Harmonizing Voices with WikiTutur

Wiktionary and Lingua Libre Collaboration for Language Preservation

Arjuna Fiqrillah - Ni Putu Diah Asyanti
arjunafiqrillah@gmail.com
diahasyanti@gmail.com

WIKIMEDIA ESEAP CONFERENCE
KOTA KINABALU 2024
COLLABORATION BEYOND THE HORIZON

ESEAP KK24

Scan here!

https://bit.ly/WT-ESEAP

Wiktionary

WIKIMEDIA ESEAP CONFERENCE
KOTA KINABALU 2024
COLLABORATION BEYOND THE HORIZON

Wiktionary

Wikipedia sister project for online multilingual dictionary (including words, phrases, proverbs, linguistic reconstructions, etc.)

- define words, phrases, terms, abbreviation
- thesaurus and linguistic reconstructions
- translation
- pronunciation?

a multilingual free encyclopedia

Wiktionary

['wɪkʃənrɪ] *n.*,
a wiki-based Open Content dictionary

Wikt ['wɪkt]

Writing

- Descriptive
- Informative
- Collaborative
- Licensed CC-BY-SA, free/open reuse rights

Words that can be written in Wiktionary:

- common and rare/archaic words
- new words: slang
- proven that it is recorded (attested)
- idiomatic phrase
- technical, scientific words
- word variations: old/new orthography, dialect variations

See more:

- https://id.wiktionary.org/wiki/Wikikamus:Kebijakan_inklusi

Wiktionary Advantages for Local Communities

ESEAP KK24

- Multilingual
- Easy for newbie contributor
- No mandatory for citation
- Providing discussion space for each lemma
- In incubator, easier for passing Langcom than other project
- No need special knowledge for the contributors
- Helping contributors to understand more about their language

Wiktionary (January 2024)

ESEAP KK24

Global Data

Languages (active)	Entries	Users	Active Users
168	38 mil	7,1 mil	6.812

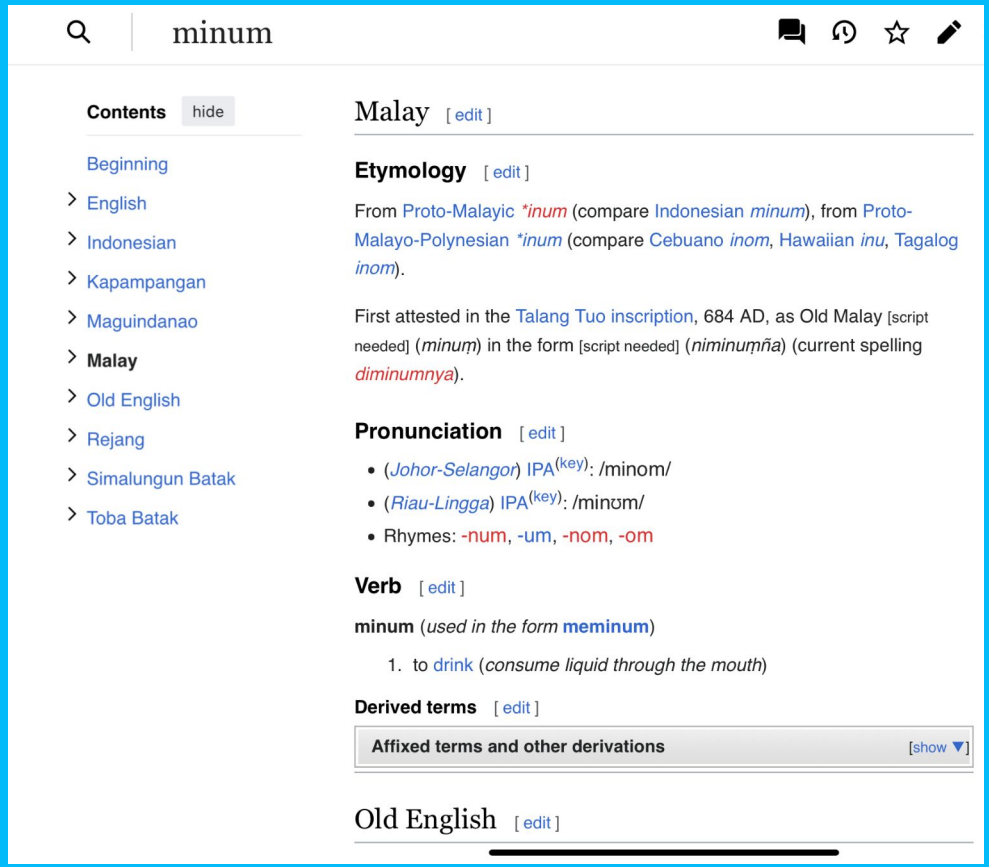
	Entries	Users	Active Users
English	8 mil	4 mil	2307

https://commons.wikimedia.org/wiki/Data:Wikipedia_statistics/data.tab

Pronunciation?

Most of lemmas in wiktionary not attaching the pronunciation

If attached, mostly use IPA.



Audio in Wiktionary

ESEAP KK24

WIKIKAMUS

https://id.wiktionary.org/wiki/kuda

Audio **makes readers easier** to know the word pronunciation

Audio **makes contributors easier** to contribute how to pronounce words

WIKTIONNAIRE

https://fr.wiktionary.org/wiki/kuda#Indon%C3%A9sien

Lingua Libre

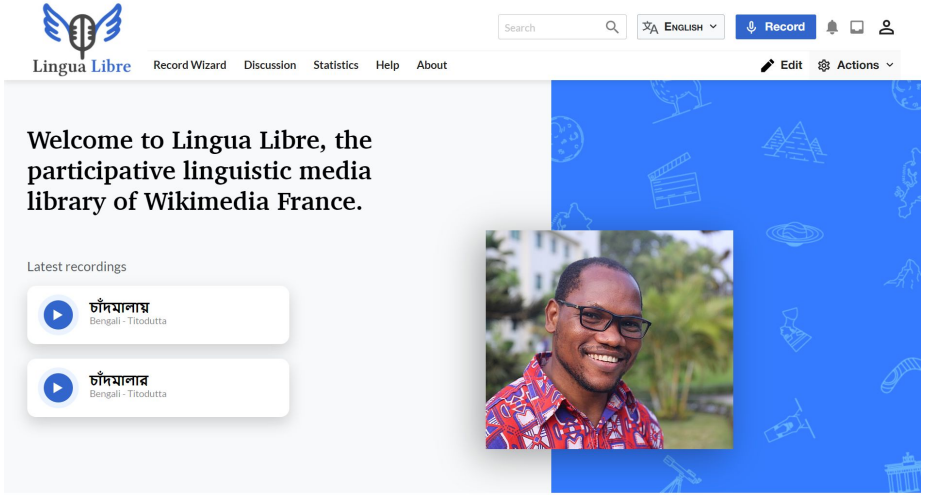
ESEAP KK24

WIKIMEDIA ESEAP CONFERENCE
KOTA KINABALU 2024
COLLABORATION BEYOND THE HORIZON

Lingua libre

Lingua libre is a collaborative project and recording tool initiated by Wikimedia France

https://lingualibre.org



Lingua Libre (April 2024)

ESEAP KK24

GLOBAL

Recordings

1 mil

Languages

200

Speakers

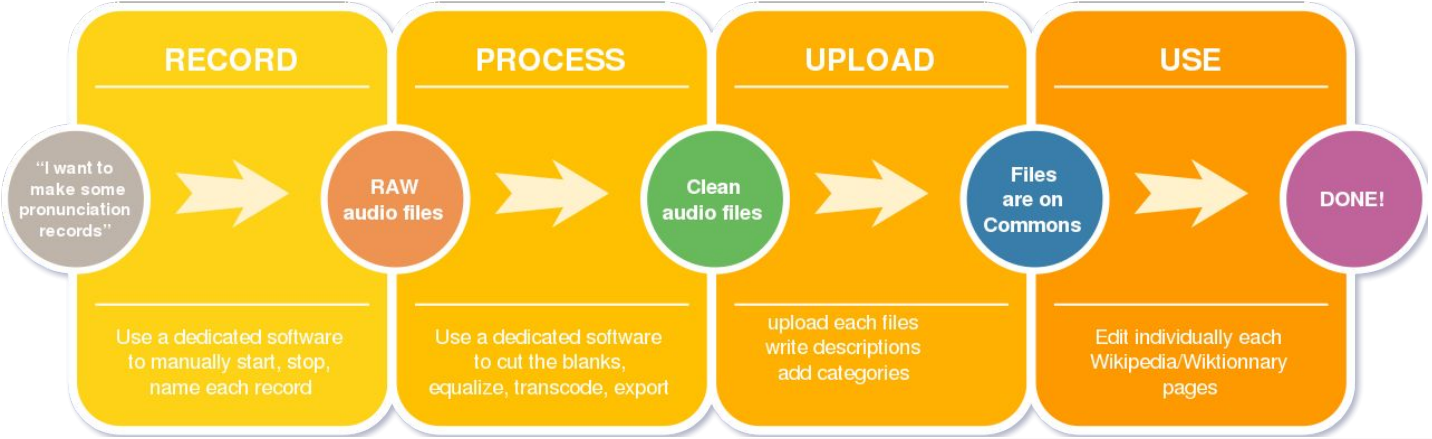
1.400

1. Central Bikol (2655)
2. Mandarin (2747)
3. Javanese (2285)
4. Sundanese (2164)
5. Indonesian (1823)
6. Bikol languages (1745)
7. Pa'O language (1716)
8. Vietnamese (1572)
9. Minangkabau (1529)
10. Balinese (1361)

11. Banjar (1355)
12. Teochew (772)
13. Mon (688)
14. Madura (672)
15. Palembang (671)

ESEAP
KK24

Why Lingua Libre?



past

present

Open license

Ringkasan [sunting]

Deskripsi	File pengucapan audio dari proyek **Lingua Libre**
Bahasa	Minangkabau
Transkripsi	kambiang
Tanggal	14 Mei 2023
Sumber	Karya sendiri
Pembuat	• Pembicara : Ardzun • Perekam : Ardzun

Lisensi [sunting]

Berkas ini dilisensikan di bawah lisensi Creative Commons Atribusi-Berbagi Serupa 4.0 Internasional.

Anda diizinkan:

- **untuk berbagi** – untuk menyalin, mendistribusikan dan memindahkan karya ini
- **untuk menggubah** – untuk mengadaptasi karya ini

Berdasarkan ketentuan berikut:

ESEAP KK24

WikiTutur

WIKIMEDIA ESEAP CONFERENCE
KOTA KINABALU 2024
COLLABORATION BEYOND THE HORIZON

WikiTutur

WikiTutur is a language preservation program by recording vocabulary pronunciation through the use of the Wiktionary and Lingua Libre. This project was previously run in Indonesia by volunteers from the Jakarta Wikimedia Community and in collaboration with other local Wikimedia communities in Indonesia.

Wiktionaries in Indonesia

ESEAP KK24

Launched

- Indonesia (155.829 lema)
- Banjar (2.212 lema)
- Gorontalo (9.815 lema)
- Javanese (61.019 lema)
- Mandailing (3.037 lema)
- Malay (16.454 lema)
- Minangkabau (11.365 lema)
- Nias (4.565 lema)
- Sundanese (3.335 lema)

Incubator

- Aceh
- Balinese
- Bugis
- Betawi

Wikikamus
Kamus bebas

Pronunciation projects in Indonesia

ESEAP KK24

WikiDialek (2020)
Minangkabau, Focused on Dialect , Video

WikiPandir (2022)
Banjar, Focused on Dialect,
Audio Lingua Libre

WikiMelayu (2023)
Malay, Culture and Language
Audio Lingua Libre

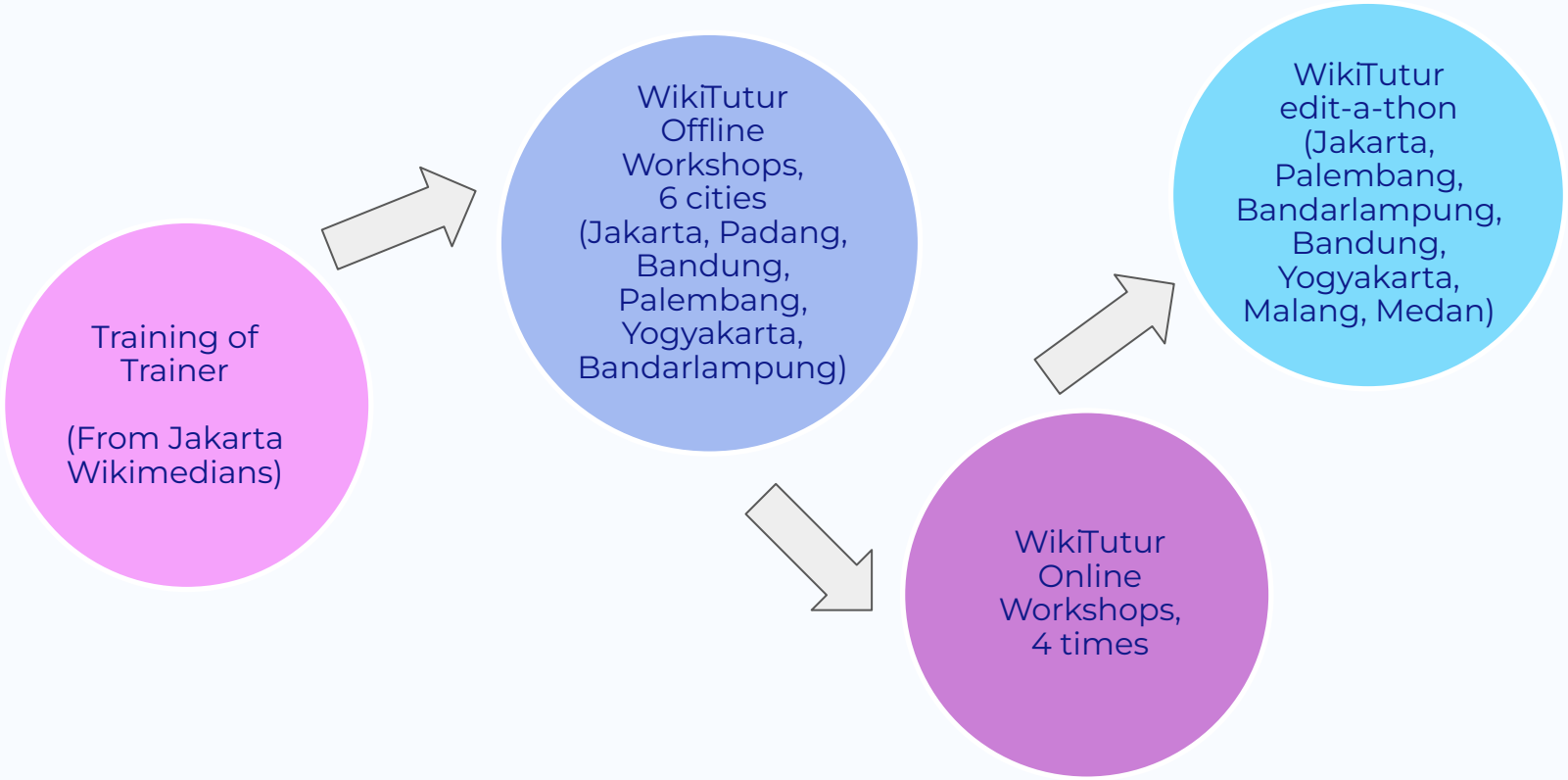
WikiSwara (2021)
Javanese, Audio

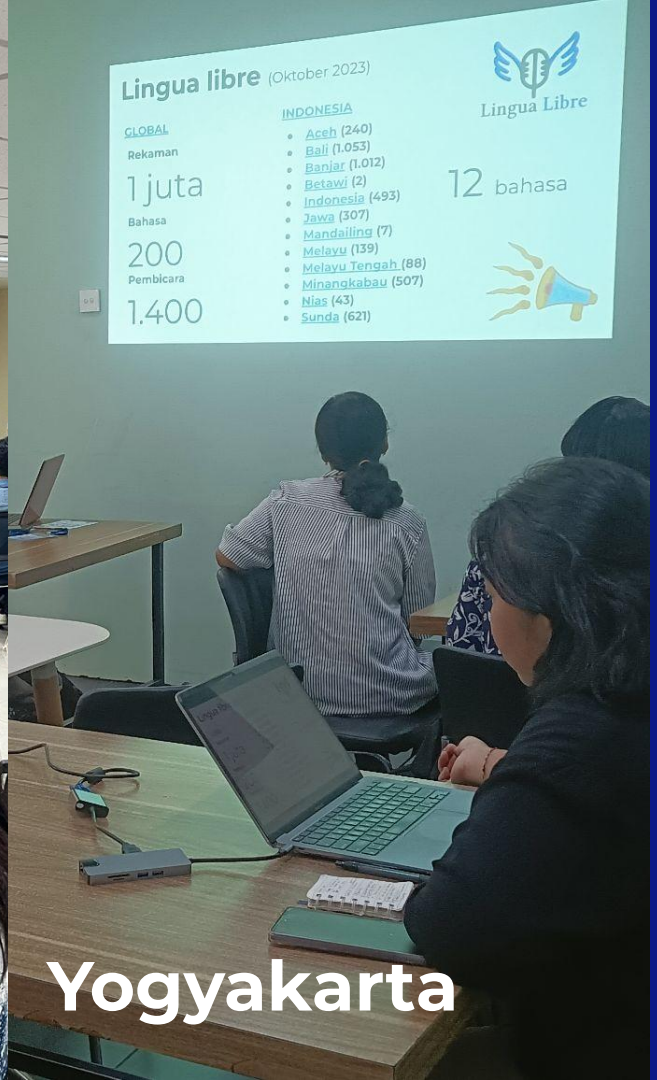
WikiSora (2022)
Sundanese,
Audio Lingua Libre

WikiTutur
(on going)

https://id.wiktionary.org/wiki/Wikikamus:ProyekWiki_WikiTutur

Timeline


Training of Trainer
(From Jakarta Wikimedians)
WikiTutur Offline Workshops, 6 cities (Jakarta, Padang, Bandung, Palembang, Yogyakarta, Bandarlampung)
WikiTutur edit-a-thon (Jakarta, Palembang, Bandarlampung, Bandung, Yogyakarta, Malang, Medan)
WikiTutur Online Workshops, 4 times



ESEAP KK24

WikiTutur Offline Outreach

Padang

Bandung

ESEAP KK24

WikiTutur Offline Outreach

ESEAP
KK24

WikiTutur Offline Outreach

Lingua Libre (Before WikiTutur)

ESEAP KK24

INDONESIA

- Aceh (240)
- Balinese (1.053)
- Banjar (1.012)
- Betawi (2)
- Indonesian (493)
- Javanese (307)
- Mandailing (7)
- Malay (139)
- Middle Malay (88)
- Minangkabau (507)
- Nias (43)
- Sundanese (621)

12

languages

Lingua Libre

Results (April 2024)

8147

Audio files

3772

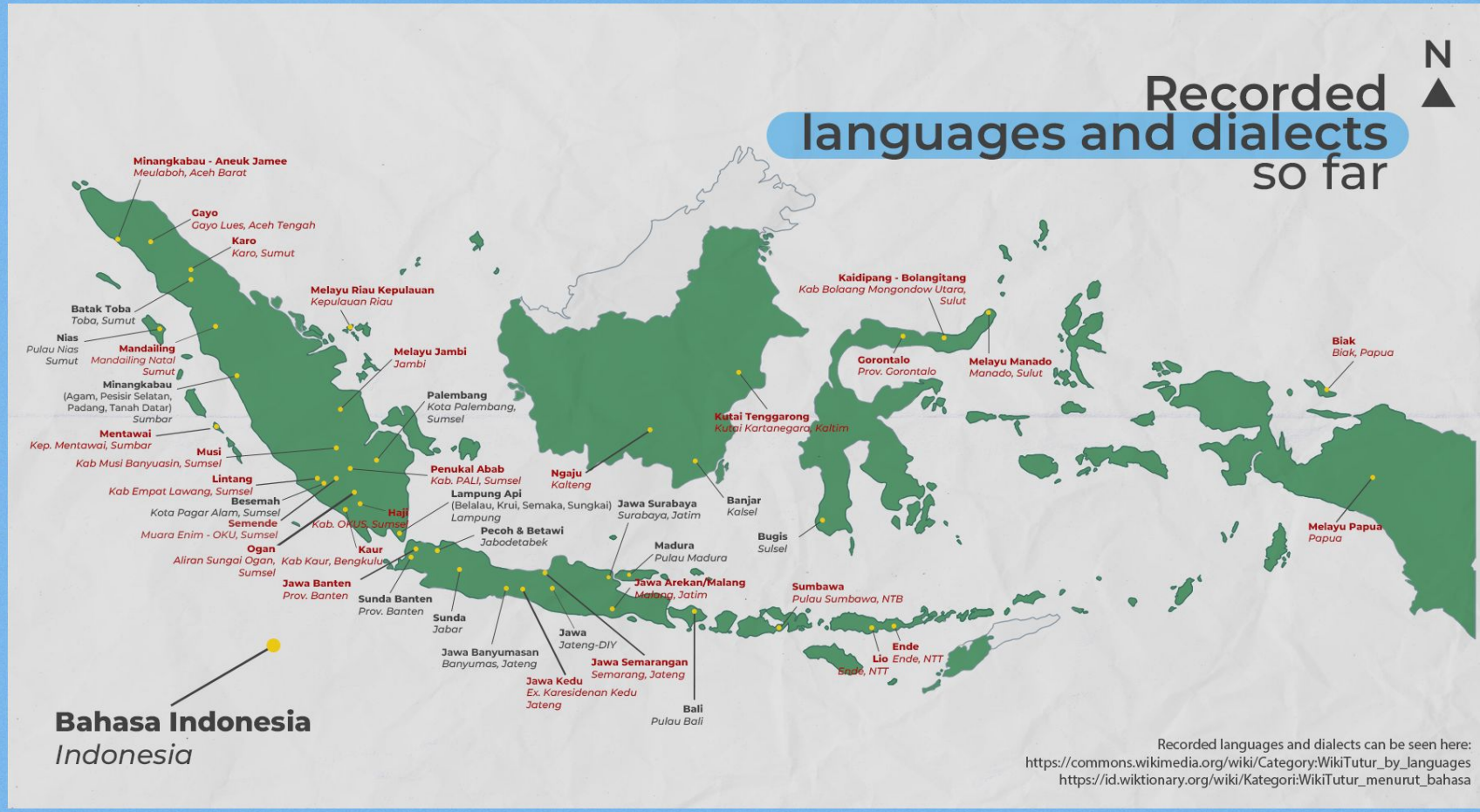
Lemmas on WikiKamus (Wiktionary in Bahasa Indonesia)

54

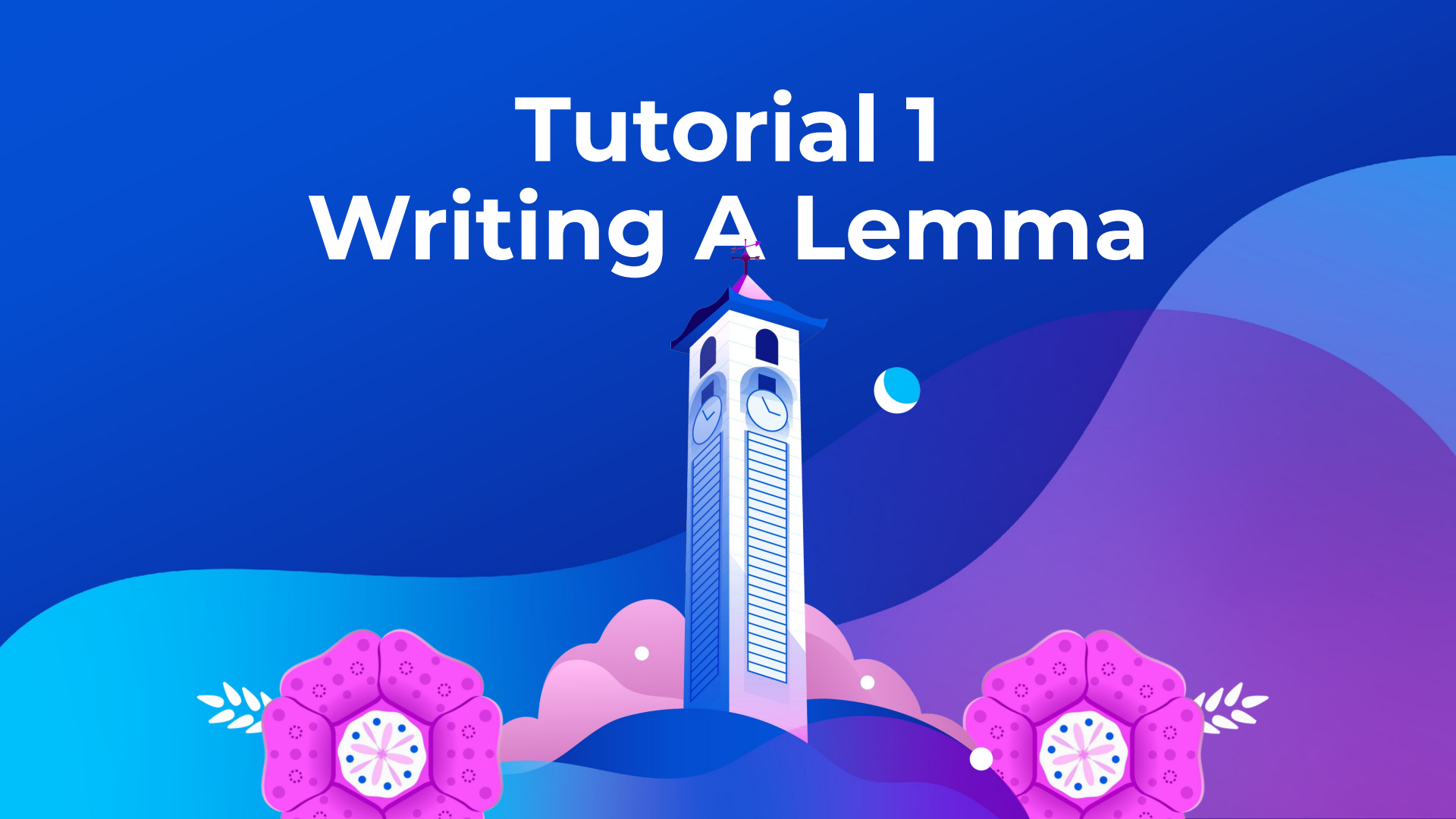
Languages/dialects

174

Contributors


N
Recorded languages and dialects so far
Minangkabau - Aneuk Jamee
Meulaboh, Aceh Barat
Gayo
Gayo Lues, Aceh Tengah
Karo
Karo, Sumut
Batak Toba
Toba, Sumut
Nias
Pulau Nias Sumut
Mandailing
Mandailing Natal Sumut
Minangkabau
(Agam, Pesisir Selatan, Padang, Tanah Datar)
Sumbar
Mentawai
Kep. Mentawai, Sumbar
Musi
Kab Musi Banyuasin, Sumsel
Lintang
Kab Empat Lawang, Sumsel
Besemah
Kota Pagar Alam, Sumsel
Semende
Muara Enim - OKU, Sumsel
Ogan
Aliran Sungai Ogan, Sumsel
Jawa Banten
Prov. Banten
Melayu Riau Kepulauan
Kepulauan Riau
Melayu Jambi
Jambi
Palembang
Kota Palembang, Sumsel
Penukal Abab
Kab. PALI, Sumsel
Haji
Kab. OKUS, Sumsel
Kaur
Kab Kaur, Bengkulu
Sunda Banten
Prov. Banten
Lampung Api
(Belalau, Krui, Semaka, Sungkai)
Lampung
Pecoh & Betawi
Jabodetabek
Sunda
Jabar
Jawa Banyumasan
Banyumas, Jateng
Jawa Kedu
Ex. Karesidenan Kedu
Jateng
Jawa
Jateng-DIY
Jawa Semarangan
Semarang, Jateng
Bali
Pulau Bali
Ngaju
Kalteng
Kutai Tenggarong
Kutai Kartanegara, Kaltim
Banjar
Kalsel
Jawa Surabaya
Surabaya, Jatim
Madura
Pulau Madura
Jawa Arekan/Malang
Malang, Jatim
Bugis
Sulsel
Kaidipang - Bolangitang
Kab Bolaang Mongondow Utara, Sulut
Gorontalo
Prov. Gorontalo
Melayu Manado
Manado, Sulut
Sumbawa
Pulau Sumbawa, NTB
Ende
Ende, NTT
Lio
Ende, NTT
Biak
Biak, Papua
Melayu Papua
Papua
Bahasa Indonesia
Indonesia
Recorded languages and dialects can be seen here:
https://commons.wikimedia.org/wiki/Category:WikiTutur_by_languages
https://id.wiktionary.org/wiki/Kategori:WikiTutur_menurut_bahasa

Tutorial 1
Writing A Lemma



Wiktionary Tutorial - New Lemma

ESEAP KK24

Difference between Wikipedia and Wiktionary in page names:

padang

Entry Discussion Citations

English [edit]

Etymology [edit]

From Malay *padang*.

Noun [edit]

padang (*plural* **padangs**)

1. Malaysian grassland

Anagrams [edit]

- padnag

The up and lower case at the beginning of word matters

Padang

3 languages

Entry Discussion Citations Read Edit View history

See also: padang

English [edit]

Etymology [edit]

From Indonesian *Padang*

English Wikipedia has an article on: Padang

Proper noun [edit]

Padang

1. The capital and largest city of West Sumatra, Indonesia

Derived terms [edit]

- Padang Besar
- Padang Panjang
- Padang Sidempuan

Translations [edit]

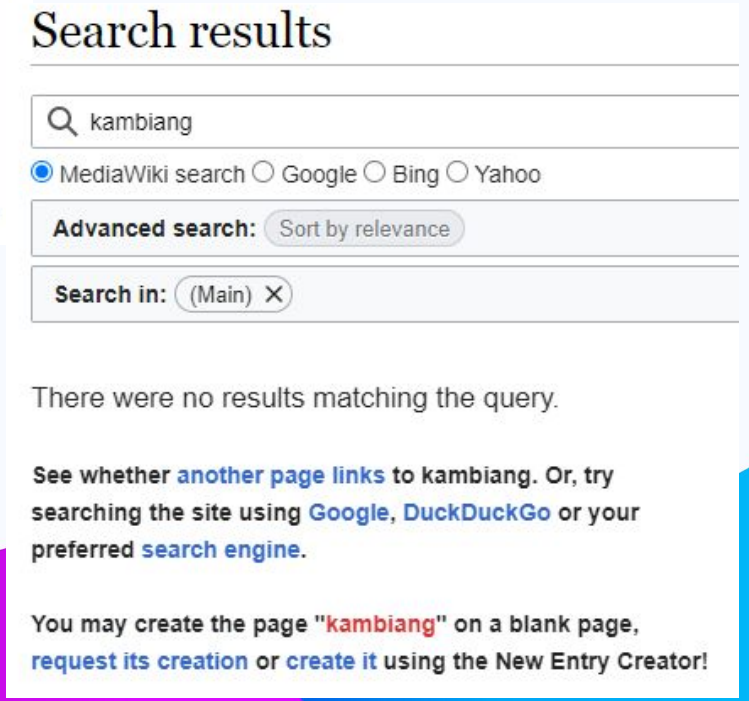
Lemma, in general, usually use *lowercase*

Entries that refer to names such as geographic location or organization can use *upper case*

Wiktionary Tutorial - New Lemma

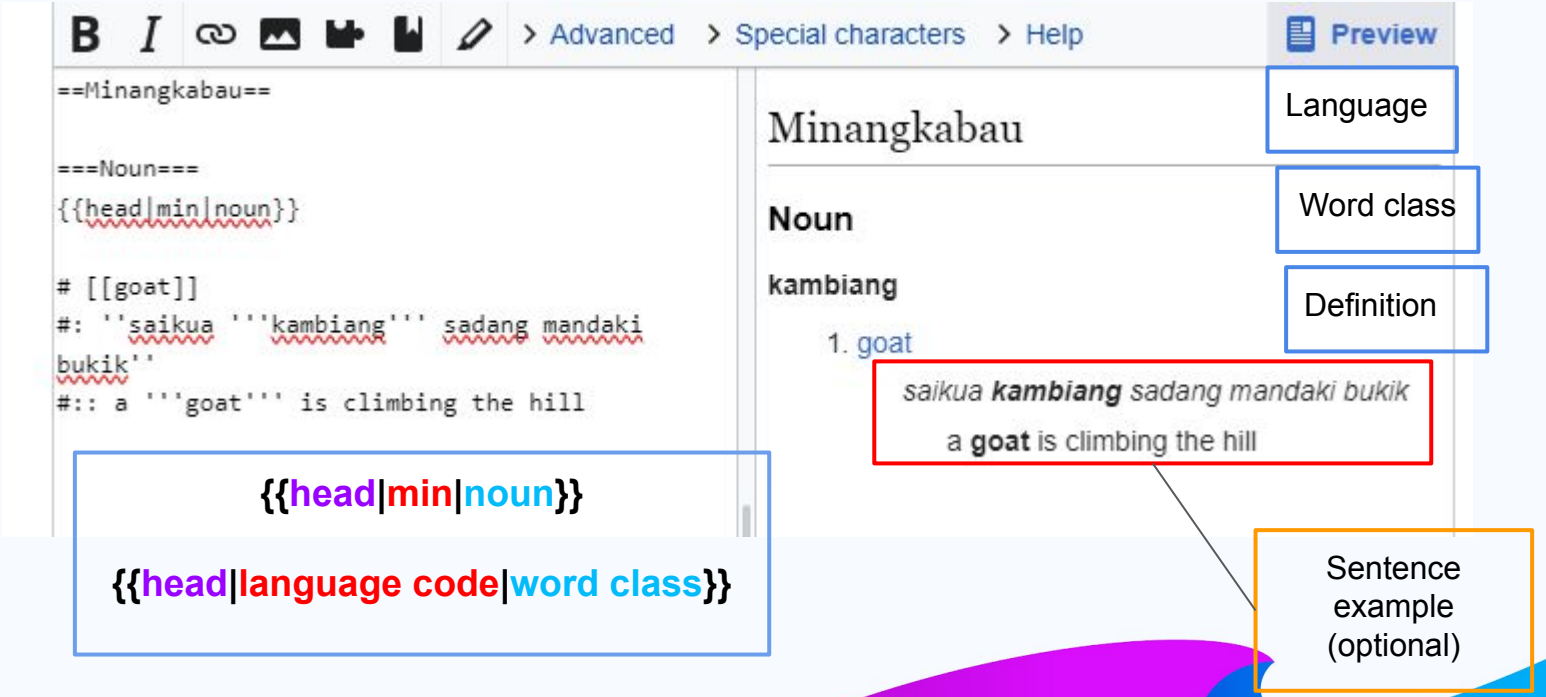
ESEAP KK24

1 Look up the word, click the word in **red**



Wiktionary Tutorial - New Lemma

ESEAP KK24



2 Input | Edit Wiktionary with **source editing**

The alternative way:
Choose the entry template based on the word class you want

See whether another page links to kambiang. Or, try searching the site using Google, DuckDuckGo or your preferred search engine.

You may create the page "kambiang" on a blank page, request its creation or create it using the New Entry Creator!

These entry templates may help when adding English words:

Basic Template with tutorial.	**Noun** Pick up that cross.	**Plural** Move those crosses here.	
Adjective He was very cross.	**Adverb** He said it very crossly.	**Comparative** She was even crosser.	**Superlative** He was the crossest.
Verb Why did he cross the road?	**3rd person** When she crosses.	**Participle** Is he crossing?	**Past** She crossed the road.

Change what is need

Creating kambiang

Entry Discussion Citations Create

Noun-specific instructions

You have selected the "Noun" preload template to help you create a new entry. This text below is what you will modify to create this Wiktionary entry.

1. Please replace `{{rfdef|en}}` below with the definition of your noun. Remember that in English, a noun is a person, a place or a thing. If you meant to enter an adjective (something that describes properties of a noun) or a verb (an action word; something you do), then you can start over here to pick a new starting template. Remember to leave the "#" at the start of the line. Please enter your definition as a full sentence, starting with a capital letter, ending with a period.
2. Click [Show preview] then [Publish page]. You can preview as many times as you like before saving.

- If you are new to Wiktionary, please see Help:Starting a new page, or use the sandbox for experiments. Also make sure your entry meets our criteria for inclusion.

B I › Advanced › Special characters › Help Preview

```
==English==

===Noun===
{{en-noun}}

# {{rfdef|en}}
```

English

Noun

kambiang (*plural* **kambiangs**)

1. *This term needs a definition. Please help out and **add a definition**, then remove the text* `{{rfdef}}`.

Define an entry

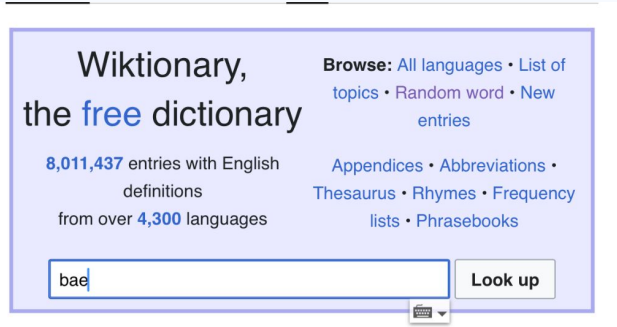
How to define a word/entry:

- Use sentence
- Use synonym
- Use antonym
- Combine the two

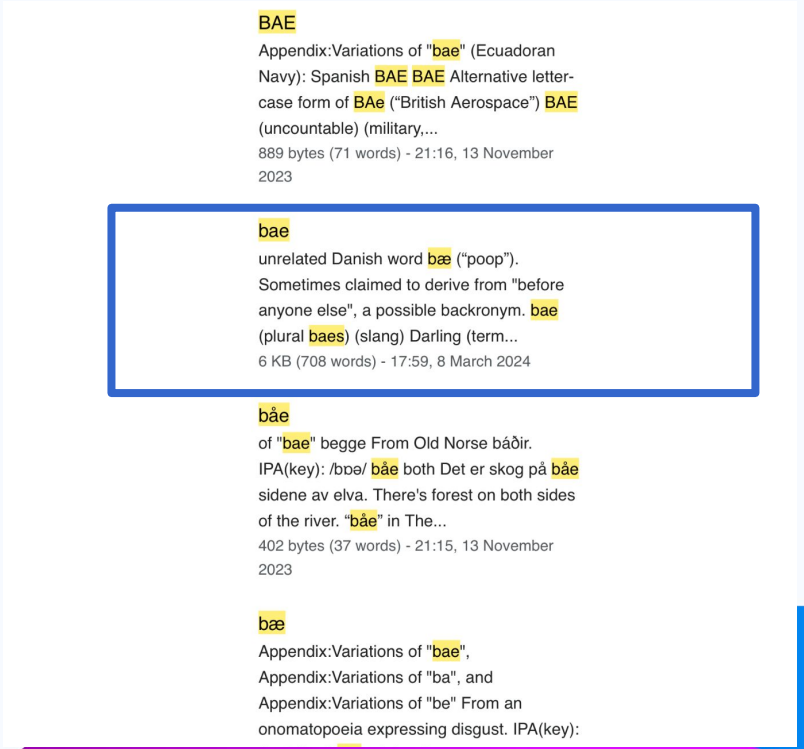
Entry in other languages:

At the minimum, translate the entries into English

Wiktionary Tutorial - Existed



1 Search for the target word and select the entry

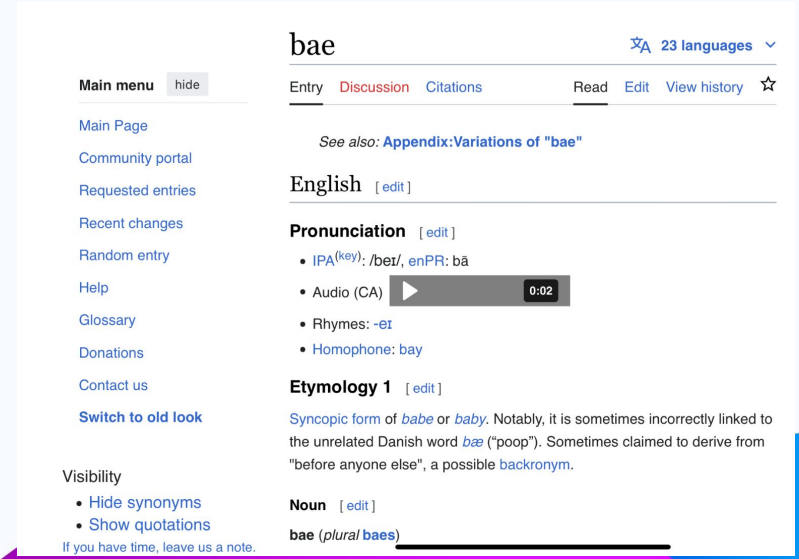


Wiktionary Tutorial - Existed

ESEAP KK24

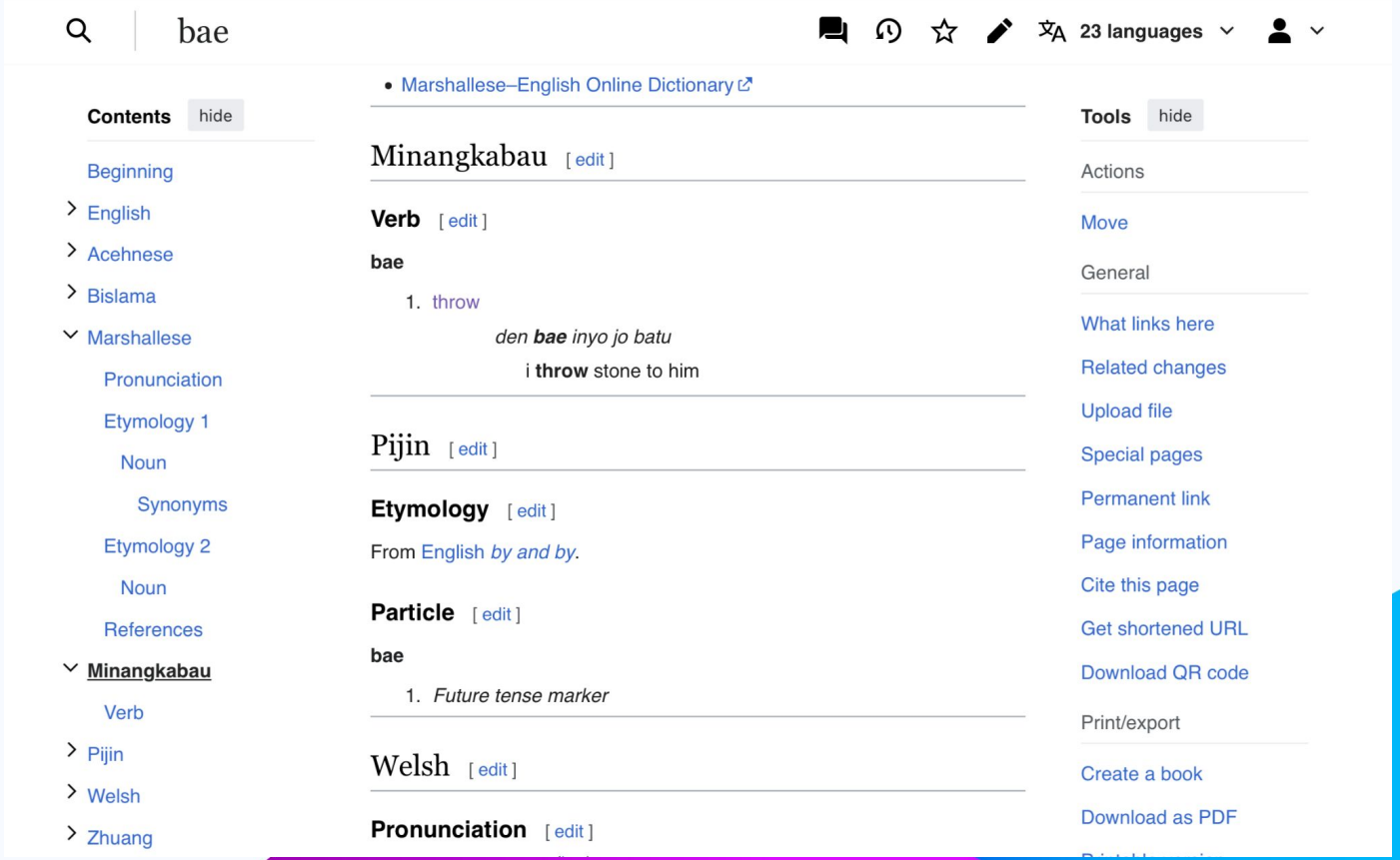
2

If you want to add a certain language entry, make sure if it already exists or not. And then click **Edit**



3

Input. According to the order of language name alphabetically. English will always be on the top (site language)



Indonesian [edit]

Etymology 1 [edit]

From Dutch *appel*, from Old Dutch **appel*, **appul*, from Proto-Germanic **aplaz*, from Proto-Indo-European **h₁ébōl* ("apple, apple-tree").

apêl

Pronunciation [edit]

- IPA(key): ['apəl]
- Rhymes: -pəl, -əl, -l

Noun [edit]

apêl (*plural* **apêl-apêl**, *first-person possessive* **apelku**, *second-person possessive* **apelmu**, *third-person possessive* **apelnya**)

1. apple, *Malus domestica*.

Alternative forms [edit]

- epal (*Brunei*, *Malaysia*, *Singapore*)

Descendants [edit]

- → Ternate: apel

Etymology 2 [edit]

From Dutch *appel* ("roll call, appeal"), previously spelled *appèl*, from French *appel*, from Old French *apeler*, from Latin *appellō*. Cognate of Afrikaans *appèl* ("roll call; appeal").

Pronunciation [edit]

- IPA(key): ['apɛl]

https://en.wiktionary.org/wiki/apel#Indonesian

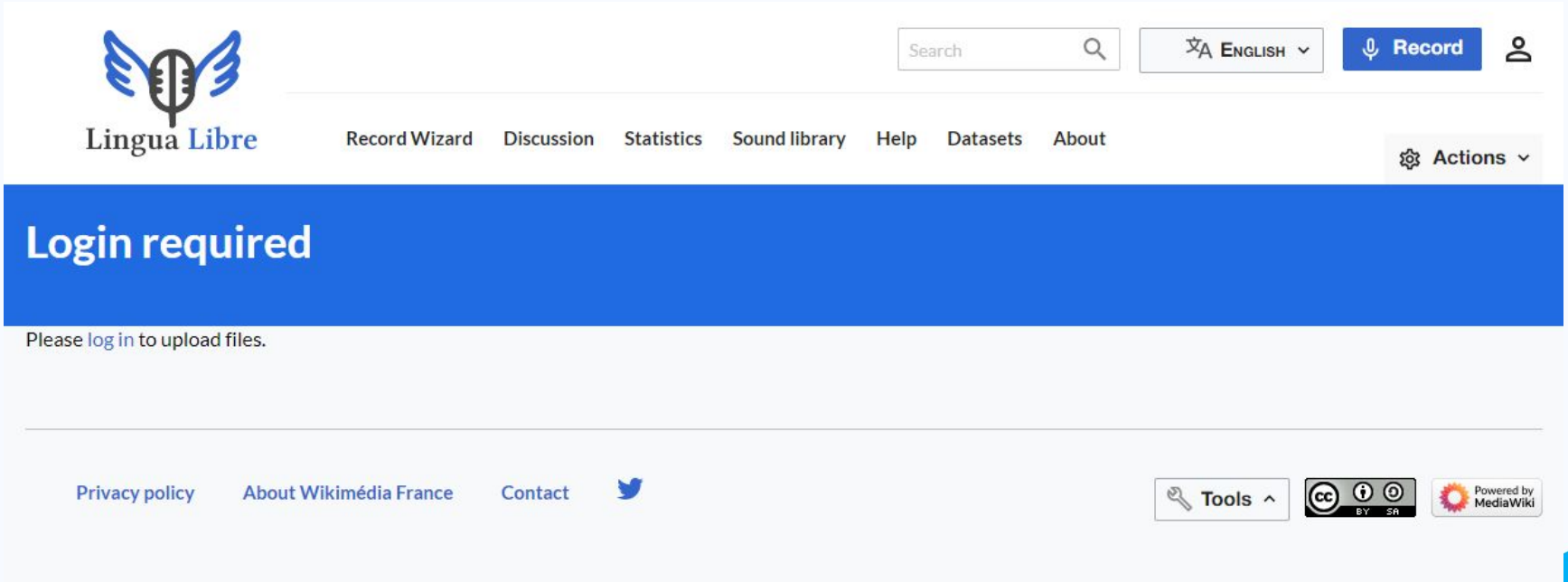
If an entry has more than 1 meaning, then:

- If it's on the **same** word class, you can just add the meaning under the same heading level
- If it's on the **same** word class, but not **homophone**, you should write the word definition **in the other heading**
- If it's on the **different** word class, just **add the new heading**

Tutorial 2
Recording with LL

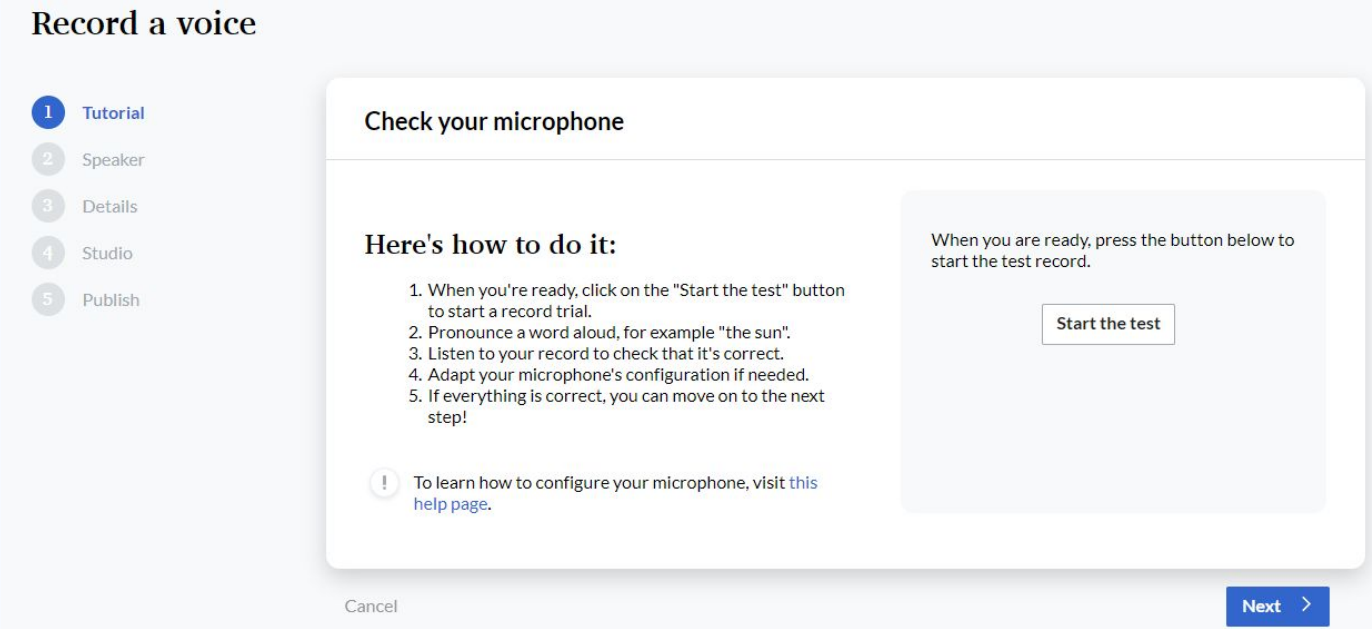
Lingua Libre tutorial

ESEAP KK24



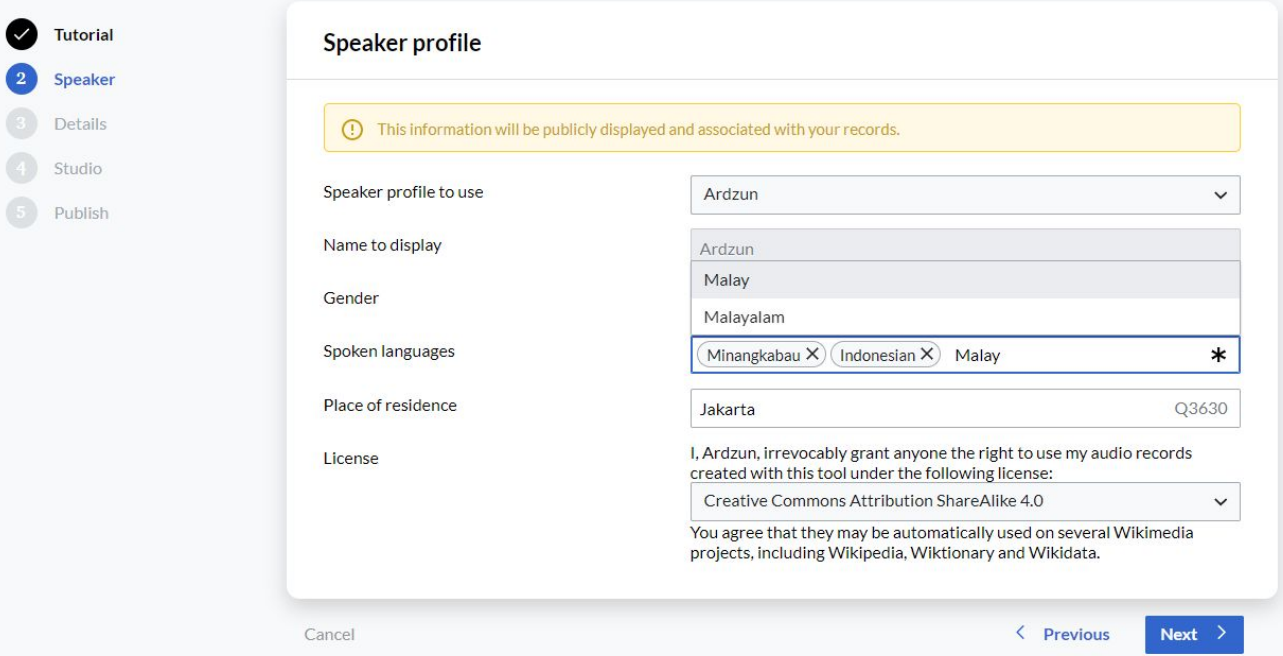
1 Login

Lingua Libre tutorial



2 Test record

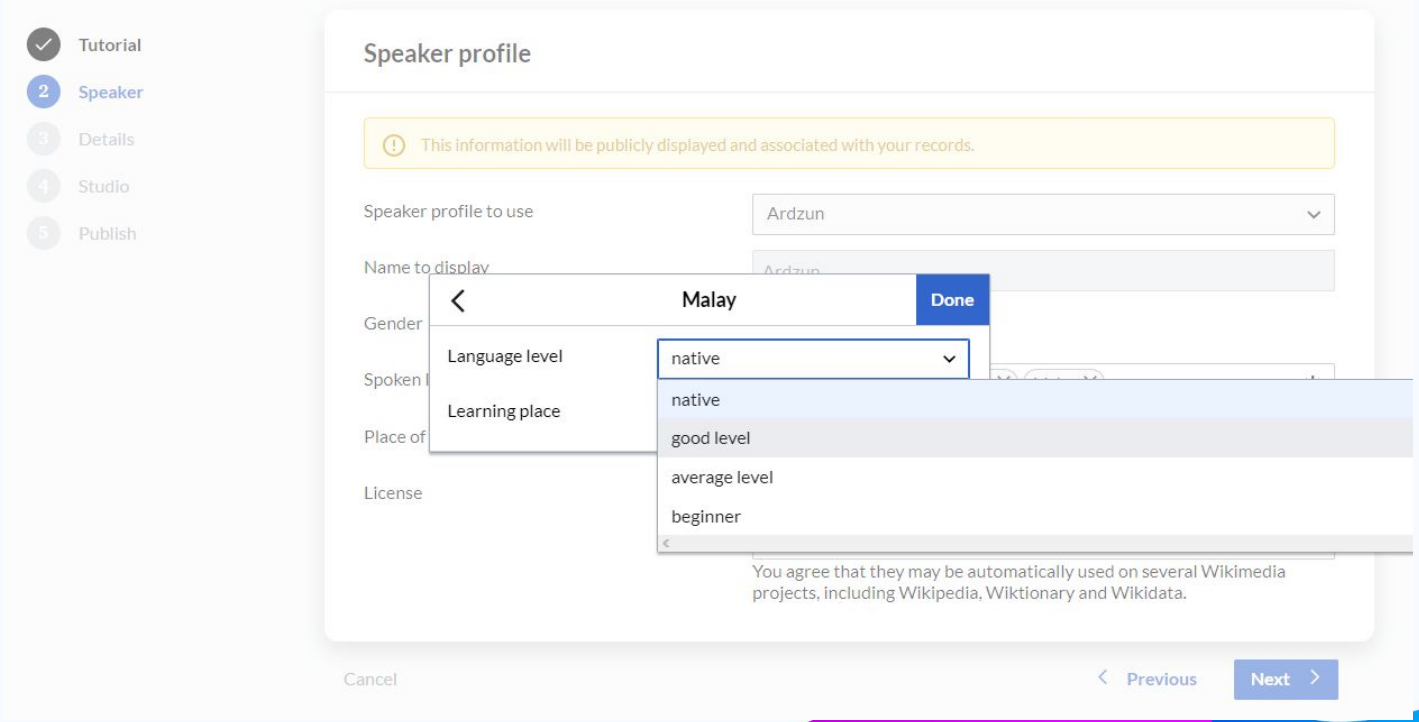
Lingua Libre tutorial



If your language not yet available, you can ask the LL administrator or you can add the language by yourself (the explanation will be on the next slide)

3 Input the speaker' data

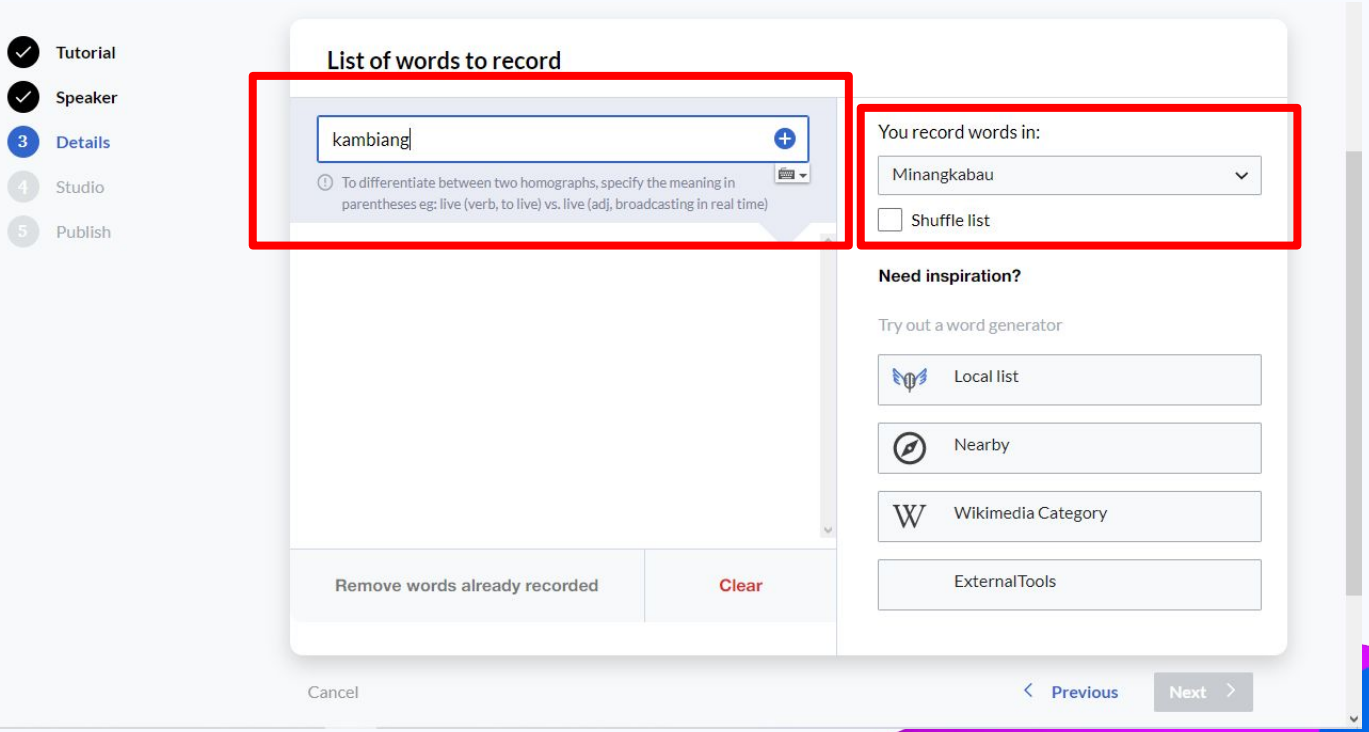
Lingua Libre tutorial



4 Input the languages level and the region where you learned the languages

Lingua Libre tutorial

ESEAP KK24

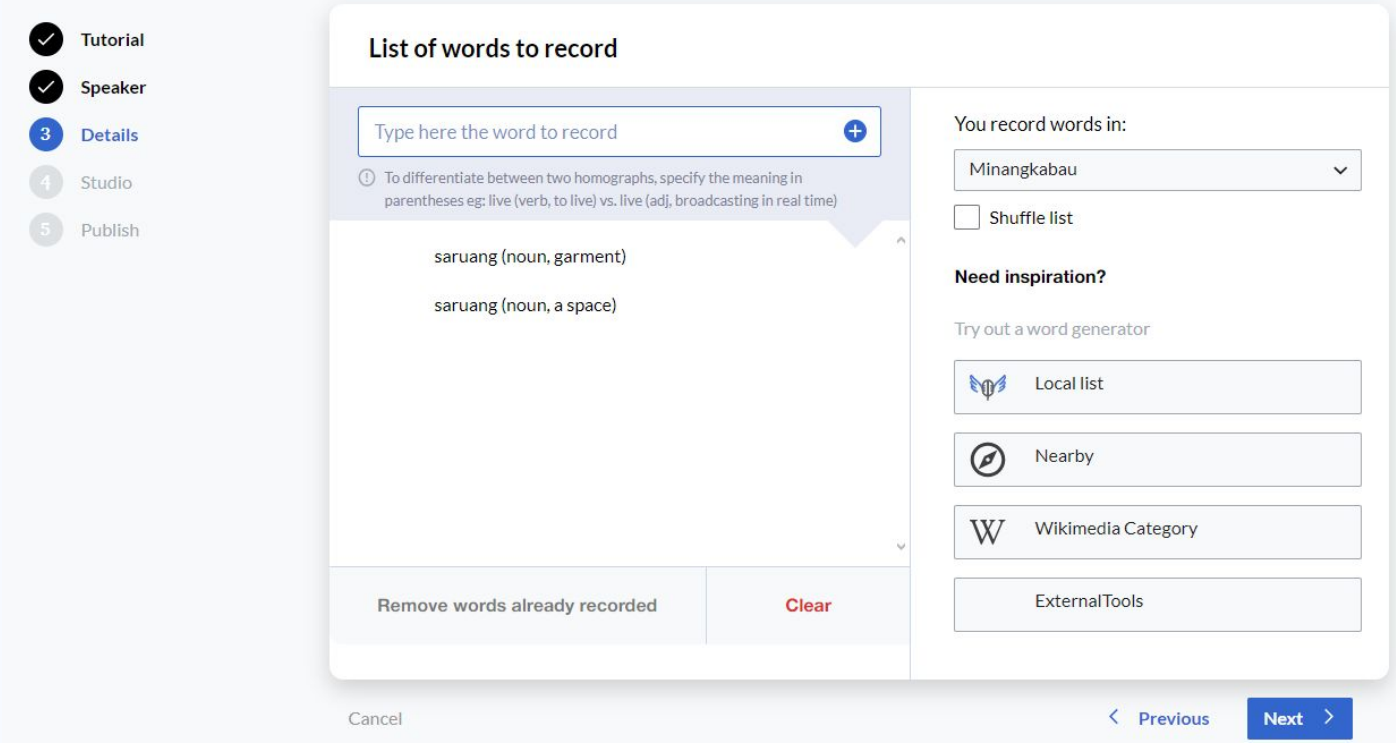


Use only lowercase, except geographical name, organization name, etc

5 List the words

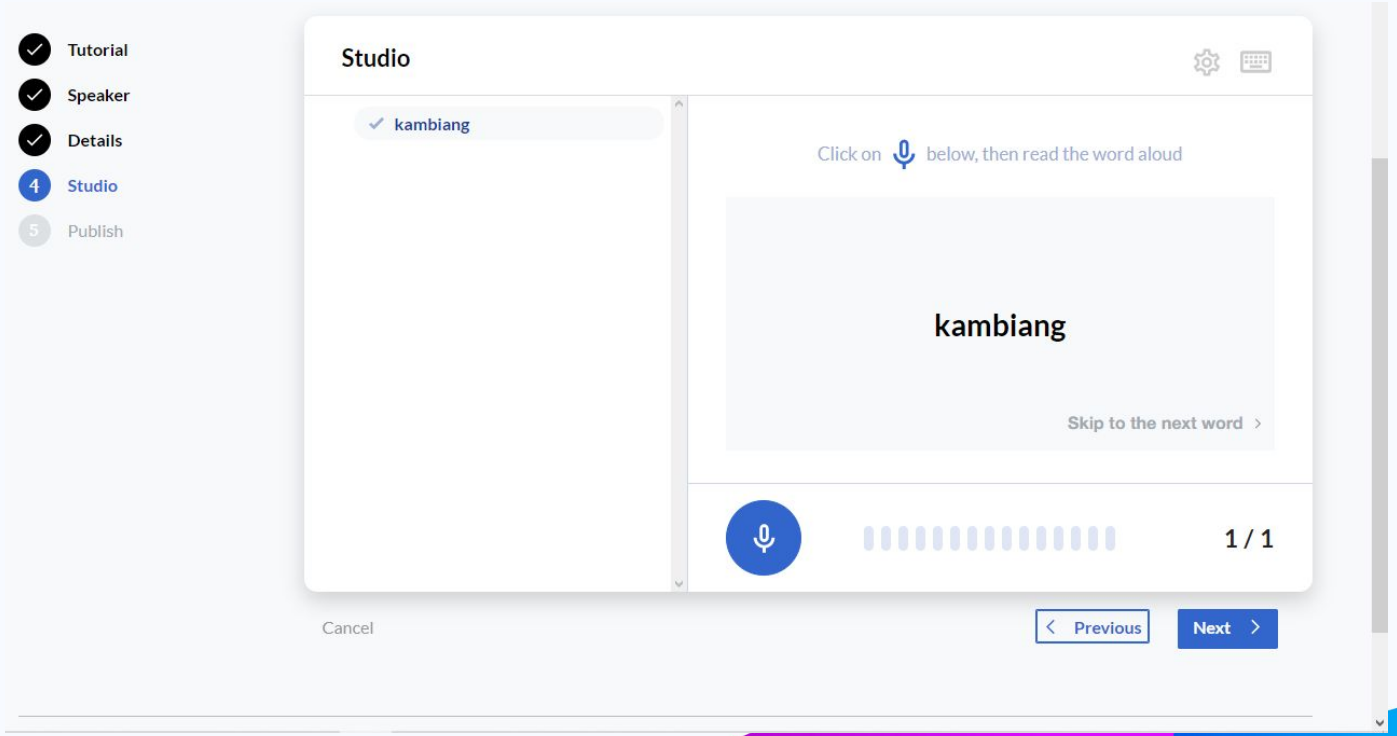
Lingua Libre tutorial

ESEAP KK24



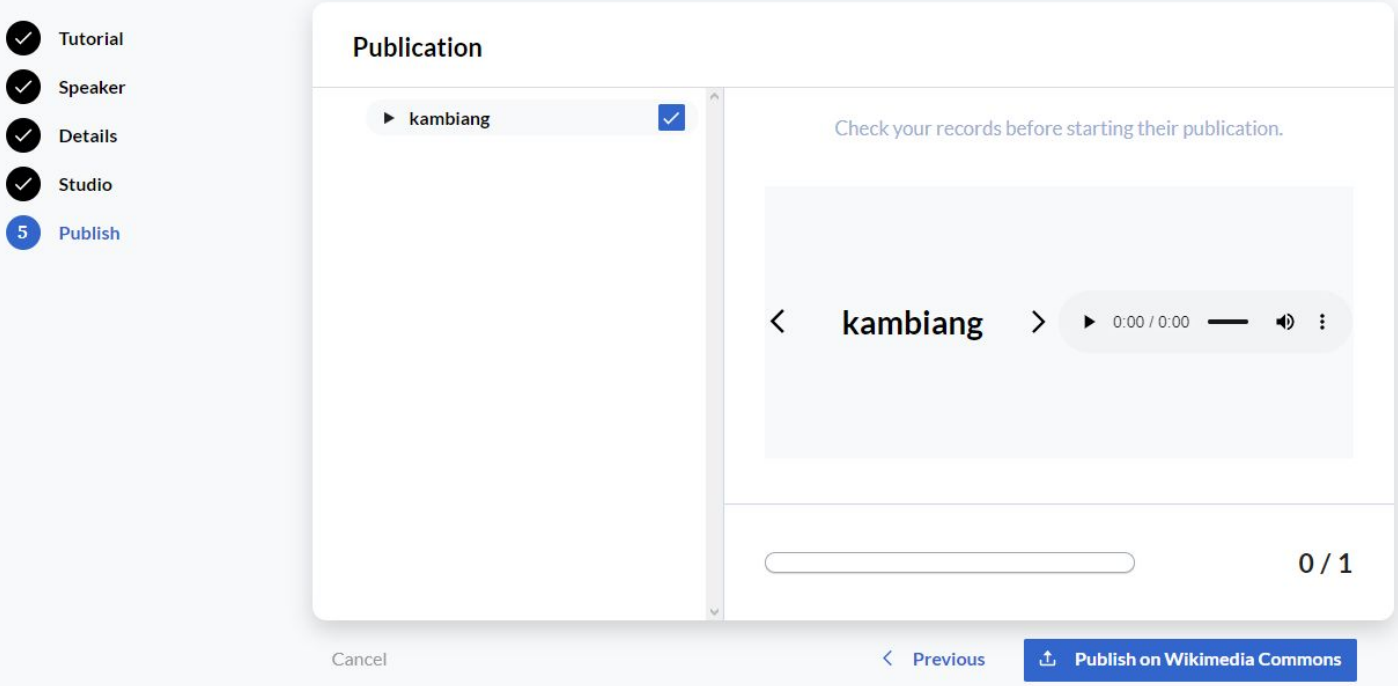
If homograph but not homophone

Lingua Libre tutorial



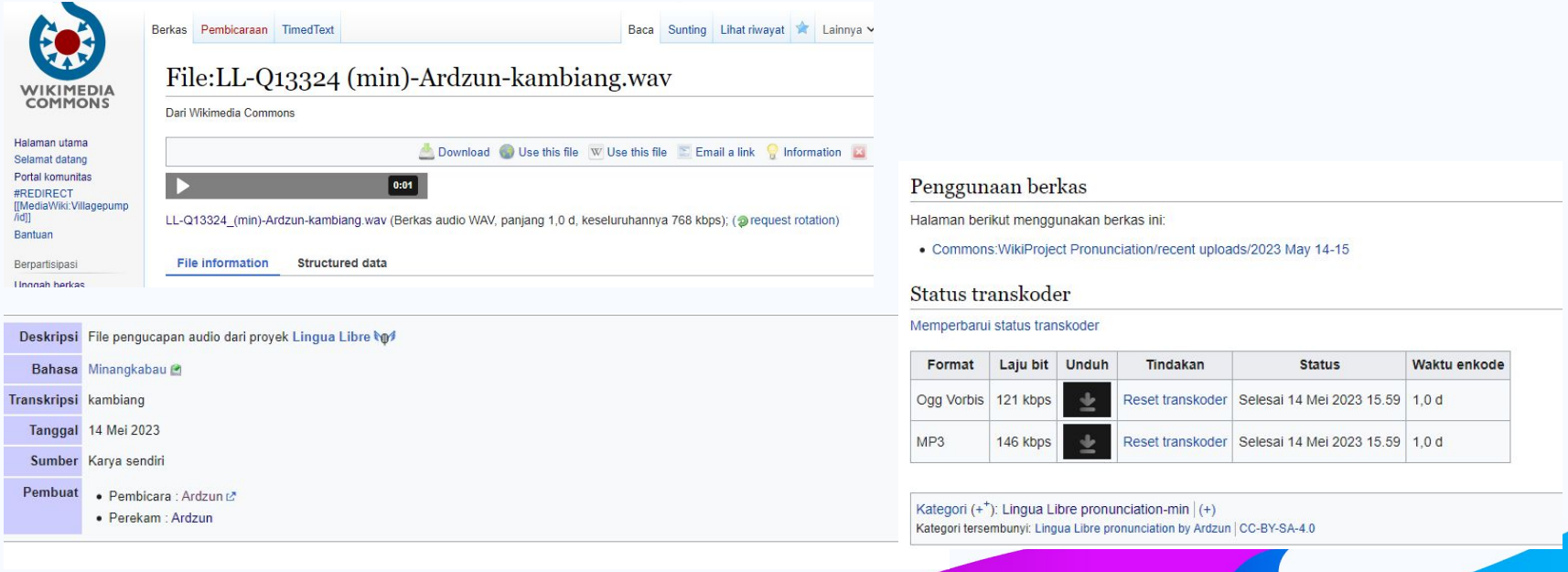
6 Start to record the words

Lingua Libre tutorial



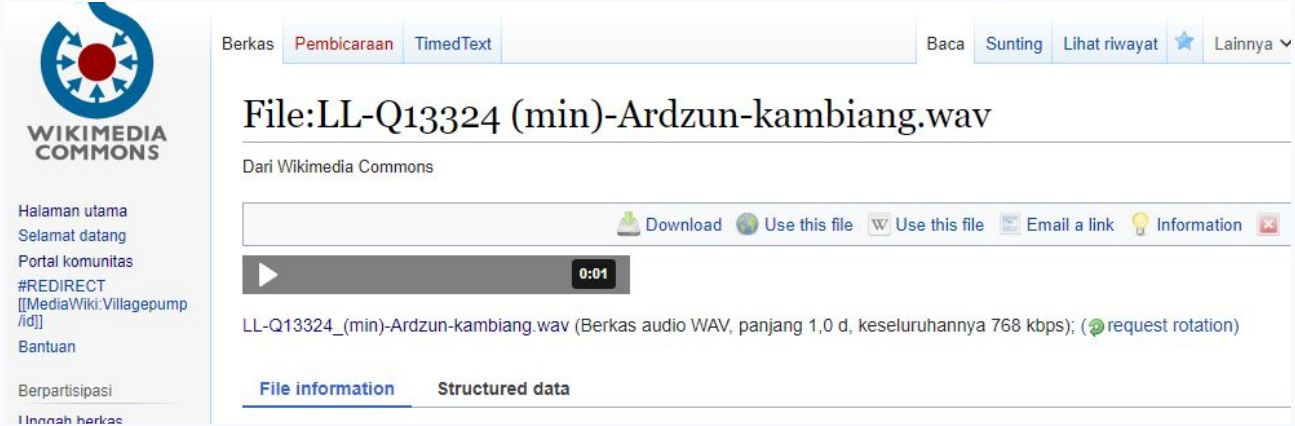
7 Review before publish

Lingua Libre tutorial



8 Audio file is ready to use!

Lingua Libre tutorial

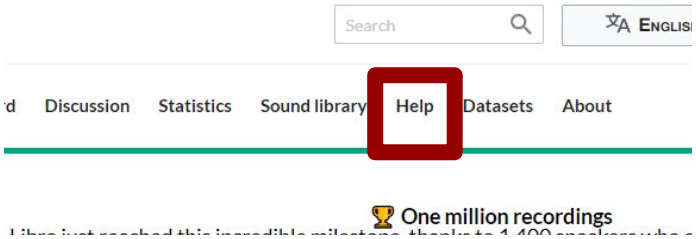


LL-*Language query* on Wikidata (language code)-user-word.wav

File name format (automatically)

Adding new languages/dialects

1

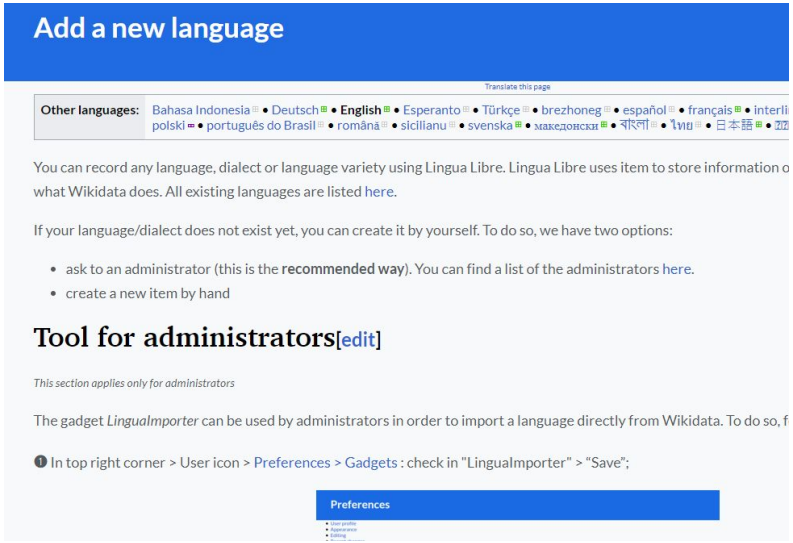


2

Record

- Your first record
- Choosing a microphone
- Configure your microphone
- RecordWizard full Manual
- Interface
- Add a new language
- Create your own lists

3



Add a new language

Other languages: Bahasa Indonesia • Deutsch • English • Esperanto • Türkçe • brezhoneg • español • français • polski • português do Brasil • română • sicilianu • svenska • македонски • বাংলা • ไทย • 日本語

You can record any language, dialect or language variety using Lingua Libre. Lingua Libre uses item to store information o
what Wikidata does. All existing languages are listed here.

If your language/dialect does not exist yet, you can create it by yourself. To do so, we have two options:

- ask to an administrator (this is the **recommended way**). You can find a list of the administrators here.
- create a new item by hand

Tool for administrators[edit]

This section applies only for administrators

The gadget *LinguaImporter* can be used by administrators in order to import a language directly from Wikidata. To do so, f

❶ In top right corner > User icon > Preferences > Gadgets : check in "LinguaImporter" > "Save";

Preferences

https://lingualibre.org/wiki/Help:Add_a_new_language

Adding new languages/dialects

Create a new language by hand[edit]

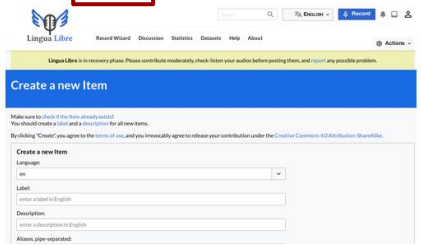
It is recommended to create a new language by asking to an administrator.

I know what I'm doing [Collapse]

So that, all labels filled on Wikidata, that contain the language name in several languages, will be imported. If you create a language by hand, you will have to translate the name of the language by yourself.

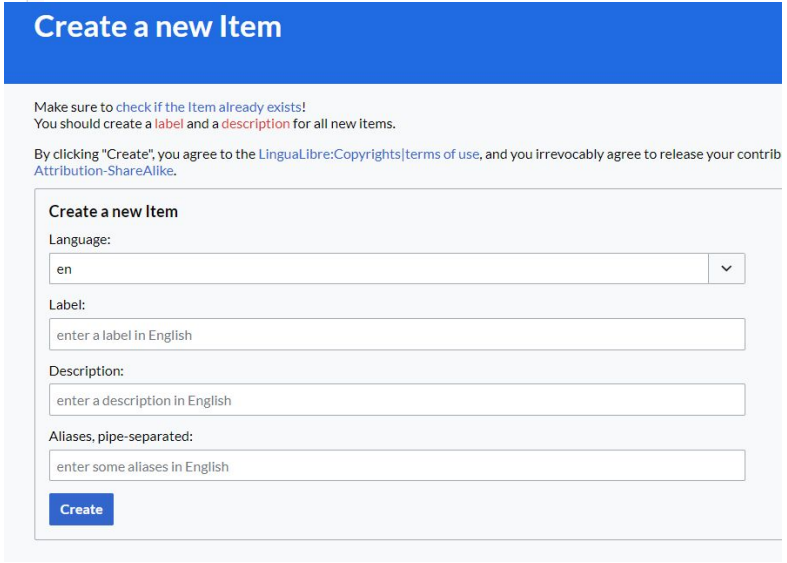
If you still want to create a new language by yourself, follow these steps:

- Click on this link in order to create a new item. You arrive on the page below.



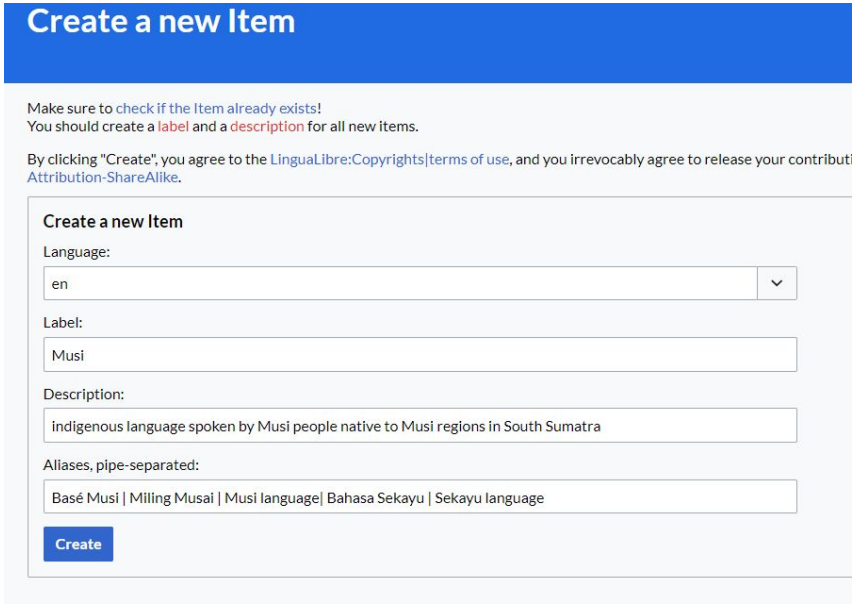
4

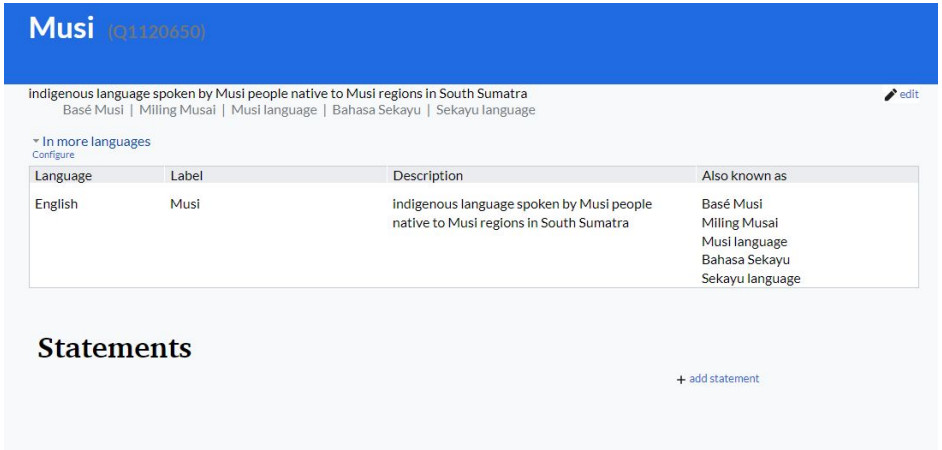
5



Adding new languages/dialects

Example: Musi language
https://lingualibre.org/wiki/Q1120650



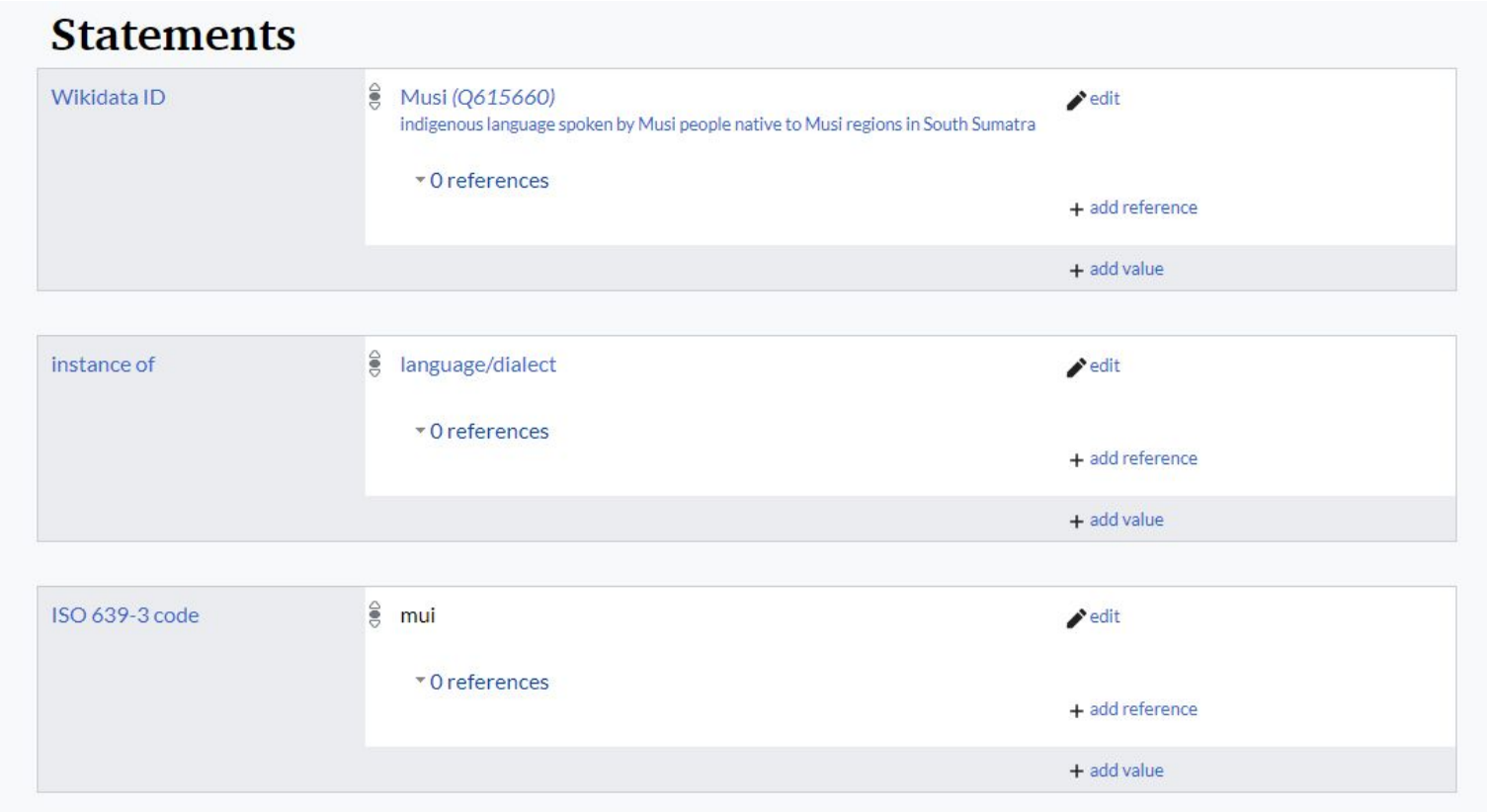


Add statement:

- instance of (P2) fill by the language/dialect (Q4)
- Wikidata ID (P12) fill by the identifier on Wikidata, start by Q…..
- ISO 639-3 code (P13) fill by the language code , exception for the dialect

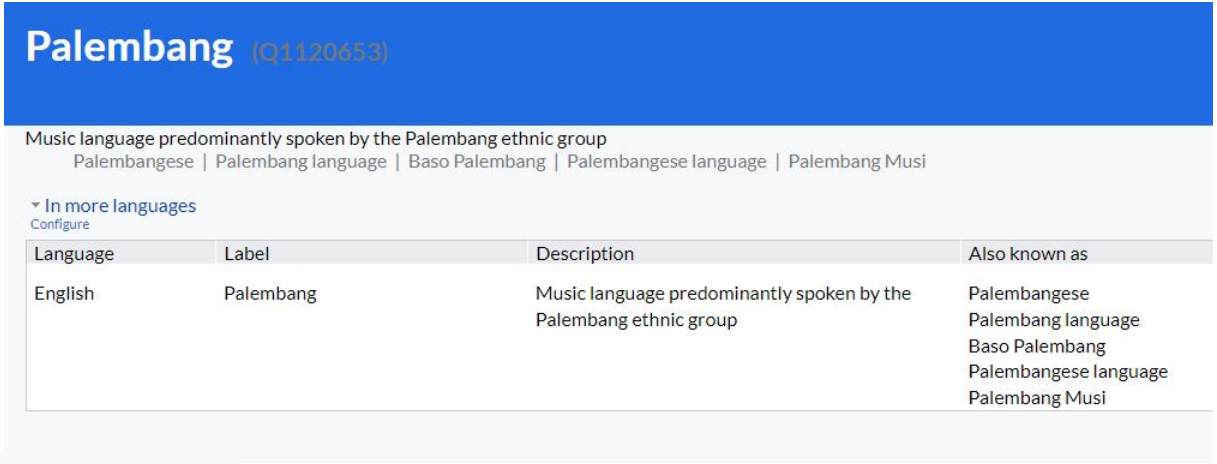
Adding new languages/dialects

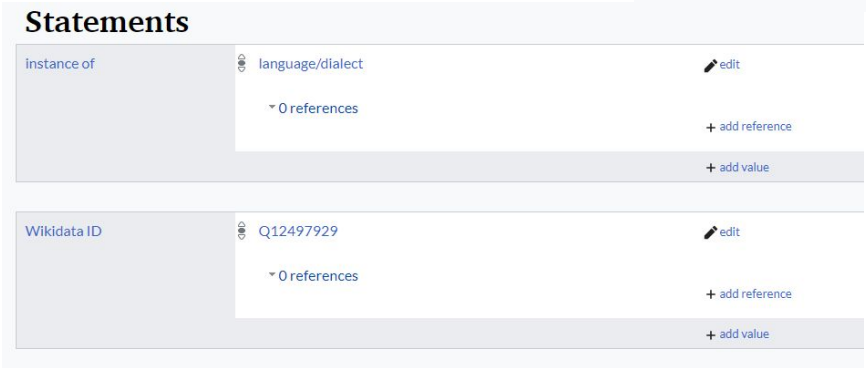
Example: Musi language in Lingua Libre



Adding new languages/dialects

Example: Palembang language https://lingualibre.org/wiki/Q1120653





Tutorial 3
Attaching The Audio

Tutorial Wiktionary - Audio

kambiang

Entry Discussion Citations

Minangkabau [edit]

Noun [edit]

kambiang

1. goat

*saikua **kambiang** sadang mandaki bukik*

a **goat** is climbing the hill

Categories: Minangkabau lemmas | Minangkabau nouns

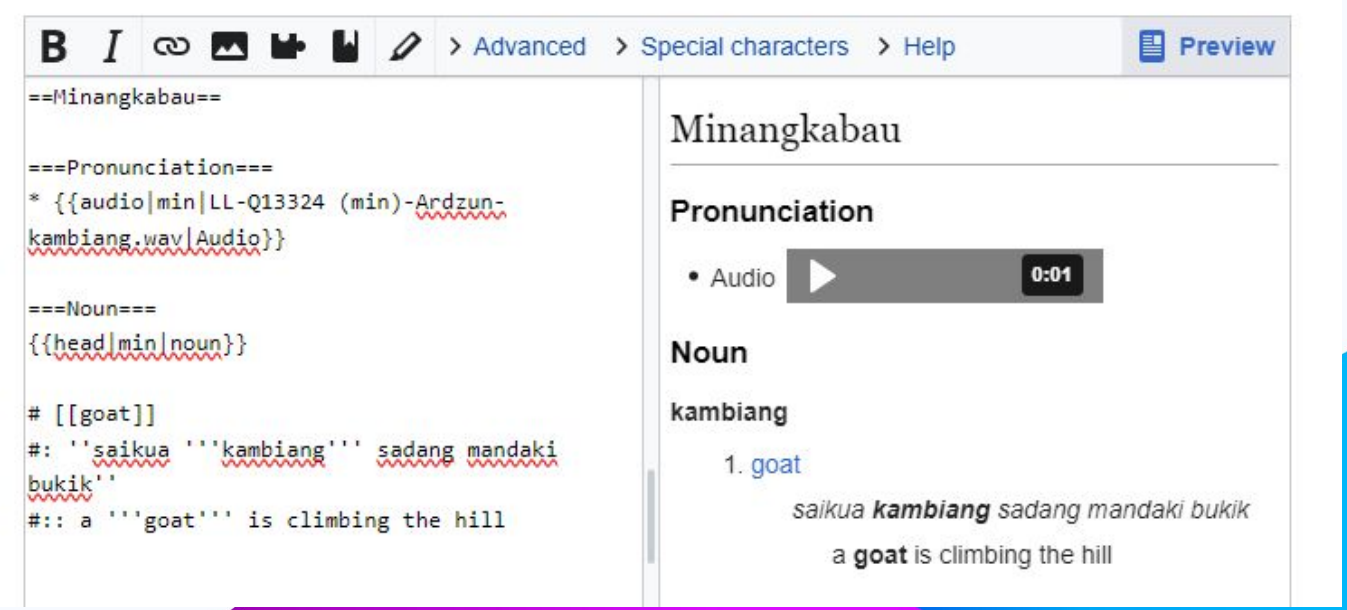
1 Click [edit]

Tutorial Wiktionary - Audio

2 Inserting the audio

===Pronunciation===

* {{audio|language code|file name|caption}}



Summary of the template

==Language ==

=== Pronunciation ===
* {{audio|language code|file name|caption}}

===Word class ===
{{head|language code|word class}}
definition in English
#: ”sentence example in your language”
#:: translation of the sentence example in English

Making new entry

ESEAP KK24

Entry: *kambiang*, in Minangkabau

```
== Minangkabau ==


=== Pronunciation ===
* {{audio|min|LL-Q13324 (min)-Ardzun-kambiang.wav|Audio}}

===Noun ===
{{head|min|noun}}
# [[goat]]
#: ''saikua '''kambiang''' sadang mandaki bukik''
#:: a '''goat''' is climbing the hill
```

Thank you!

Follow @wikijkt on Instagram

ESEAP
KK24