yield search results related to morphemes inside words, and alphabetical considerations are therefore inconvenient. As Figure 3 illustrates, searching for the morpheme 'd3um' (eat) in the Sora Living Dictionary yields a list of results that contains the 'd3um' inside of words and phrases, and not just at the beginning of an entry.

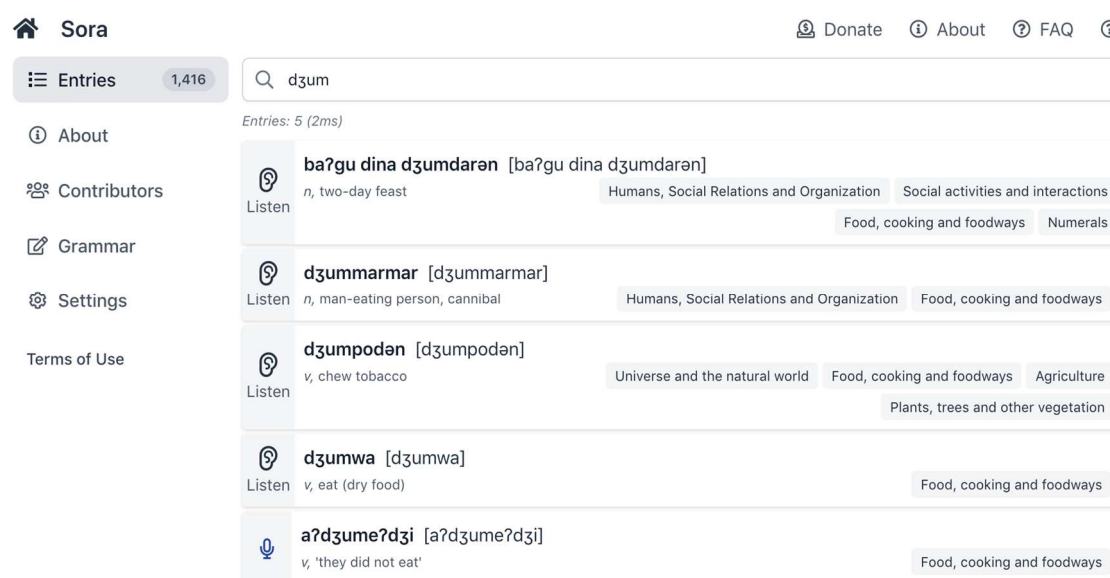


Figure 3: Search results for Sora morpheme 'd3um' (eat) in the Sora Living Dictionary

Source: <https://livingdictionaries.app/sora/entries/list>

The Living Dictionary website interface is currently available for use in English, Spanish, French, Portuguese, Hebrew, Russian, Bahasa Indonesia, Malay and KiSwahili, with Modern Standard Arabic, Tagalog, Zulu, Shona, Amharic, Hausa, Hindi, Assamese, Odia and Bengali interfaces coming online in 2021. A dictionary user can click on the top-right "Language" button to toggle between interface languages to display the website in the available languages (see Figures 4, 5 and 6). All functionality and features, including extensive dropdown menus for semantic domains and parts of speech are represented in the various interface languages. The website remembers the user's choice of language interface preference and automatically displays the website in

this language upon the user's return. At any point in navigating the web platform, the user may toggle between interface languages without having to leave the website at all. The platform also allows for nearly three hundred built-in glossing languages, covering most languages that function as a local, national or regional language of wider communication. The Living Dictionaries not only elevate threatened languages but allow for them to be explored in multilingual online environments, tailored for the usage needs of specific communities.

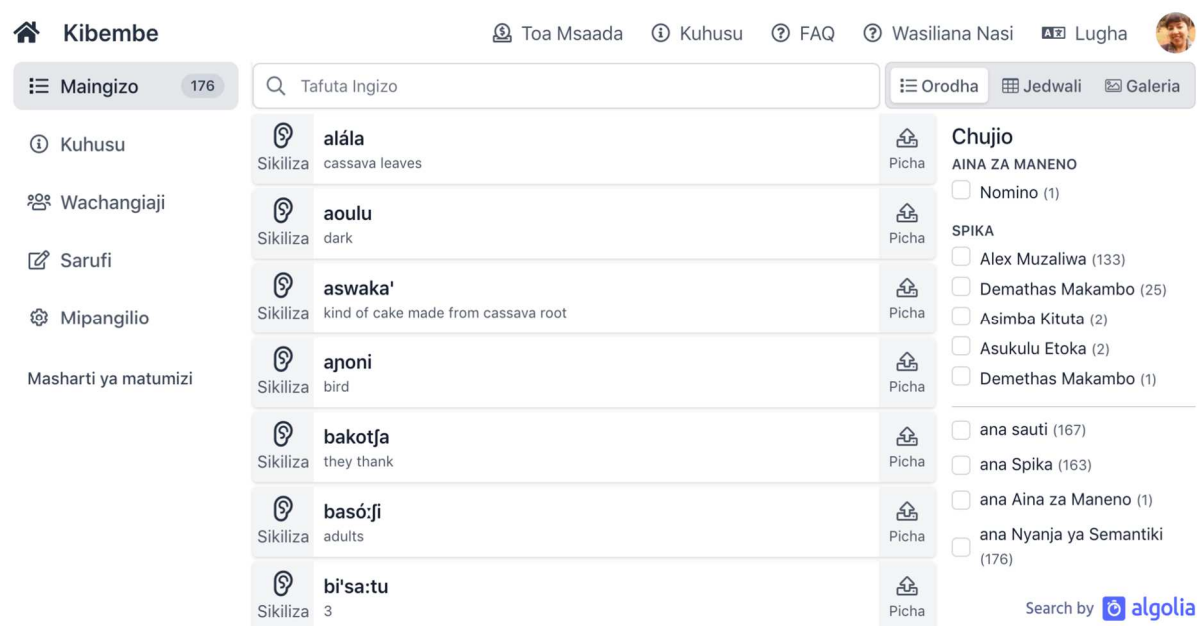


Figure 4: The Kibembe Living Dictionary displayed in the KiSwahili interface.

Source: <https://livingdictionaries.app/kibembe/entries/list>

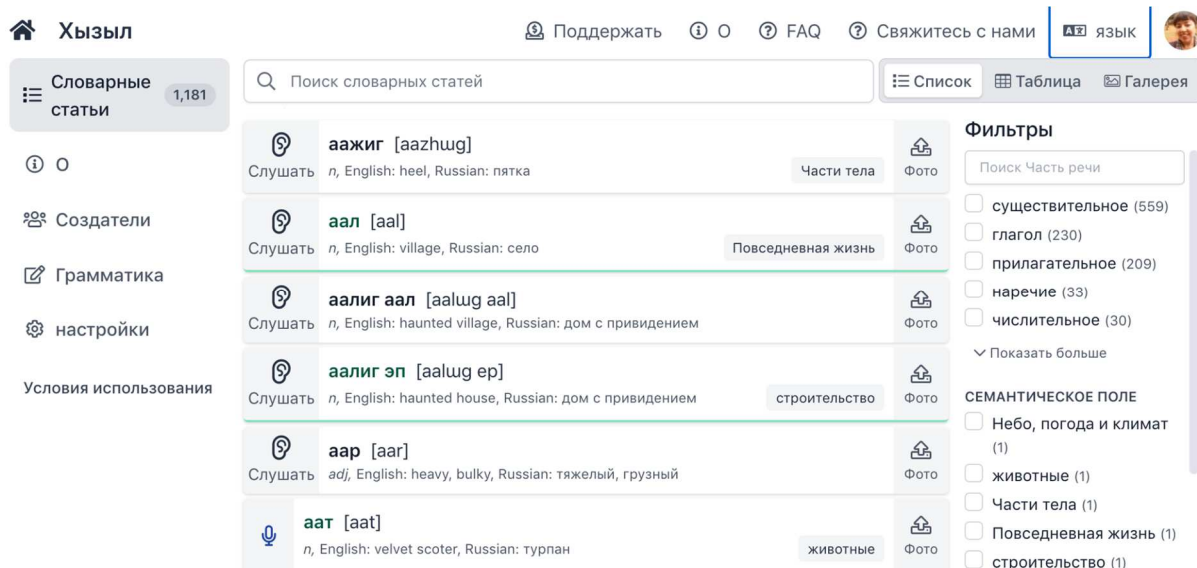


Figure 5: The Xyzyl Living Dictionary, displayed in the Russian (Cyrillic) interface.

Source: <https://livingdictionaries.app/xyzyl/entries/list>

Diccionario de lengua Tehuelche (aonekko 'a'ien) Donar Acerca de FAQ Contáctenos Idioma

Entradas 227 Búsqueda

Filtros

Buscar Partes de discurso

- ☐ sustantivo (93)
- ☐ verbo (34)
- ☐ frase (29)
- ☐ adverbio (10)
- ☐ pronombre (4)
- Mostrar más

HABLANTE

- ☐ Dora Manchado (215)
- ☐ javier domingo (11)
- ☐ Hay Audio (226)
- ☐ Hay Hablante (226)
- ☐ Hay Partes de discurso (182)

	'an wallaqqn	
Escuchar	fr, inglés: how nice, español: qué bonita	Foto
	'anglesh	
Escuchar	fr, inglés: English, español: inglés	Foto
	'anjieshm	
Escuchar	v, inglés: to dance, to buck, español: bailar, corcovear	Foto
	'appe	
Escuchar	s, inglés: plate, bowl, español: plato, cuenco	Foto
	'appe tcháan 'eu sh wáanko	
Escuchar	fr, inglés: wash the dishes before you leave, español: lavá los platos antes de irte	Foto
	'auer kkom	
Escuchar	fr, inglés: plucked, español: desplumado	Foto
	'elkke	
Escuchar	s, inglés: grandson, español: nieto	Foto

Figure 6: The Tehuelche (aonekko'a'ien) Living Dictionary displayed in the Spanish interface.

Source: <https://livingdictionaries.app/80CcDQ4DRyiYSPIWZ9Hy/entries/list>

Each dictionary can also include up to five glossing languages so that users may search for terms across regionally dominant and other relevant languages. For example, for Living Dictionaries for the tribal languages of the Munda family of India, glossing languages include English, Hindi and Oḍia (and sometimes other languages like Assamese or Bengali) so that users may search for terms in various languages. A Living Dictionary can also display up to five writing systems for an entry, which is useful for dictionaries where multiple competing scripts are used to represent a language. An example of one such project is the Birhor Living Dictionary (see Figure 7 for a sample entry), which is a multilingual resource that contains multiple glossing languages (English, Oḍia and Hindi) and multiple scripts (Devanagari and Oḍia). Another project, the Sora Living Dictionary, also includes an array of scripts and glossing languages (see Figure 8). In short, Living Dictionaries are designed explicitly with maximal inclusivity and unrestricted multilingualism in mind.

Birhor

Entries

150

About

Contributors

Grammar

Settings

Terms of Use

← Back

daʔ-re dʰer dʒug duɾup-mi

Devanagari Script
दाः-रे ढेर जुग दुड़ुप-मि

Oriya Script
ଦା:-ରେ ଢେଉ ଜୁଗ ଦୁଡୁପ-ମି

Phonetic
[daʔ-re dʰer dʒug duɾup-mi]

Oriya Gloss
ଦୀର୍ଘ ସମୟ ପର୍ଯ୍ୟନ୍ତ ପାଣିରେ ବସିବା

English Gloss
sit in water for a long time

Hindi Gloss
पानी में लम्बे समय तक बैठना

Part of Speech
intransitive verb

Semantic Domain
Physical Actions and States

Morphology
daʔ-re dʰer dʒug duɾup-mi

Interlinearization
water-LOC much time sit-2SG.SUBJ

Figure 7: *Entry view* for the phrase ‘sit in water for a long time’ (Birhor Living Dictionary).

Source: <https://livingdictionaries.app/birhor/entries/2SIhhZQdAr8ZfLaXI8f9>

kənsodəŋji [kənsodəndʒi] কান্‌চ'ড্‌নজি কঁৱৱাংখিল্লঙঙ
 Listen *n*, Assamese: কুকুৰা কেইতা, English: dogs, Hindi: कुत्ता, Oriya: କୁକୁର ଗୁଡ଼ିକ

Figure 8: The Sora word “dogs” in the Sora Living Dictionary (displayed in *list view*).

Source: <https://livingdictionaries.app/sora/entries/wPKHVIbyQgJhEVI1mCcI>

Figures 7 and 8 also show the types of information that can be provided for each entry in a Living Dictionary: headword, phonetic transcription, representation in different scripts, glosses into different languages, part of speech, semantic domain, morphology, interlinearisation, dialect name, audio recording and image file. All are optional metadata depending on the needs of the user, except for the headword. Not displayed in these entries are other optional fields, such as a sample sentence that contains the entry headword alongside a gloss of the sample sentence.

Living Dictionaries may be adjusted depending on what data the user wants to see. They may be viewed through three different types of visualisation: *list view*, *table view* and *gallery view* (settings that are available near the top right-hand corner of the “Entries” page). Each different setting provides the user with different ways of visualising and navigating the data inside the dictionary. *List view* (Figure 9) displays the data in a traditional dictionary list, *table view* (Figure 10) shows a spreadsheet of data, and *gallery view* (Figure 11) only pulls in entries with accompanying images.

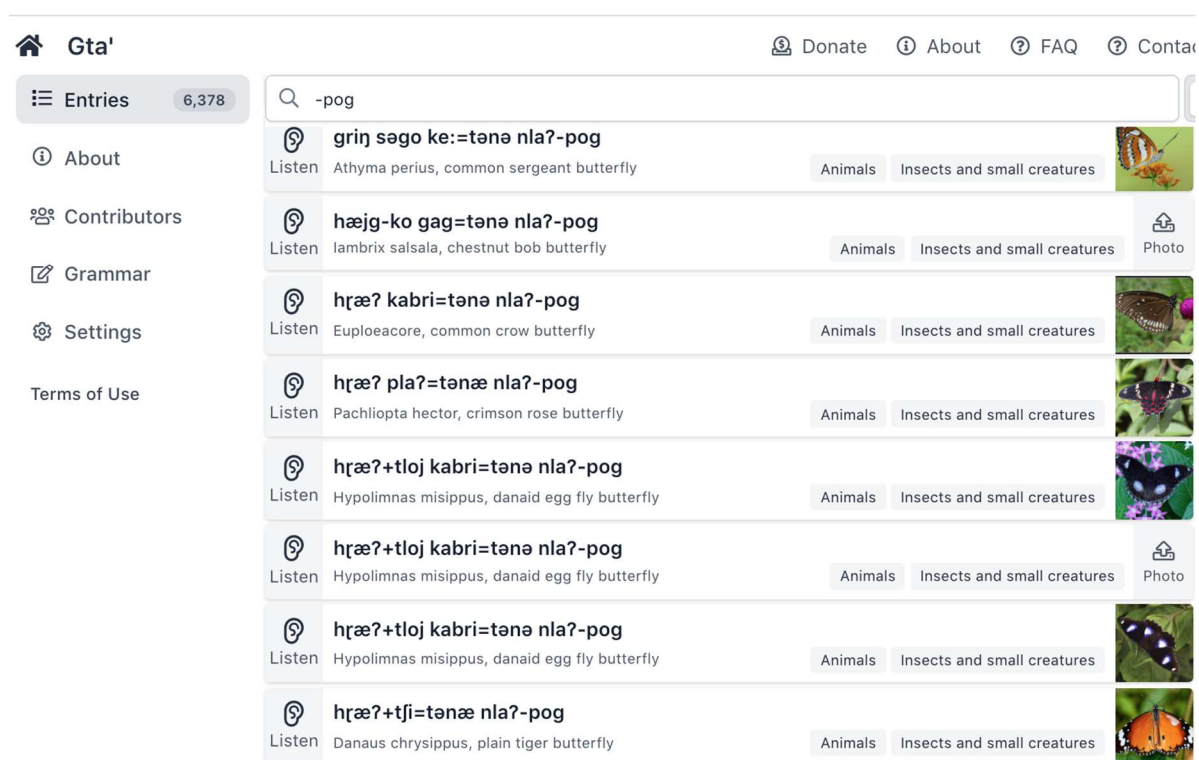


Figure 9: *List view* display of Gta? morpheme -pog ‘bug’ in the Gta? Living Dictionary

Source: <https://livingdictionaries.app/gta/entries/list>

Gta' Donate About FAQ

Entries 6,378

Entries: 103 (1ms)

Lexeme/Word/Phrase	Semantic Domain	Speaker
grin sɛgɔ ke:=tənə nla?-pog	Animals, Insects and small creatures	Lachmu Raspeda
hæjg-ko gag=tənə nla?-pog	Animals, Insects and small creatures	Lachmu Raspeda
hɛɛ? kabri=tənə nla?-pog	Animals, Insects and small creatures	Budra Raspeda
hɛɛ? pla?=tənə nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani
hɛɛ?+tloj kabri=tənə nla?-pog	Animals, Insects and small creatures	Budra Raspeda
hɛɛ?+tloj kabri=tənə nla?-pog	Animals, Insects and small creatures	Budra Raspeda
hɛɛ?+tloj kabri=tənə nla?-pog	Animals, Insects and small creatures	Budra Raspeda
hɛɛ?+tji=tənə nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani
hɛɛ?+tji=tənə nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani
hɛɛ?-roj=tənə nla?-pog	Animals, Insects and small creatures	Lachmu Raspeda
hɛɛ?=tənə nla?-pog	Animals, Insects and small creatures	Budra Raspeda
hɛɛ?=tənə nla?-pog	Animals, Insects and small creatures	Lachmu Raspeda
hɛɛ?+tji=tənə nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani
ikuɾa?+ikuɾa? tloj=tənə nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani
kabra=tənə nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani
kabra=tənə nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani
kabri nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani
kabri nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani
kabri=tənə nla?-pog	Animals, Insects and small creatures	Lachmu Raspeda
kimlin anor nla?-pog	Animals, Insects and small creatures	Dosorup Kirsani

Figure 10: *Table View* display of Gta? morpheme -pog 'bug' in the Gta? Living Dictionary

Source: <https://livingdictionaries.app/gta/entries/list>

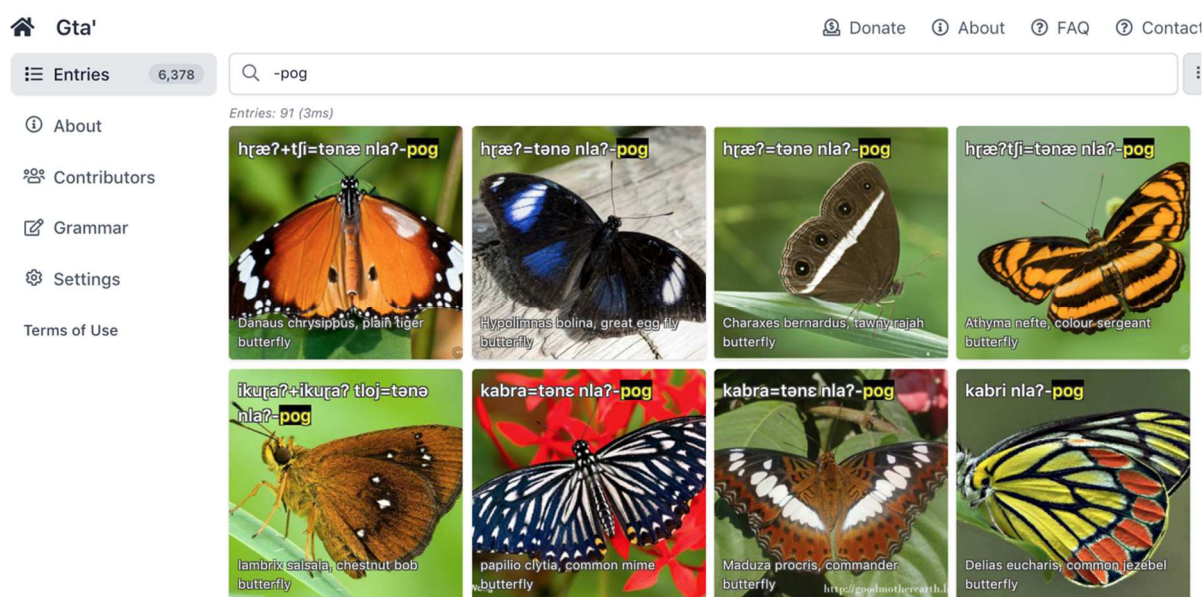


Figure 11: *Gallery view* display of Gta? morpheme -pog 'bug' in the Gta? Living Dictionary

Source: <https://livingdictionaries.app/gta/entries/list>

Search and use of entries in a platform such as Living Dictionaries are freed of the linear constraints of traditional dictionaries. As Lew and de Schryver (2014: 350) put it “[t]he user of a digital dictionary is no longer constrained by either the formal (spelling or phonology) or semantic criteria as the organizing principle. It is now perfectly possible to combine formal and semantic relations and utilise both types in

navigating the lexical material.” One key feature we have included in this are the tagging of entries according to semantic domains. The use of semantic domains as an organisational search principle is grounded in insights of cognitive linguistics (Langacker, 1987; Clausner & Croft, 1999; see also Bowers & Romary, 2018: 97) and allows for the generation of specific subsets of lexical entries to facilitate instruction in formal or informal educational settings in language revitalisation programs. Semantic domains are a sensitive issue because they often overlap and may be difficult to delineate. Our system allows for flexibility, and thus there is no limit to the number of semantic domains that can be used to tag entries. Users can also search by one or various semantic domain “filters” to yield tailored sets of results related to their domains of inquiry.

Ideologies of what is a ‘proper’ linguistic variety to be used are not relevant to the Living Dictionaries. Decisions guiding what dialects are represented (or not) within a Living Dictionary are community-driven. A digital dictionary may be created for any variety, whether it is oral or signed, recognised as a separate distinct language or ‘just’ a dialect, patois, Creole, pidgin, or any other lectal designation. Living Dictionaries can accommodate as many dialects or variants as desired by the community members creating the tool. For example, Zapotec and Mixteco communities in Mexico may wish to have a separate dictionary for each dialect, and therefore each dictionary will contain data from a specific dialect rather than showcasing multiple dialects. In the example below, the Mexican/American research team that created the first-ever Living Dictionary for the inactive indigenous language Opata (the name given to two closely related Uto-Aztecan tongues, Tegüima and Eudeve) decided it was best to group resources for both Opata varieties into one dictionary. They accomplished this by tagging the entries with the dialect names Tegüima and Eudeve (Figure 12).



Figure 12: An entry from the Opata Living Dictionary tagged as the “Tegüima” dialect.

Source: <https://livingdictionaries.app/opata/entries/yK1Yi17Fivn37BWDMMima>

6. Usage and Remote Collaboration

In terms of usage, there are currently close to 300 activists working on over 200 different Living Dictionaries on the platform, and more joining every week. In terms of dictionary size, recently created Living Dictionaries contain anywhere from a handful to several hundred entries, while many other Living Dictionaries that have been developed over the course of many years contain over 10,000 entries. Altogether, the platform contains over 250,000 entries and is growing each week.

One of the strengths of the Living Dictionaries is that they allow people to hear pronunciations of the words and phrases (Figure 13). We strongly encourage dictionary managers to upload audio files, or record audio content directly into the platform when possible, by using the microphone on their desktop or mobile device. If a dictionary manager does not speak the language fluently, we encourage them to locate a fluent speaker who can record audio entries later. Each dictionary, and each entry within a dictionary, is shareable with a unique URL that can be easily shared on social media or hyperlinked on other websites.

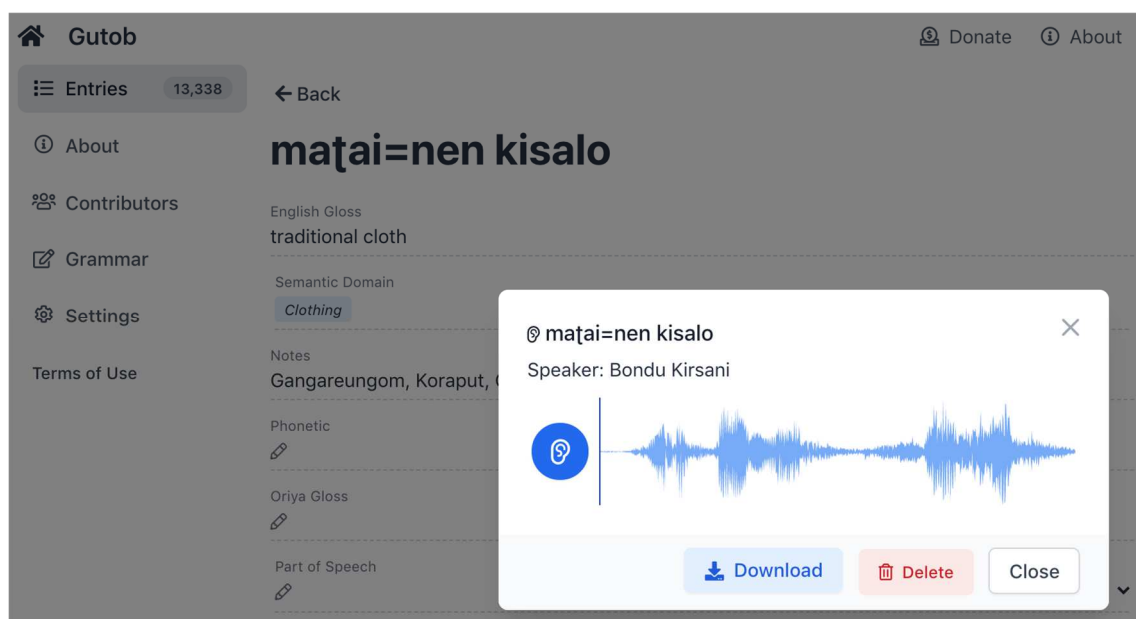


Figure 13: The audio waveform entry for “maṭai=nən kisalo” in the Gutob Living Dictionary

Source: <https://livingdictionaries.app/gutob/entries/KTJzdxbcYxtZsjRI2fTt>

Remote collaboration is possible and encouraged on the platform. Many existing Living Dictionaries have collaborators who work on different aspects of the work: some work on the text entries while others undertake the recording of the words and phrases based on the written data has been added to the system. There is no limit to the number of collaborators in a Living Dictionary. A dictionary manager may invite other collaborators to join the dictionary directly through the platform itself by using the “Invite Manager” or “Invite Collaborator” feature. Dictionary managers may add, edit or delete content. Contributors are project collaborators who can also add and edit but

cannot delete any content. The latter feature is designed for students and interns who may be working on the project as digital assistants, and they need to be able to safely work on content without deleting any of it by accident. The Living Dictionaries platform is engineered to have multiple collaborators logged into the system and editing a dictionary project at the same time, in real-time. The collaborators can be working remotely in different places in the world and see the exact same changes that are being made without even having to refresh their browsers, within seconds. There is no limit to the number of people who can be logged into a project at once, but we suggest that a team coordinates its strategy so that multiple people are not trying to edit the exact same entries at the same time.

7. The Future of Living Dictionaries

The platform is built to make ongoing relevant contributions to an increasingly dynamic world. As such, we continue to innovate and roll out new features on a regular basis. In 2021, we are releasing an updated International Phonetic Alphabet (IPA) Chart Picker on the platform, so that users may easily locate and select phonetic characters when they are creating (or editing) entries. It will be a great help for activists who need to be able to type effectively in IPA without leaving the platform. This year, we will also be launching our video integration feature, in which dictionary managers can directly record videos within entries, or link to existing YouTube videos, without ever leaving the platform. We are also working on displaying links to ecological databases within entries about species, which will help create a global network linking linguistic knowledge to other relevant databases. In 2020, we collaborated with the Ethno-Ornithology World Atlas to discuss and enact ways in which traditional ecological knowledge about birds can better interface with existing scientific online resources. Our intention is to keep these kinds of interdisciplinary discussions flowing so that our platform may become increasingly encyclopaedic over time. We also regularly meet with indigenous leaders, experts and scholars to discuss new opportunities for collaboration and avenues for language revitalisation that include Living Dictionaries.

Our long-term development roadmap includes expanding and improving features on the platform like speed optimisation, offline mode functionality, audio analysis and rolling out important new features such as export functionality (so that dictionary managers can retrieve their data in CSV, XML, JSON, PDF and other formats) and further multimedia integration. Based on user feedback, we intend to explore ways to integrate lists of culturally specific prompts by allowing users to draw from existing elicitation lists to start their dictionary projects from scratch. Users have also requested the implementation of an image API (Application Programming Interface) that would allow them to use relevant copyright-free images from sources such as Creative Commons directly in the platform. We intend to expand storage capacity exponentially over time and implement language localisation of the dictionary interface into two dozen additional dominant languages to serve the widest audience of endangered

language activists possible. We plan to implement notifications to increase real-time contributions and collaboration between users and begin regional campaigns to attract hundreds of new users and contributors worldwide. This will be done by demonstrating the software at regional and international gatherings of linguists and language activists to maximise the potential user groups as well as rolling out comprehensive training videos and webinars in various languages to assist contributors on the Living Dictionary platform.

In summary, the Living Tongues Institute has developed practical, web-based software (found at the URL LivingDictionaries.app) that can help people build a dictionary from the ground up. Moving forward, our team will continue to build and refine this framework for global application and deploy the platform at scale to serve all the world's endangered languages. This project can help mitigate the global language extinction crisis by opening the door to linguistic documentation for all, expanding access to cultural equity and self-determination. As an online platform that presently houses dictionaries for over 200 languages, it utilises the safety and flexibility of remote collaboration between dictionary managers. We are committed to maintaining this platform for decades to come so that the work of language activists may live on and benefit our descendants, community stakeholders, educators and scholars.

Acknowledgments

We gratefully acknowledge the support of the following grants that helped support the creation of Living Dictionaries for Sora, Remo, Birhor, Gtaʔ, Gutob and other Munda languages spoken in India: “Sora Typological Characteristics: Towards a Re-Evaluation of South Asian Human History” (National Science Foundation Linguistics Grant Award NSF/BCS #1844532, 2019-2022), “Documenting the Fragile Knowledge Domains of the Birhor People” (The Zegar Family Foundation, 2018-2020), “Documentation of Gutob, an endangered Munda language” (National Science Foundation, Documentation of Endangered Languages Grant Award #1500092, 2015-2017), “Documentation of Hill Gtaʔ—a seriously endangered Munda language” (National Endowment for the Humanities, Documentation of Endangered Languages Grant Award #PD-50025-13, 2013-2015) and “Documentation of Remo” (National Science Foundation, Documentation of Endangered Languages Grant Award #0853877, 2010-2011). We acknowledge web developer Jacob Bowdoin for his extensive contributions in programming the Living Dictionaries platform, and we thank our Living Tongues colleagues Dr. Luke Horo, Dr. Bikram Jora and Opino Gomango for their many contributions to the Living Dictionaries for Munda languages mentioned above. We acknowledge the work of Dr. K. David Harrison and Jeremy Fahringer who collaborated with us previously on a different desktop version of this platform, Talking Dictionaries (spearheaded by Living Tongues Institute for Endangered Languages and housed at Swarthmore College). We recognise the efforts of many volunteers and interns who have contributed to making the Living Dictionary platform multilingual and who have edited and updated data on the platform. For the interface languages

featured in this article, we thank the linguists and translators Dr. Michael Karani (KiSwahili interface), Dr. Denis Tokmashev (Russian interface), Amanda Chao Benbassat and Mónica Bonilla Parra (Spanish interface), Crisofia Langa da Camara (Portuguese interface) Yustinus (Yanche) Ghanggo Ate (Bahasa Indonesia interface), Nur Hidayah Binte Sunaryo (Bahasa Melayu interface) and Dana Melaver and Daniel Bögre-Udell (Hebrew interface) for their assistance. Lastly and very importantly, we acknowledge the tireless and ongoing efforts of the hundreds of language activists, speakers and scholars who are building their Living Dictionaries on our platform.

References

- Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Vol. 1. Stanford, CA: Stanford University Press.
- Anderson, G. D. S. (2017). Consequences of Russian Linguistic Hegemony in (Post-)Soviet Colonial Space. In Ramazan Korkmaz and Gürkan Dogan (eds.) *Endangered Languages of the Caucasus and Beyond*, pp. 1-16. Leiden: Brill.
- Prinsloo, D. J., Prinsloo, J. V. & Prinsloo, D. (2017). African Lexicography in the Era of the Internet. *The Routledge Handbook of Lexicography*. Pedro A. Fuertes Olivera (Ed.). London: Routledge, pp. 487-502.
- Dimitrova, L., Koseska(-Toszewa), V., Dutsova, R. & Panova, R. (2009). Bulgarian-Polish Online Dictionary — Design and Development. In V. Koseska-Toszewa, L. Dimitrova & R. Roszko (eds.) *Representing Semantics in Digital Lexicography: Innovative Solutions for Lexical Entry Content in Slavic Lexicography*. Warsaw, Institute of Slavic Studies, Polish Academy of Sciences, pp. 78-88.
- Prinsloo, D. J. (2019). A perspective on the past, present and future of lexicography with specific reference to Africa. In M. Gürlek, A. N. Çiçekler and Y. Taşdemir (eds.) *Asialex 2019*. Istanbul. 217-230.
- Rundell, M. (2012). It works in practice but will it work in theory? The uneasy relationship between lexicography and matters theoretical. In R. Vatvedt Fjeld & J. M. Torjusen (eds.). *Proceedings of the 15th Euralex International Congress*. 7-11 August 2012. Oslo.
- Clausner, T. C. & Croft, W. (1999). Domains and image schemas. *Cognitive Linguistics* 10, pp. 1-32.
- Lew, R. & de Schryver, G.-M. (2014). Dictionary users in the digital revolution. *International Journal of Lexicography* 27(4), pp. 341-59.
- Grin, F. (2003). The economics of language planning. *Current Issues in Language Planning* 4, pp. 1-66.
- Van Wyk, E.B. (1995). Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5, pp. 82-96.
- Whalen, D. H., Moss M. & Baldwin D. (2016). Healing through language: Positive physical health effects of indigenous language use. *F1000Research* 2016, **5**, pp. 852.

Websites:

ilc.cnr.it/ccurl2014 Accessed at: <http://www.ilc.cnr.it/ccurl2014/> (8 April 2021)
livingdictionaries.app. Accessed at: <https://livingdictionaries.app/> (8 April 2021)
slll.cass.anu.edu.au. Accessed at: <http://slll.cass.anu.edu.au/centres/andc/australex-2019> (8 April 2021)

Dictionaries:

Akumbu, Pius. (2021). *Babanki Living Dictionary*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/babanki>

Anderson, G.D.S., Jora, B. & Gomango, O. with Raspeda, B. Raspeda, P., Raspeda, L., Raspeda, K., Raspeda, K. Kirsani, D., Majhi, B., Majhi, S., Golpeda, B., & Mondal, I. (2021). *Gta' Living Dictionary*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/gta>

Anderson, G.D.S., Jora, B. & Gomango, O. with Sisa, T and Sisa, K., Kirsani, B., Kirsani, R. & Mondal, I. (2021). *Gutob Living Dictionary*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/gutob>

Anderson, G.D.S. & Harrison, K.D. with Muzaliwa, A., Makambo, D., Kituta A. & Etoke, A. (2021). *KiBembe Living Dictionary*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/kibembe>

Anderson, G.D.S., Gomango, O., Harrison, K.D. & Horo, L. with Gomango, S., Gomango, M., Gomango, O., Roita, Z., Roita, M., Sabar, T., Mondal, I. (2021). *Sora Living Dictionary*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/sora>

Anderson, G.D.S., Tokmashev, D., Fahringer, J., Harrison, K.D. & Tabatkin, M.M. (2021). *Xyzyl Living Dictionary*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/xyzyl>

Cortés Torres, F.D, Mantenuto, I., Espinoza, A.D, Pimentel Rojas, I., Anderson, G.D.S & Daigneault, A.L. (2021). *San Sebastián del Monte Mixtec (Tò'on Ndà'vi) Living Dictionary*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/san-sebastian-del-monte-m>

Jora, B., Anderson, G.D.S., Lakra, S. K., & Markki, A. with Birhor, A., Birhor, B., Birhor, B. Birhor, K., Birhor, K. Birhor, M., Birhor, M. Birhor, S. (2021). *Birhor Living Dictionary*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/birhor>

Fundación TEGUIMA-OPATA & Daigneault, A.L. (2021). *Opata Living Dictionary*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/opata>

Manchado, D., Domingo, J., Duval, N., Arias García, N. Berger, T., Daigneault, A.L. (2021). *Diccionario de lengua Tehuelche (aonekko 'a'ien)*. Living Tongues Institute for Endangered Languages. <https://livingdictionaries.app/80CcDQ4DRyiYSPIWZ9Hy>

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

