

Lingua Libre

What is it and how to take part in it?

A workshop by [[User:WikiLucas00|Lucas Prégaldiny]]

WIKIMANIA





Lingua Libre

alias “LiLi”

1. The project

2. Demonstration

3. Impact

4. Your turn!

1. The project in a nutshell

Goals of Lingua Libre

- Document oral linguistic diversity
- Feed the Wikimedia projects with audio content
- Bring new linguistic communities to Wikimedia projects



Origins: a multidisciplinary team

- The chapter [Wikimédia France](#)
- Contributors from the French-speaking Wiktionary ([Wiktionnaire](#))
- Linguists
- Speakers of some French regional languages
- The creator of [Shtooka](#), a PC software for recording pronunciations

- A [Project Grant](#) provided to [0x010C](#)
- Funds from the [French Ministry of Culture](#)



**WIKIMEDIA
FRANCE**



**MINISTÈRE
DE LA CULTURE**

*Liberté
Égalité
Fraternité*

How does it work?

- [Lingua Libre](#) is a wiki, runs on MediaWiki (1.35) and [BlueLL](#) is its custom skin
- A [Wikibase](#) is installed, allowing SPARQL queries
- The core of LiLi is a MW extension based on Voice Activity Detection ([RecordWizard](#))
- The files are sent to Commons via OAuth on behalf of the user ([Upload2Commons](#))
- [Datasets](#) can also be downloaded in .zip files (Work in progress)

Advantages

Without
Lingua Libre



With



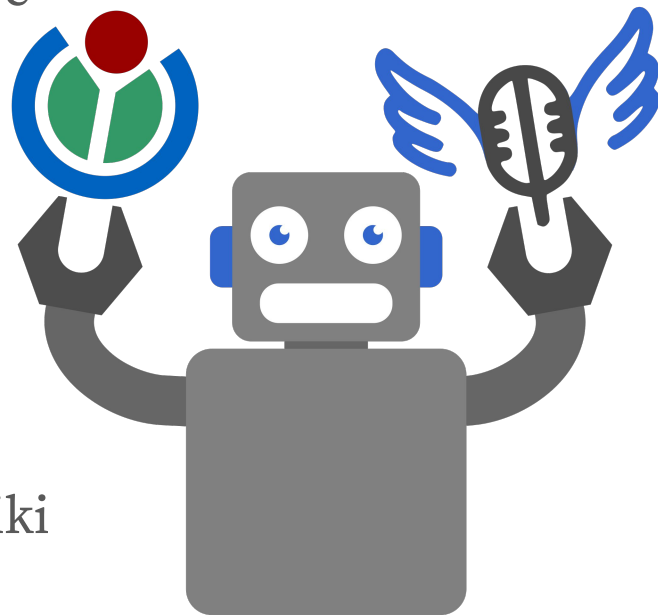
Lingua Libre Bot

What is its purpose?

- Integrates Commons-hosted Lingua Libre files to local wikis
- Coded in Python, see [GitHub](#)

How to have it on your wiki?

- Fill in a request for your wiki on the [LL:Bot](#) page
- Start a community discussion on your wiki and a vote for its Bot status



2. Demonstration

Logging in

A Wikimedia account is required, as Lingua Libre uses an OAuth token

It means that it will edit Commons on your behalf

The image shows a screenshot of the Lingua Libre Record Wizard interface. At the top left is the Lingua Libre logo, a stylized microphone with wings. Below it, the text "Lingua Libre" is displayed. To the right, the text "Record Wizard" is visible. A large blue banner on the left side of the interface reads "Login required" in white text, with the subtext "Please log in to upload files." below it. A white arrow points from this banner towards the right side of the screen. On the right side, there is a "Special page" header and a "WIKIMEDIA COMMONS" logo. Below the logo is a list of navigation links: "Main page", "Welcome", "Community portal", "Village pump", "Help center", "Participate", "File", "Changes", "Random file", and "Contact us". At the top right of the interface, there are links for "English" and the user name "WikiLucas00". A modal dialog box is open in the foreground, titled "OAuth". It contains the text: "Hi WikiLucas00, In order to complete your request, **Lingua Libre** needs permission to perform the following actions on your behalf on *commons.wikimedia.org*:". Below this text are two bullet points: "• **Interact with pages** Create, edit, and move pages" and "• **Interact with media** Upload, replace, and move files". At the bottom of the dialog, there is a "Privacy policy" link, a blue "Allow" button, and a "Cancel" button.



Access page


Lingua Libre Record Wizard Discussion Statistics Help About

Search 🔍 🗣️ ENGLISH 📺 Record 🔔 👤 ✎ Edit 👥 Users

Welcome to Lingua Libre, the participative linguistic media library of Wikimedia France.

Latest recordings

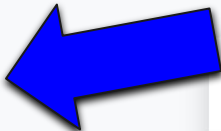
-  **চাঁদমালয়**
Bengali - Titodutta
-  **চাঁদমালার**
Bengali - Titodutta



Test your microphone

Record a voice

- 1 Tutorial
- 2 Speaker
- 3 Details
- 4 Studio
- 5 Publish



Check your microphone

Here's how to do it:

1. When you're ready, press the button above to start a record trial.
2. Pronounce a word aloud, for example "the sun".
3. Listen to your record to check that it's correct.
4. Adapt your microphone's configuration if needed.
5. If everything is correct, you can move on to the next step!

! To learn how to configure your microphone, visit [this help page](#).

When you are ready, press the button below to start the test record.

Start the test

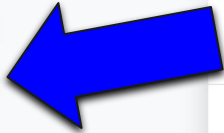
Cancel

Next >

Fill a speaker profile

Record a voice

- ✓ Tutorial
- 2 **Speaker**
- 3 Details
- 4 Studio
- 5 Publish



Speaker profile

! This information will be publicly displayed and associated with your records.

Speaker profile to use

Test ▾

Name to display

Test

Gender

Male Female Other

Spoken languages

French × *

Place of residence

Paris

License

I, Test, irrevocably grant anyone the right to use my audio records created with this tool under the following license:

Creative Commons Attribution ShareAlike 4.0 ▾

You agree that they may be automatically used on several Wikimedia projects, including Wikipedia, Wiktionary and Wikidata.

Cancel

< Previous

Next >

If a language is not available

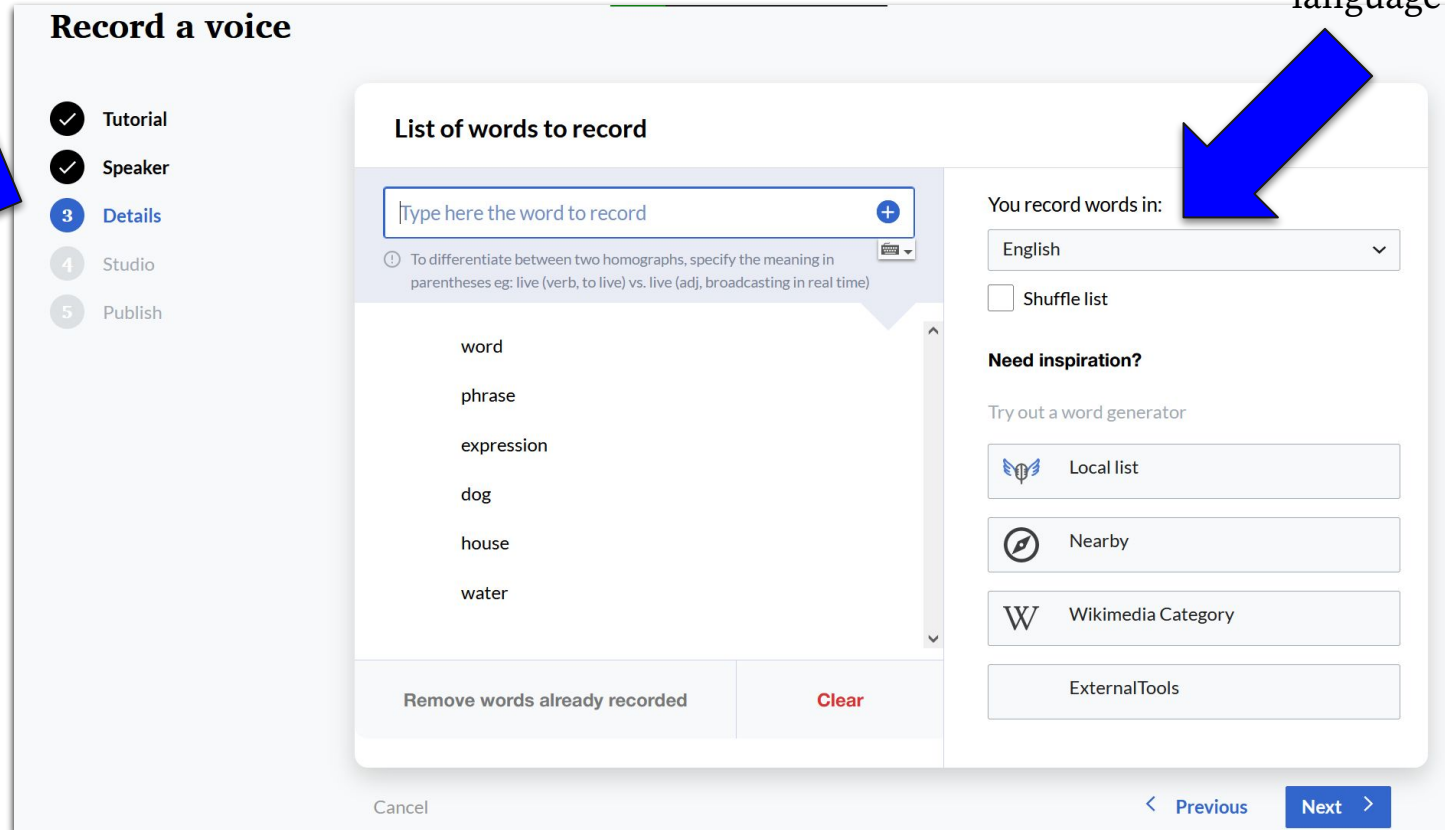
→ Ask an administrator via their talk page or the chat room, they will add it rapidly

[List of admins](#)

Several ways to choose the words you will record

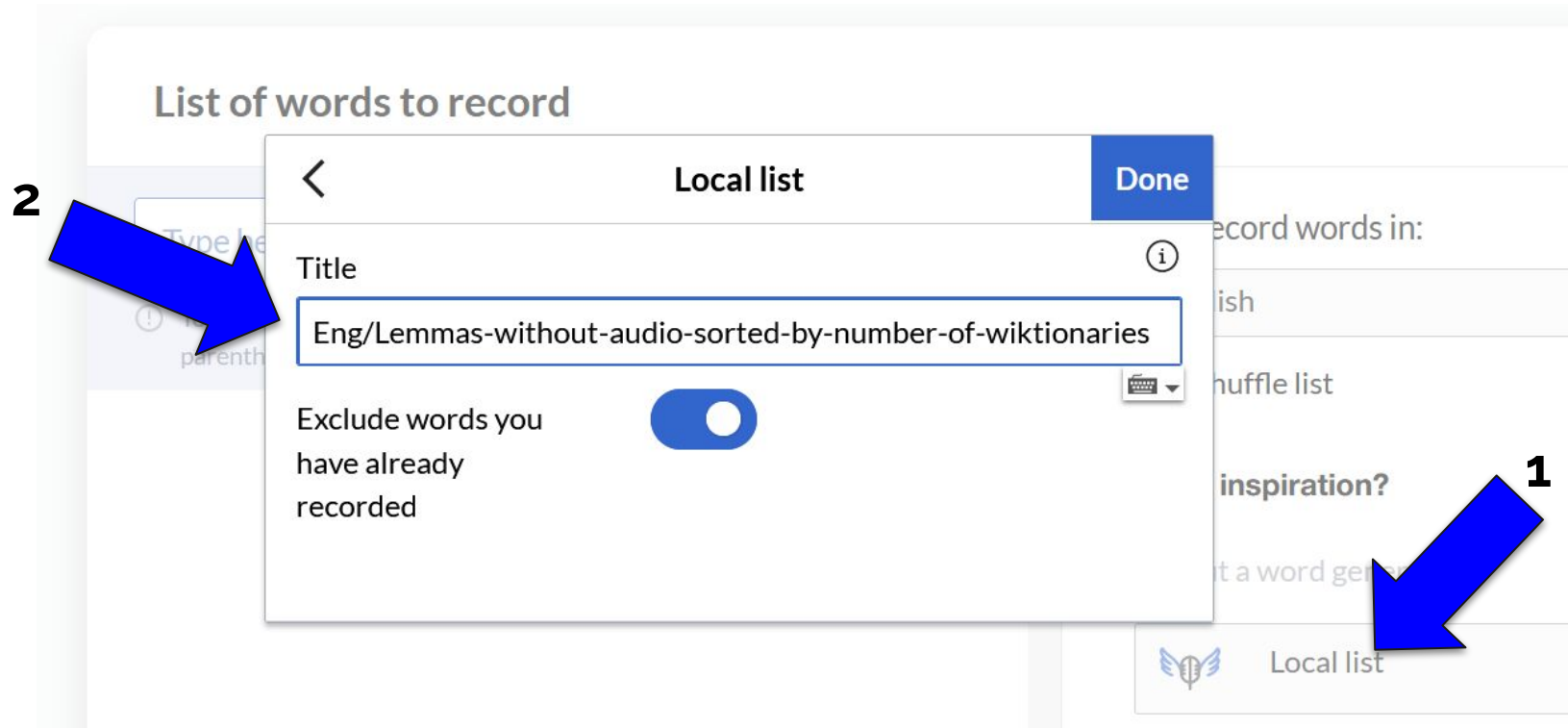
A. Add the words one by one

double-check the language you will record



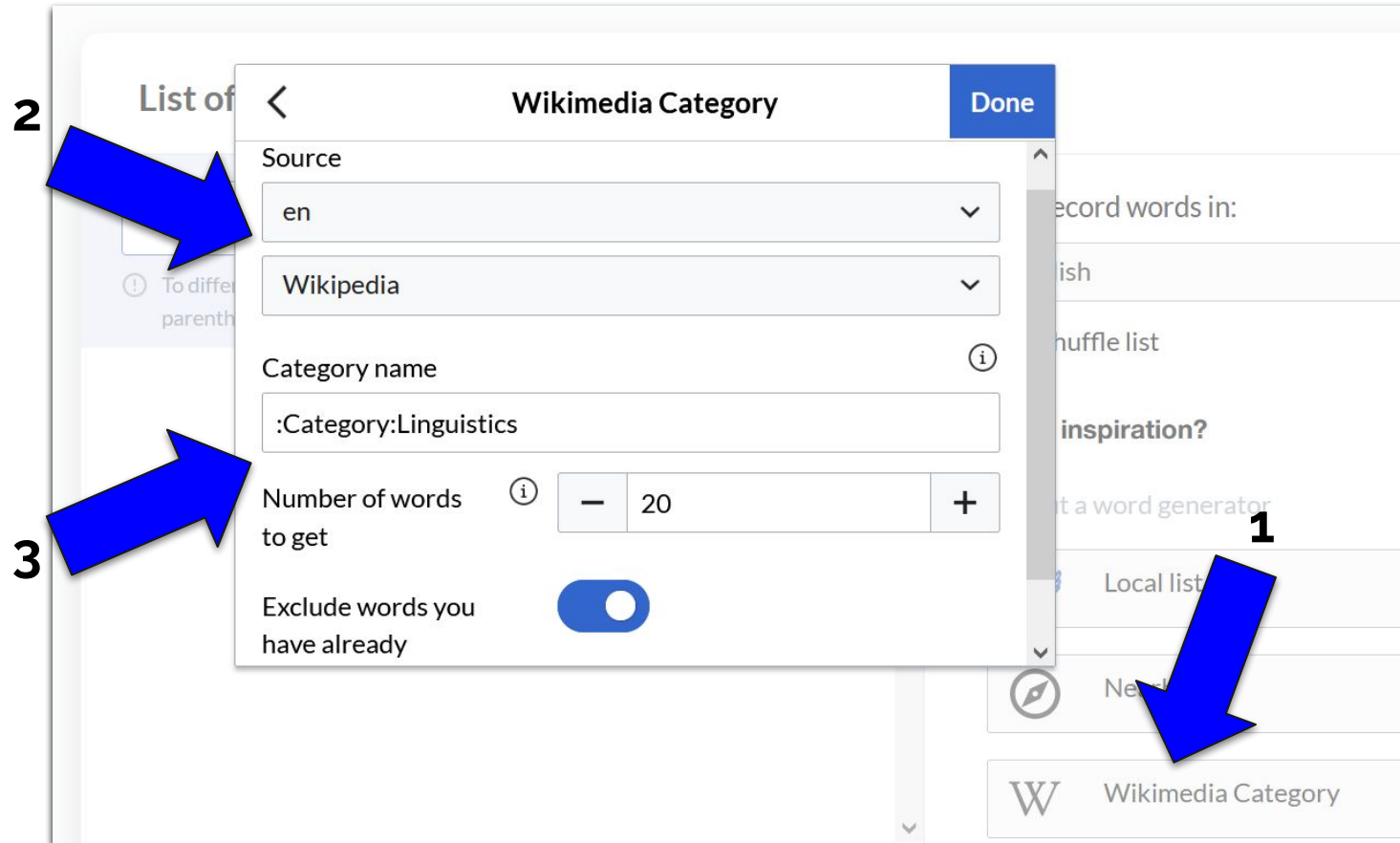
The screenshot shows the 'Record a voice' interface. On the left, a vertical sidebar contains five steps: 'Tutorial' (checked), 'Speaker' (checked), 'Details' (selected with a blue circle and number 3), 'Studio' (number 4), and 'Publish' (number 5). A large blue arrow points to the 'Details' step. The main area is titled 'Record a voice' and contains a 'List of words to record' section. This section has a text input field with the placeholder 'Type here the word to record' and a plus icon. Below the input is a warning icon and text: 'To differentiate between two homographs, specify the meaning in parentheses eg: live (verb, to live) vs. live (adj, broadcasting in real time)'. A list of words is shown: 'word', 'phrase', 'expression', 'dog', 'house', and 'water'. At the bottom of this list are two buttons: 'Remove words already recorded' and 'Clear'. To the right of the list is a section 'You record words in:' with a dropdown menu set to 'English' and a 'Shuffle list' checkbox. Below this is a 'Need inspiration?' section with the text 'Try out a word generator' and five buttons: 'Local list', 'Nearby', 'Wikimedia Category', and 'ExternalTools'. At the bottom of the interface are 'Cancel', '< Previous', and 'Next >' buttons. A second large blue arrow points to the 'English' dropdown menu.

B. Use an existing local list



Local lists can be manually created, or by bot, like [Olaf's bot lists for +70 languages](#)

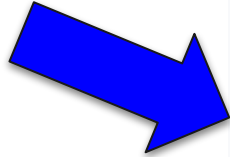
C. Use a category from Wikipedia or Wikt.



Recording studio

Record a voice

- ✓ Tutorial
- ✓ Speaker
- ✓ Details
- 4 Studio**
- 5 Publish



Studio

- ▶ language
- lexicon
- outline of linguistics
- Noam Chomsky
- linguistic description
- theoretical linguistics
- pleonasm
- metalanguage
- Object language
- gloss
- rhema
- conditional sentence
- topic-comment
- possible world

Click on \$1 below, then read the word aloud

language

Skip >

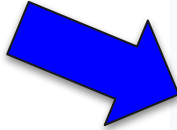
0 / 20

Cancel < Previous Next >

Check-listen before you submit

Record a voice

- ✔ Tutorial
- ✔ Speaker
- ✔ Details
- ✔ Studio
- 5 Publish



Publication

- language
- lexicon
- ▶ language classification
- language documentation
- phrase
- sentence

Check your records before starting their publication.

< language classification >

0:02 / 0:02

0 / 4

Cancel [Previous](#) [Publish on Wikimedia Commons](#)



Verify the uploading, retry if needed

Record a voice

- ✓ Tutorial
- ✓ Speaker
- ✓ Details
- ✓ Studio
- 5 Publish

Publication

- ⚠ language ✓
- lexicon
- ▶ language classification
- ✓ language documentation ✓
- ✓ phrase ✓
- ✓ sentence ✓

Check your records before starting their publication.

< language classification > 0:02 / 0:02

3 / 4
⚠ 1

Cancel [Check your uploads on Wikimedia Commons](#) **↻ retry failed upload** > Record more words

Check your uploads on Commons



[Main page](#)
[Welcome](#)
[Community portal](#)
[Village pump](#)
[Help center](#)

[Participate](#)
[Upload file](#)
[Recent changes](#)
[Latest files](#)
[Random file](#)
[Contact us](#)

[Tools](#)
[User uploads](#)
[User contributions](#)
[Logs](#)
[Email this user](#)
[Mute preferences](#)
[View user groups](#)
[Special pages](#)
[Printable version](#)
[Perform batch task](#)
[JS Wiki Browser](#)

🌐 English 🧑 WikiLucas00

Special page

Uploads by WikiLucasoo

[For WikiLucas00 \(talk · contributions · Move log · block log · uploads · Abuse filter log\)](#)

This special page shows all uploaded files that have not been deleted; for those see the [upload log](#).

File list

Items per page:

50

Username:

WikiLucas00

Include old versions of files

Go

⏪ First page ⏩ Previous page Next page ⏪ La:

▼ Date	Name	Thumbnail	Size
21:34, 13 August 2021	LL-Q1860 (eng)-WikiLucas00-language.wav (file)		46 KB
21:32, 13 August 2021	LL-Q1860 (eng)-WikiLucas00-sentence.wav (file)		46 KB
21:32, 13 August 2021	LL-Q1860 (eng)-WikiLucas00-phrase.wav (file)		38 KB
21:32, 13 August 2021	LL-Q1860 (eng)-WikiLucas00-language documentation.wav (file)		70 KB

In case a mispronounced recording has been send, you only need to record it again on Lingua Libre

Categorized by language and user on Commons

WIKIMEDIA COMMONS

Main page
Welcome
Community portal
Village pump
Help center

Participate
Upload file
Recent changes
Latest files
Random file
Contact us

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Perform batch task
RSS feed
Nominate category for discussion
JS Wiki Browser

Print/export
Create a book
Download as PDF
Printable version
In Wikipedia

English WikiLucas00 Talk Preferences Beta Watch

Category: **Discussion** View Edit History More Search Wikia

Category:Lingua Libre pronunciation-ron

Categories (+*): [Lingua Libre pronunciation \(-\) \(±\) \(l\) \(t\)](#) | [Romanian pronunciation \(-\) \(±\) \(l\) \(t\) \(+\)](#)

From Wikimedia Commons, the free media repository

Pronunciation audio files recorded with [Lingua Libre](#), in Romanian.

Media in category "Lingua Libre pronunciation-ron"

The following 200 files are in this category, out of 17,926 total.

(previous page) (next page)

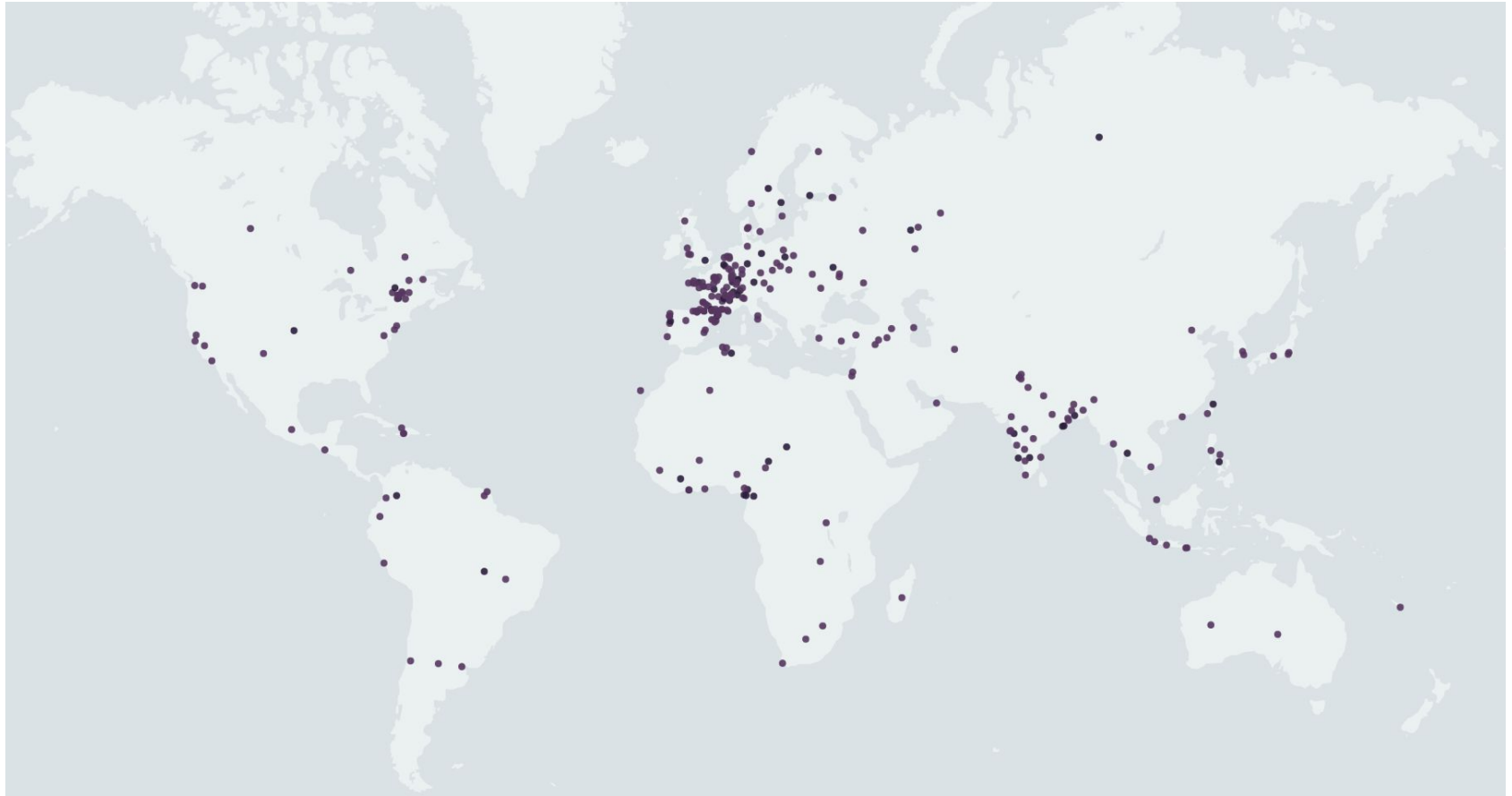
LL-Q7913 (ron)-KlaudiuMihaila-1-Decembrie.wav 1.4 s; 121 KB	LL-Q7913 (ron)-KlaudiuMihaila-2-Mai.wav 1.1 s; 97 KB	LL-Q7913 (ron)-KlaudiuMihaila-23-August.wav 1.7 s; 145 KB	LL-Q7913 (ron)-Andreea-Teodora-a o lăsa baltă.wav 1.8 s; 171 KB	LL-Q7913 (ron)-Andreea-Teodora-a se așeza.wav 1.6 s; 147 KB	LL-Q7913 (ron)-Andreea-Teodora-a se chinui.wav 1.5 s; 139 KB
LL-Q7913 (ron)-Andreea-Teodora-a se căsători.wav 1.4 s; 131 KB	LL-Q7913 (ron)-Andreea-Teodora-a se datora.wav 1.6 s; 147 KB	LL-Q7913 (ron)-Andreea-Teodora-a se descurca.wav 1.5 s; 139 KB	LL-Q7913 (ron)-Andreea-Teodora-a se enerva.wav 1.7 s; 155 KB	LL-Q7913 (ron)-Andreea-Teodora-a se indigna.wav 1.5 s; 139 KB	LL-Q7913 (ron)-Andreea-Teodora-a se lăsa de.wav 1.5 s; 139 KB
LL-Q7913 (ron)-Andreea-Teodora-a se multumi cu.wav	LL-Q7913 (ron)-Andreea-Teodora-a se mândri cu.wav	LL-Q7913 (ron)-Andreea-Teodora-a se ocupa de.wav	LL-Q7913 (ron)-Andreea-Teodora-a se plânge.wav	LL-Q7913 (ron)-Andreea-Teodora-a se preface.wav	LL-Q7913 (ron)-Andreea-Teodora-a se pricepe.wav

https://commons.wikimedia.org/wiki/Category:Lingua_Libre_pronunciation-ron

3. Impact

A. The community

Map of Lingua Libre's speakers (>550)



An international community

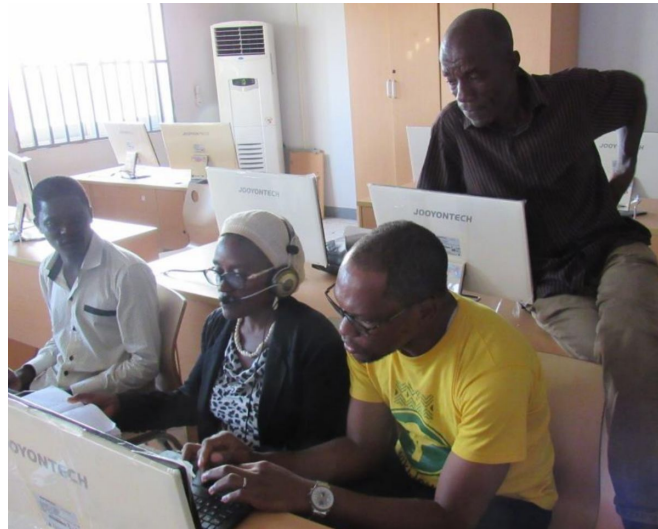


**Recording session in
Atikamekw (Canada)**

Photo by Benoit Rochon,
CC BY-SA 3.0

Recording some languages of Cameroon

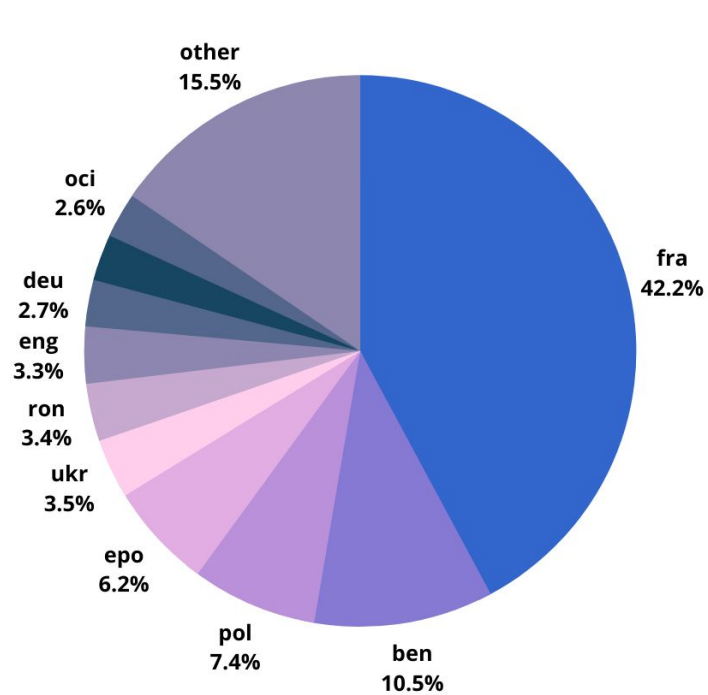
Photo by Geueur
[original pdf], CC BY-SA 4.0



**Team of Wikimedia France and
LiLi contributors**

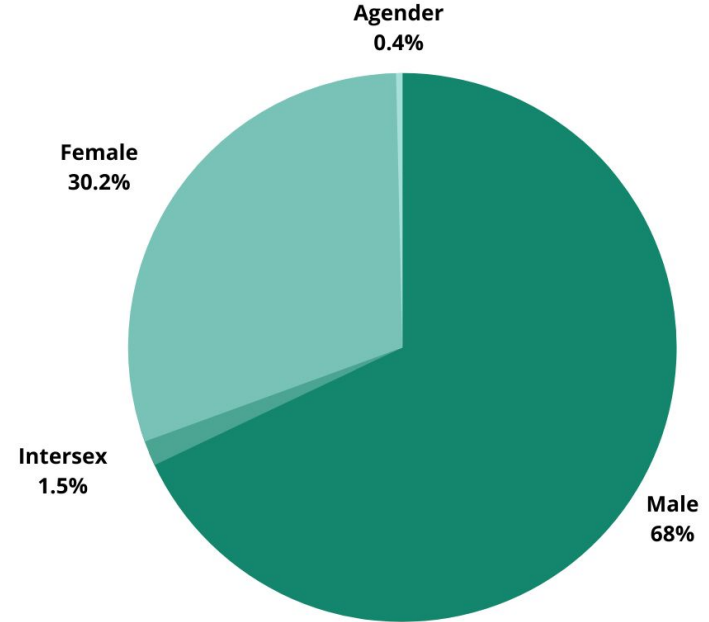
Photo by 0x010C,
CC BY-SA 4.0

Statistics: Languages and Genders



Languages

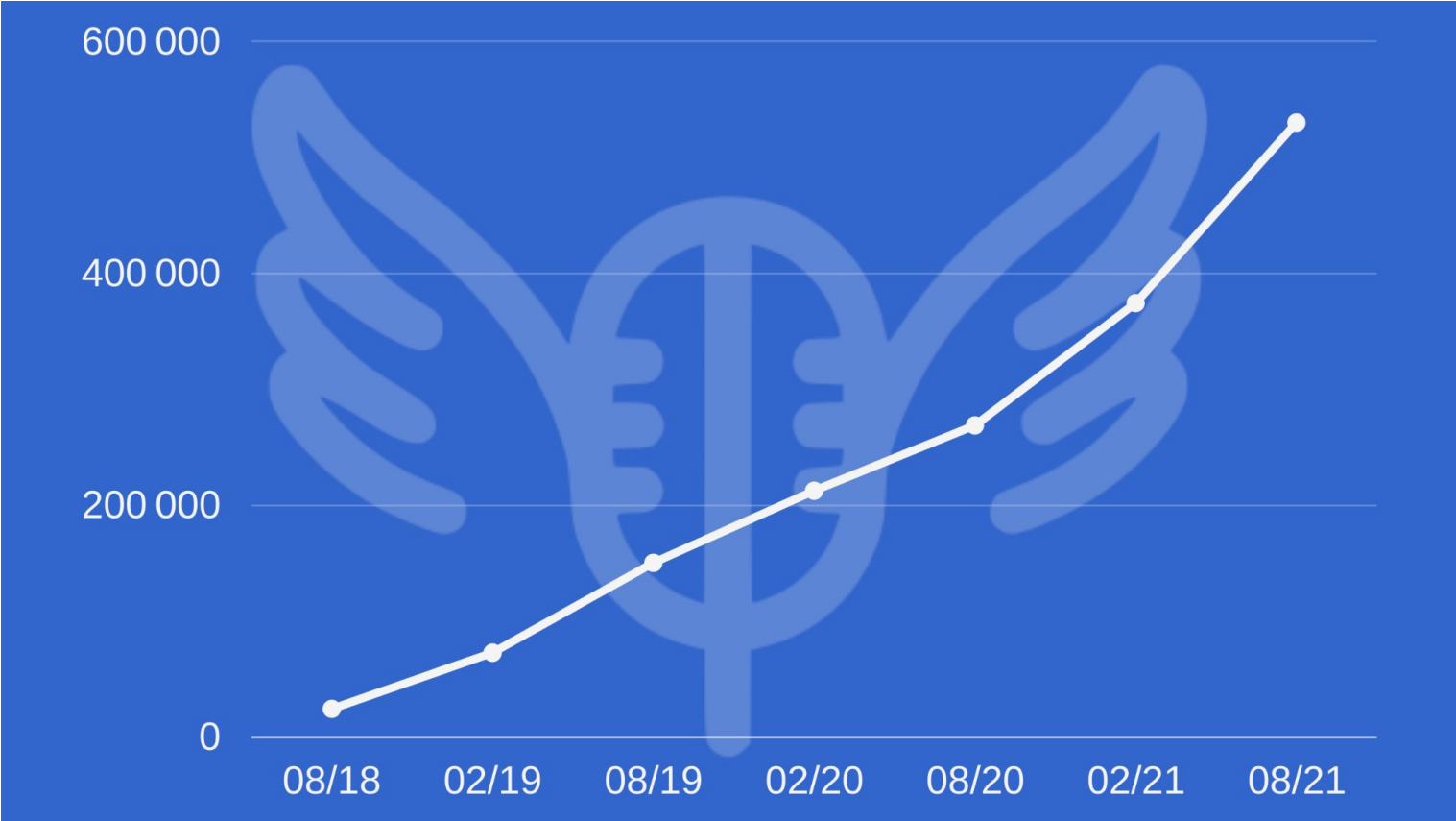
Total of 530k recordings, 127 languages



Genders

Total of 550 speakers

Statistics



B. Use of the recordings

Examples of use on the Wikimedia projects

Illustrate the pronunciation of words and names on different projects:
Wikipedias, Wiktionaries, Wikidata, Wikivoyage, Wikisource...

Page:SIL 1996 Ajilabal.djvu/35

This page has not been proofread.

Chica ntiel ri números de 1 a 20 pa kach'ab'al

Los números de 1 a 20 en Cakchiquel

1 jun	▶	11 julajuj	▶
2 ca'ye'	▶	12 cab'alajuj	▶
3 oxé'	▶	13 oxlajuj	▶
4 caje'	▶	14 cajlajuj	▶
5 vu'uo'	▶	15 vulajuj	▶
6 vaki'	▶	16 vaklajuj	▶
7 vuku'	▶	17 vuklajuj	▶
8 vakxaki'	▶	18 vakxaklajuj	▶
9 b'ilijje'	▶	19 b'ilijlajuj	▶
10 lajuj	▶	20 juvinak	▶

Chica ntiel ri números de 1 a 20 pa kach'ab'al
Los números de 1 a 20 en Cakchiquel

1 <input type="checkbox"/> jun	11 <input type="checkbox"/> julajuj
2 <input type="checkbox"/> ca'ye'	12 <input type="checkbox"/> cab'alajuj
3 <input type="checkbox"/> oxé'	13 <input type="checkbox"/> oxlajuj
4 <input type="checkbox"/> caje'	14 <input type="checkbox"/> cajlajuj
5 <input type="checkbox"/> vu'uo'	15 <input type="checkbox"/> vulajuj
6 <input type="checkbox"/> vaki'	16 <input type="checkbox"/> vaklajuj
7 <input type="checkbox"/> vuku'	17 <input type="checkbox"/> vuklajuj
8 <input type="checkbox"/> vakxaki'	18 <input type="checkbox"/> vakxaklajuj
9 <input type="checkbox"/> b'ilijje'	19 <input type="checkbox"/> b'ilijlajuj
0 <input type="checkbox"/> lajuj	20 <input type="checkbox"/> juvinak

Numbers from 1 to 20 in
Kaqchikel on [Wikisource](#)

Prononciation [modifier le wikicode]

- France : écouter « une forêt [yn fo.ʁɛ] »
▶ 0:00 MENU
- (Région à préciser) : écouter « forêt [fo.ʁɛ] »
▶ 0:00 MENU
- France (Yvelines) : écouter « forêt [fo.ʁɛ] »
▶ 0:00 MENU
- France : écouter « forêt [fo.ʁɛ] »
▶ 0:00 MENU
- France (Massy) : écouter « forêt [fo.ʁɛ] »
▶ 0:00 MENU
- France : écouter « forêt [fo.ʁɛ] »
▶ 0:00 MENU
- France (Paris) : écouter « forêt [fo.ʁɛ] »
▶ 0:00 MENU
- Suisse (Genève) : écouter « forêt [fo.ʁɛ] »
▶ 0:00 MENU

Entry [forêt](#) on the French
speaking Wiktionary

Lexeme Discussion

(L310162) **চন্দ্রেরখা**

bn

Language Bengali
Lexical category noun

Statements

has quality inanimate
- 0 references
+ add reference
+ add value
+ add statement

Senses
+ add Sense

Forms

L310162-F1 **চন্দ্রেরখা**
bn

Grammatical features: nominative case

Statements about L310162-F1

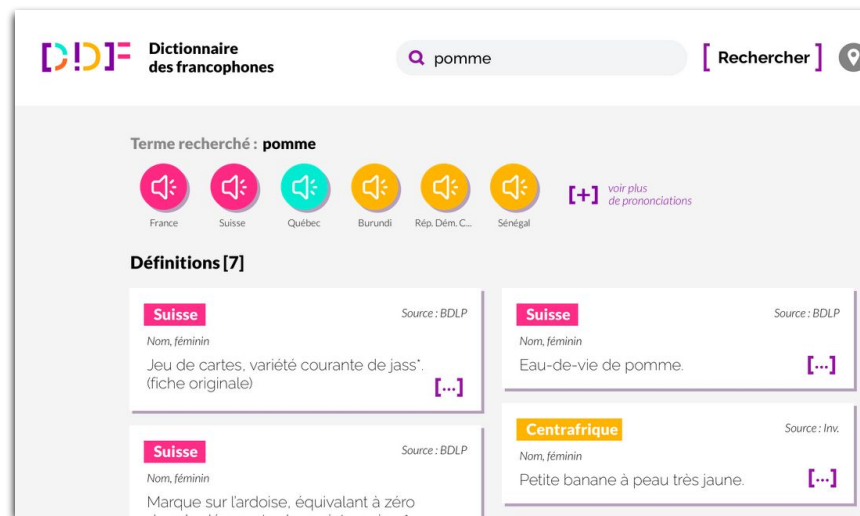
pronunciation audio

LL-Q3610 (ben)-Tisodutta-চন্দ্রেরখা.wav
1.6 s; 147 KB
language of work or name: Rarhi dialect
- 0 references
+ add reference

A [lexeme](#) in Bengali
on Wikidata

Examples of use outside of Wikimedia projects

- Some recordings are used as training data for Mozilla's speech to text engine, [DeepSpeech](#)
- [Le Dico des Ados](#), a libre dictionary for young people, records and uses some recordings.
- The [Dictionnaire des francophones](#) will soon be including Lingua Libre recordings to highlight the accents of the French-speaking world.
- Perspectives for language-learning applications



The screenshot shows the 'Dictionnaire des francophones' website. At the top, there is a search bar with the word 'pomme' and a search button labeled '[Rechercher]'. Below the search bar, the text 'Terme recherché : pomme' is displayed. Underneath, there are six circular icons representing different regions: France (pink), Suisse (purple), Québec (teal), Burundi (orange), Rép. Dém. C. (yellow), and Sénégal (yellow). To the right of these icons is a link that says '[+] voir plus de prononciations'. Below this, the section 'Définitions [7]' is shown, containing four definition cards. Each card has a regional label (Suisse or Centrafrique), the source (BDLP or Inv.), the gender (Nom, féminin), and a description of the word. The first card for 'Suisse' (BDLP) describes 'Jeu de cartes, variété courante de jass* (fiche originale)'. The second card for 'Suisse' (BDLP) describes 'Eau-de-vie de pomme'. The third card for 'Suisse' (BDLP) describes 'Marque sur l'ardoise, équivalent à zéro'. The fourth card for 'Centrafrique' (Inv.) describes 'Petite banane à peau très jaune'. Each card has a '[...]' link at the bottom right.

Thank you!
Do you have questions?

5-10 minutes Q&A before the interactive session

4. We need you!

Interactive session

A. Get involved in Lingua Libre

Several ways of getting involved

- The whole content can be translated, learn more at [Help:Translate](#)
- [Create lists](#) in your own language
- Any question, suggestion, news? The [Chat Room](#) is yours
- For technical discussions, see the [Tech Board](#)
- Help us improving the code via [GitHub](#) and [Phabricator](#)
- Request Lingua Libre Bot on a specific wiki [here](#)
- Spread the word on your own wikis
- **[Record your voice!](#)**

B. Record your voice

Record your voice: advice

- Lingua Libre works better on Firefox for the moment
- Please mind the noises around you (and turn off the sound of this session!) to avoid interferences
- Follow the steps of the Record Wizard or read the [demonstration slides](#)
- Please listen to your recording before uploading them to Commons
- Watch your files on Commons, and on the main page of Lingua Libre!



Credits

- [Young girls reading - Government primary school in Amman, Jordan](#), Tanya Habjouqa, CC-BY-SA 3.0 IGO
- [LinguaLibreAtikamekw3](#), Seeris, CC-BY-SA 4.0
- [Wikimedia Hackathon Barcelona 2018 \(07\)](#), Unuaiga, CC-BY-SA 4.0
- [Audio record workflow without Lingua Libre](#), 0x010C, CC-BY-SA 4.0
- [Audio record workflow with Lingua Libre](#), 0x010C, CC-BY-SA 4.0
- [Lingua Libre Bot icon](#), WikiLucas00, CC-BY-SA 4.0
- [2021-07 speakers map Lingua Libre \(light\)](#), WikiLucas00, CC-BY-SA 4.0
- [Lingua Libre Atikamekw at Wikimania 2017 Montreal](#), Benoit Rochon, CC-BY-SA 3.0
- [Lingua Libre at Wiki Indaba 2019](#) [extracted], Geugeor, CC-BY-SA 4.0
- [2018-11 Lingua Libre meeting](#), 0x010C, CC-BY-SA 4.0
- [2021-08 Lingua Libre graph of total recordings](#), WikiLucas00, CC-BY-SA 4.0
- [Lingua Libre Logo](#), Marc Brouillon and 0x010C, CC-BY-SA 4.0
- [Logo Wikimedia France](#), WMF / Wikimedia France, CC-BY-SA 4.0
- [Firefox logo, 2019](#), Mozilla Corporation, MPL 2.0
- Screenshots of Lingua Libre, Wikimedia Commons, Wikisource, Wiktionary, Wikidata, Dictionnaire des Francophones websites
- Graphs by WikiLucas00 - CC-BY-SA 4.0
- Presentation templates by the Wikimedia Foundation and by WikiLucas00

Thanks to [Eavq](#), [Noé](#), [Adélaïde Calais WMFr](#) and [Nicolas Lopez de Silanes WMFr](#) for their help with developing this presentation.

This document was made by WikiLucas00 - CC-BY-SA 4.0