

LLOD-2023



Qichwabase 2.0



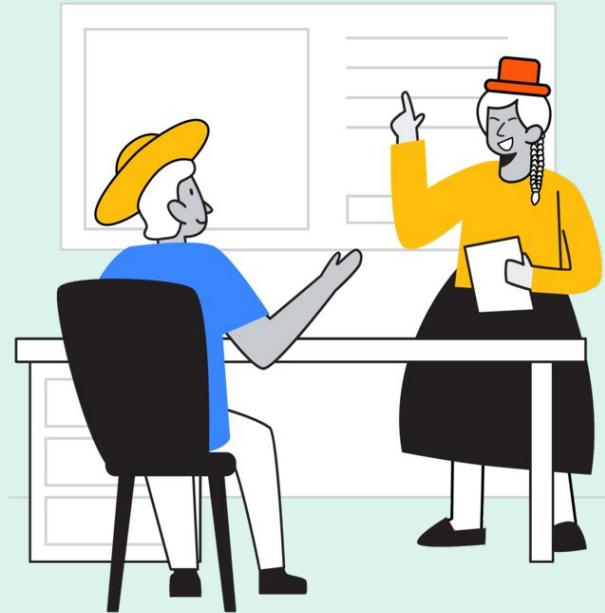
Natalia Patino & Elwin Huaman
Michael Cochez (Mentor)



June 2023

Roadmap

- 1. Ingest Language Variants**
Link variants to Wikidata
- 2. Synchronize Wikidata > Places, Variants**
Regions, cities, speakers
- 3. Register Artisans**
Quechua last names, localities, products
- 4. Develop Games**
For crowdsourcing contributions
- 5. Towards Voices**
Adding pronunciation audio to lexemes



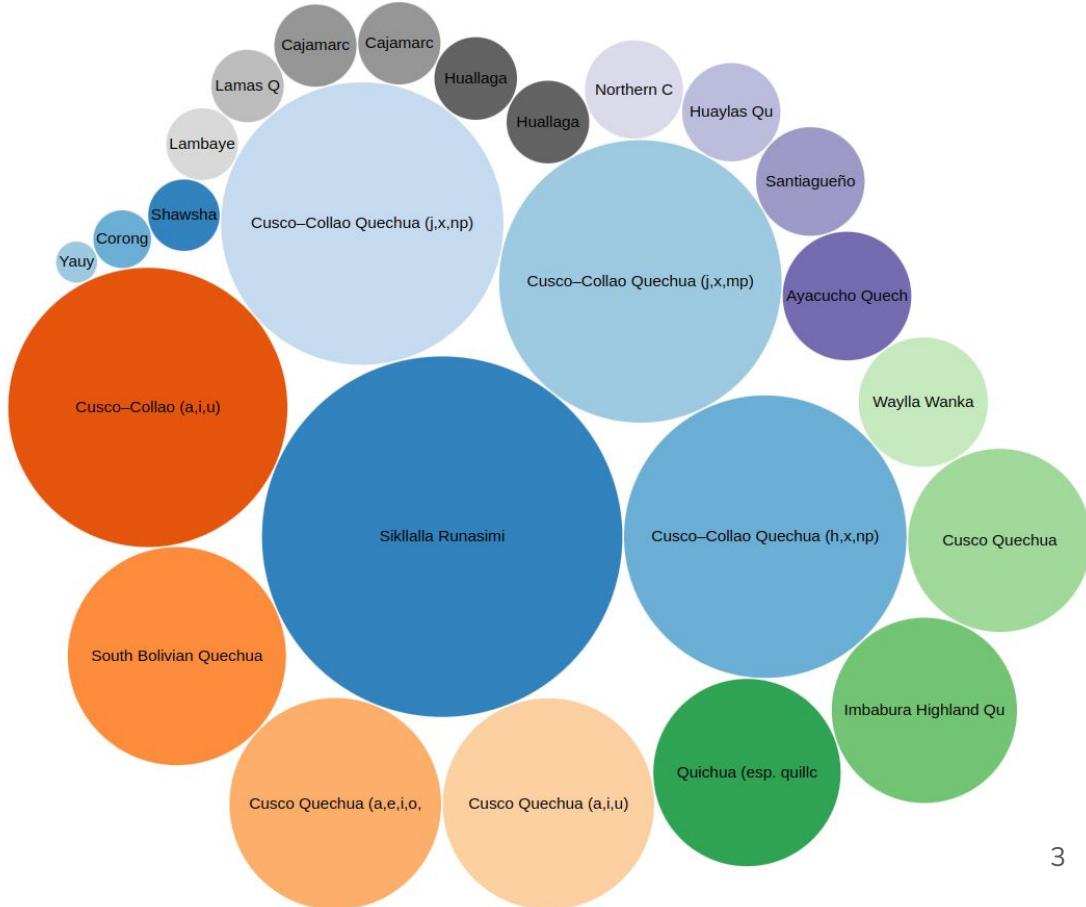
1. Ingest Language Variants : Qichwabase to Wikidata

Quechua Language Family

- 145: Sub families & dialects
- 59: Wikidata
- 25: Qichwabase

Qichwabase

- LLOD 2022
 - 553 636 triples
- Before LLOD 2023
 - 1 030 825 triples
- After LLOD 2023
 - 1 083 153 triples



1. Ingest Language Variants : Qichwabase to Wikidata

Query Service: qichwa.wikibase.cloud Examples More tools ▾ English

```
1 PREFIX qwb: <https://qichwa.wikibase.cloud/entity/>
2 PREFIX qdp: <https://qichwa.wikibase.cloud/prop/direct/>
3 PREFIX qp: <https://qichwa.wikibase.cloud/prop/>
4 PREFIX qps: <https://qichwa.wikibase.cloud/prop/statement/>
5 PREFIX qpq: <https://qichwa.wikibase.cloud/prop/qualifier/>
6 PREFIX qpr: <https://qichwa.wikibase.cloud/prop/reference/>
7 PREFIX qno: <https://qichwa.wikibase.cloud/prop/novalue/>
8
9
10 select ?lexeme ?lemma (group_concat(str(?sensegloss);SEPARATOR="; ") as ?senseglosses) ?example ?source
11
12 where {
13
14 ?lexeme a ontolex:LexicalEntry;
15   wikibase:lemma ?lemma;
16   qp:P10 [qps:P10 ?example; prov:wasDerivedFrom [qpr:P12 ?source]].
17   optional {?lexeme ontolex:sense [skos:definition ?sensegloss].}
18
19 } group by ?lexeme ?lemma ?senseglosses ?example ?source
```

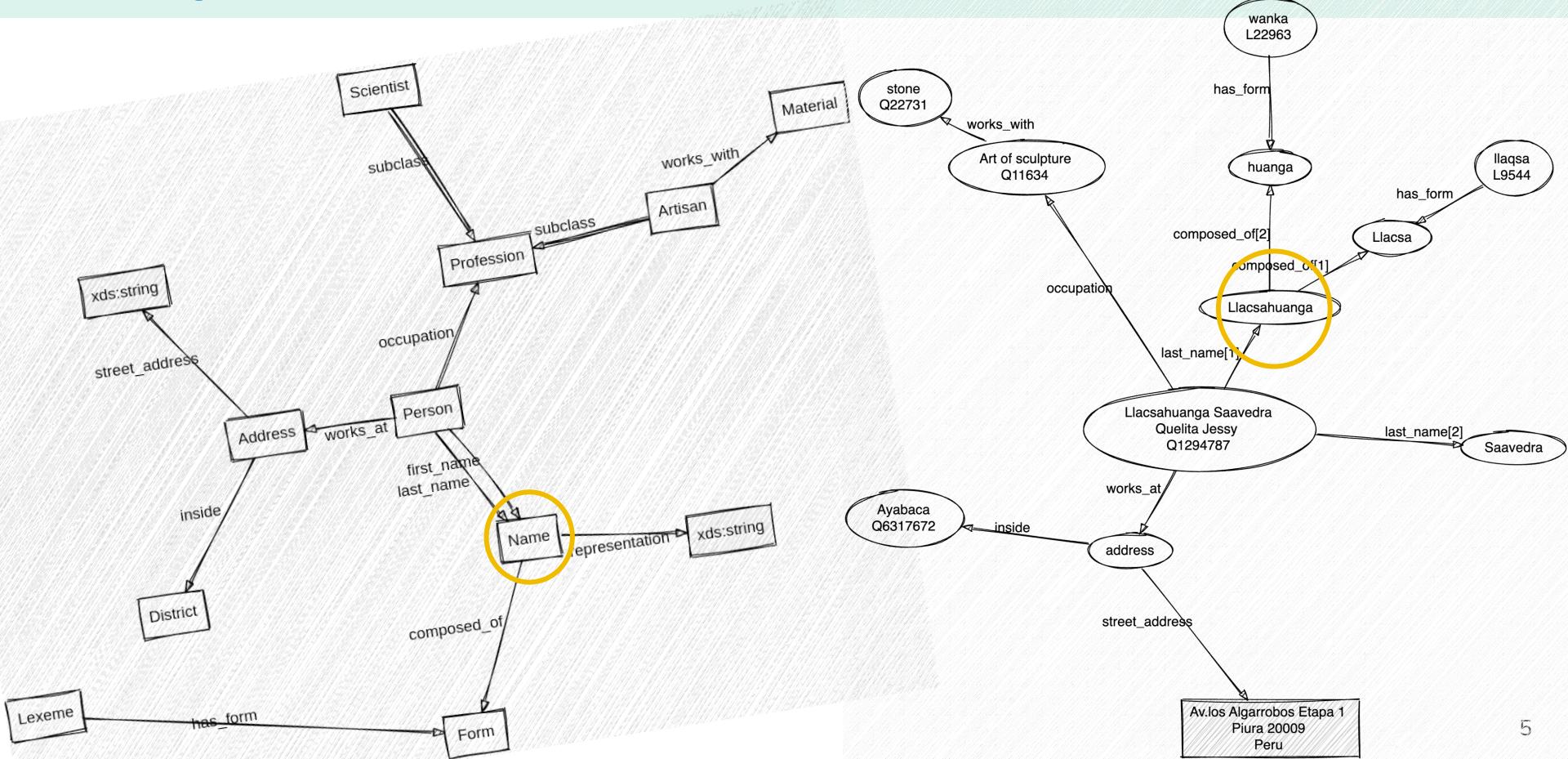
QICHWABASE 2.0

- **Statements:** 1 083 153 triples
- **Senses:** English, German, & Spanish
- **Forms:** Linking names
- **Link:** <https://qichwa.wikibase.cloud/>

Graph ▾ 3 results in 8 ms Code Download Link

lexeme	lemma	senseglosses	example	source
Q<https://qichwa.wikibase.cloud/entity/L106>	achka	viel; viele; a lot; much; many; much; mucho; mucha; muchos; muchas; harto; bastante; abundante; harto; mucho; varios; demasiado; bastante; mucho; bastante; gran cantidad; cuantioso; numeroso; gran porción; profusamente; tanto	11 ūiqin tarpu killapi 2001 p'unchawpiqa terroristakuna New York llaqtapi pitu hatun wasintintam (World Trade Center, New York Twin Towers nisqata), Washington DC llaqtapi maqanakuy ministryrip (Pentagon nisqap) huk rakinta suwakusqa antankakunawan thūñichirqan, achka waranqa runakunatam wañuchispa. Chaymantam George Walker Bush terrorismo hayu maqanakuya rimarirqan.	<https://qu.wikipedia.org/wiki/Awacha_turrikuna>
Q<https://qichwa.wikibase.cloud/entity/L251>	akllay	elección; Wahl; election; scelta; Lotterie; lottery; lotería; lotteria; elección; Auswahl; choice; elezione	Huwan hina Hisuwp'a akllay qatiqninkunapas unuchachirqanku	<https://qu.wikipedia.org/wiki/Unuchay>
Q<https://qichwa.wikibase.cloud/entity/L7724>	kunanlla kasqa	neulich; reciente; recently; recentemente	Shuk wawa kunanlla wacharishkata shitashkami shuk llakta ñanpi kay Santa Clara del Tablón llakta	<https://www.policia.gob.ec/dinapen-lrankaymanta-chapakkuna-kishpichishkami-kunanlla-wacharishkata>

1. Ingest Artisans : Modelling



2. Synchronize Wikidata : Reconciliation of places

OpenRefine Artisans Textil csv Permalink

Facet / Filter Undo / Redo 2 / 2 <

Refresh Reset all Remove all

Department of Peru: judgment change
1 choice Sort by: name count
matched 79
Facet by choice counts

Department of Peru: best candidate's score

100 — 101

Province of Peru: judgment change
2 choices Sort by: name count
matched 67
none 12
Facet by choice counts

Province of Peru: best candidate's score

100 — 101

79 rows Show as: rows records Show: 5 10 25 50 100 500 1000 rows first previous 1 next last »

Extensions Wikibase ▾

	All	qid	Department of Peru	Province of Peru	District of Peru	P30	P35	P31	P33	P34	P7
1.	Piura department		Ayabaca Province		El Carmen De La Frontera	"Q1294787"	es:"Abad Abad Maria Froliana"	33301	"Q607081"	Q2268	Q2267
2.	Piura department		Ayabaca Province		El Carmen De La Frontera	"Q1294787"	es:"Llacsahuanga Saavedra Jessy Quellin"	19702	"Q607081"	Q2268	Q2267
3.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Garcia Carrion Sheyla Isabel"	30804	"Q607081"	Q2268	Q2267
4.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Jimenez Torres Keila"	30788	"Q607081"	Q2268	Q2267
5.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Garcia Cruz Isena Yanet"	30784	"Q607081"	Q2268	Q2267
6.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Pena Garcia Gloria"	30781	"Q607081"	Q2268	Q2267
7.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Pena Garcia Ojeda Neyra"	30771	"Q607081"	Q2268	Q2267
8.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Sarmiento Esmeralda"	30769	"Q607081"	Q2268	Q2267
9.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Velasco Zurita Cristina"	30766	"Q607081"	Q2268	Q2267
10.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Garcia Guevara Maria Magdalena"	8524	"Q607081"	Q2268	Q2267
11.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Pena Garcia Gloria"	8520	"Q607081"	Q2268	Q2267
12.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Garcia Camizan"	8518	"Q607081"	Q2268	Q2267
13.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Garcia Camizan"	8517	"Q607081"	Q2268	Q2267
14.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Garcia Guevara Maria Magdalena"	8516	"Q607081"	Q2268	Q2267
15.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Pena Garcia Gloria"	8513	"Q607081"	Q2268	Q2267
16.	Piura department		Huancabamba Province		El Carmen De La Frontera	"Q1294787"	es:"Garcia Camizan"	8512	"Q607081"	Q2268	Q2267

Reconcile Start reconciling... Facets Actions Copy reconciliation data... Use values as identifiers... Add entity identifiers column...

6

2. Synchronize Wikidata : Resolving conflicting values

OpenRefine Artisans Textil csv Permalink

Facet / Filter Undo / Redo 8 / 8 <

79 rows Show as: rows records Show: 5 10 25 50 100 500 1000 rows < first < previous 1 next > last >

Extensions Wikibase ▾

All	qid	Department of Peru	Province of Peru	District of Peru	P30	P35	P31	P33	P34	P7
		Choose new match	Choose new match	Choose new match	"Q1294787"	es:"Pillco Quispe Kenely Esther"	25666	"Q607081"	Q2268	Q2267
	66.	Piura department Choose new match	Talara Province Choose new match	Los Organos District Choose new match	"Q1294787"	es:"Rosales Vinces Lucero Stefany"	43583	"Q607081"	Q2268	Q2267
	67.	Piura department Choose new match	Talara Province Choose new match	Mancora <input checked="" type="checkbox"/> Mancora District (61) <input checked="" type="checkbox"/> Marcavelica District (31) <input checked="" type="checkbox"/> Create new item	"Q1294787"	es:"Rosales Vinces Lucero Stefany"	43583	"Q607081"	Q2268	Q2267
	68.	Piura department Choose new match	Talara Province Choose new match	Mancora <input checked="" type="checkbox"/> Mancora District (61) <input checked="" type="checkbox"/> Marcavelica District (31) <input checked="" type="checkbox"/> Create new item	"Q1294787"	es:"Rosales Vinces Lucero Stefany"	43583	"Q607081"	Q2268	Q2267
	69.	Piura department Choose new match	Talara Province Choose new match	Mancora <input checked="" type="checkbox"/> Mancora District (61) <input checked="" type="checkbox"/> Marcavelica District (31) <input checked="" type="checkbox"/> Create new item	"Q1294787"	es:"Rosales Vinces Lucero Stefany"	43583	"Q607081"	Q2268	Q2267

1 choice Sort by: name count
matched 79
Facet by choice counts

1 choice Sort by: best candidate's score
100 — 101

1 choice Sort by: name count
matched 79
Facet by choice counts

1 choice Sort by: best candidate's score

Match this cell Match all identical cells Cancel

Mancora District (Q6949621)
district in Piura, Peru



3. Register Artisans : Modelling and importing artisans data

Artisan	
(WD:Q1294787)	
Label (qu)	
work location	(P36)
occupation	(P30)
name	(P35)
RNA certificate	(P31)
field of work	(P33)
technique	(P34)
described in	(P7)
given name	(P37)
surname	(P38)

QICHWABASE Search Qichwabase

Navigation Wiki tools Page tools Other projects

Llacsahuanga Saavedra Jessy Quelita (Q2271)

Item Discussion History

Artisan from Piura, Peru

given name Jessy series ordinal 1 edit
Quelita series ordinal 2 edit

surname Llacsahuanga series ordinal 1 edit
Saavedra series ordinal 2 edit

Try it: <https://qichwa.wikibase.cloud/wiki/Item:Q2271> + add statement

Results:

- 3 875 out of 55 317 (7%)

3. Register Artisans : Splitting and linking names

QICHWABASE Search Qichwabase

Navigation ▾ Wiki tools ▾

Llacsahuanga (Q2353)

Item Discussion ★

last name of a person

In more languages

Statements

instance of	surname
-------------	---------

composed of	llacs huanga
-------------	-----------------

QICHWABASE Search Qichwabase

Navigation ▾ Wiki tools ▾ Page tools ▾ Other projects ▾

Llacsahuanga Saavedra Jessy Quelita (Q2271)

Item Discussion ★ History

Artisan from Piura, Peru

given name	Jessy	edit
	series ordinal	1
Quelita	edit	
	series ordinal	2

surname	Llacsahuanga	edit
	series ordinal	1
Saavedra	edit	
	series ordinal	2

Try it:
<https://qichwa.wikibase.cloud/wiki/Item:Q2271>

+ add statement

3. Register Artisans : Linking forms to lexemes

QICHWABASE Search Qichwabase

Navigation ▾ Wiki tools ▾

Llacsahuanga (Q2353)

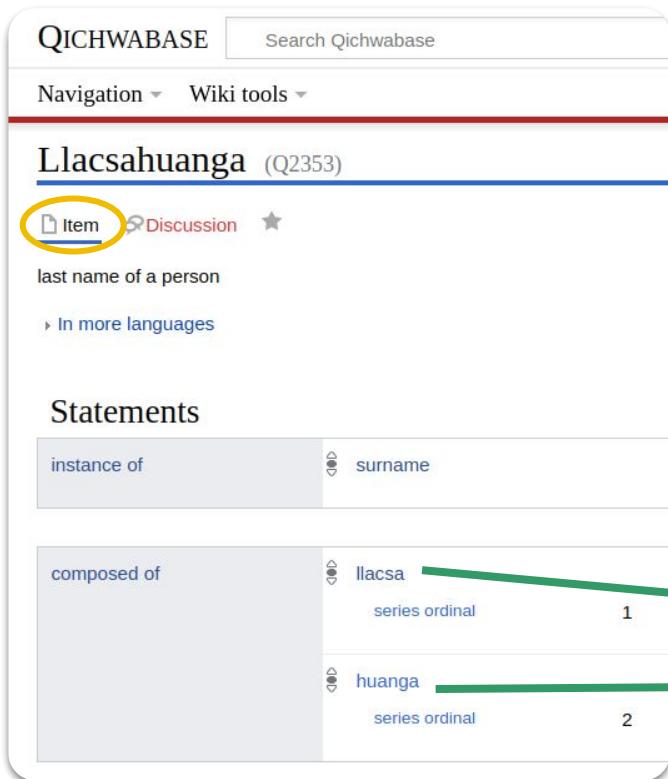
Item Discussion ★

last name of a person

In more languages

Statements

instance of	surname
composed of	<ul style="list-style-type: none">llacsas (series ordinal 1)huanga (series ordinal 2)



QICHWABASE Search Qichwabase

Navigation ▾ Wiki tools ▾

Lexeme Discussion ★

(L9544) **llaqsa**
qu

Language Quechua
Lexical category noun

Senses

L9544-S1 | Spanish bronze; metal en fusión

Statements about L9544-S1

+ add Sense

Forms

L9544-F1 | llacsas
es

Grammatical features

Statements about L9544-F1

+ add Form



QICHWABASE Search Qichwabase

Navigation ▾ Wiki tools ▾

Lexeme Discussion ★

(L22962) **wanka**
qu

Language Quechua
Lexical category noun

Forms

L22962-F1 | huanga
es

Grammatical features

Statements about L22962-F1



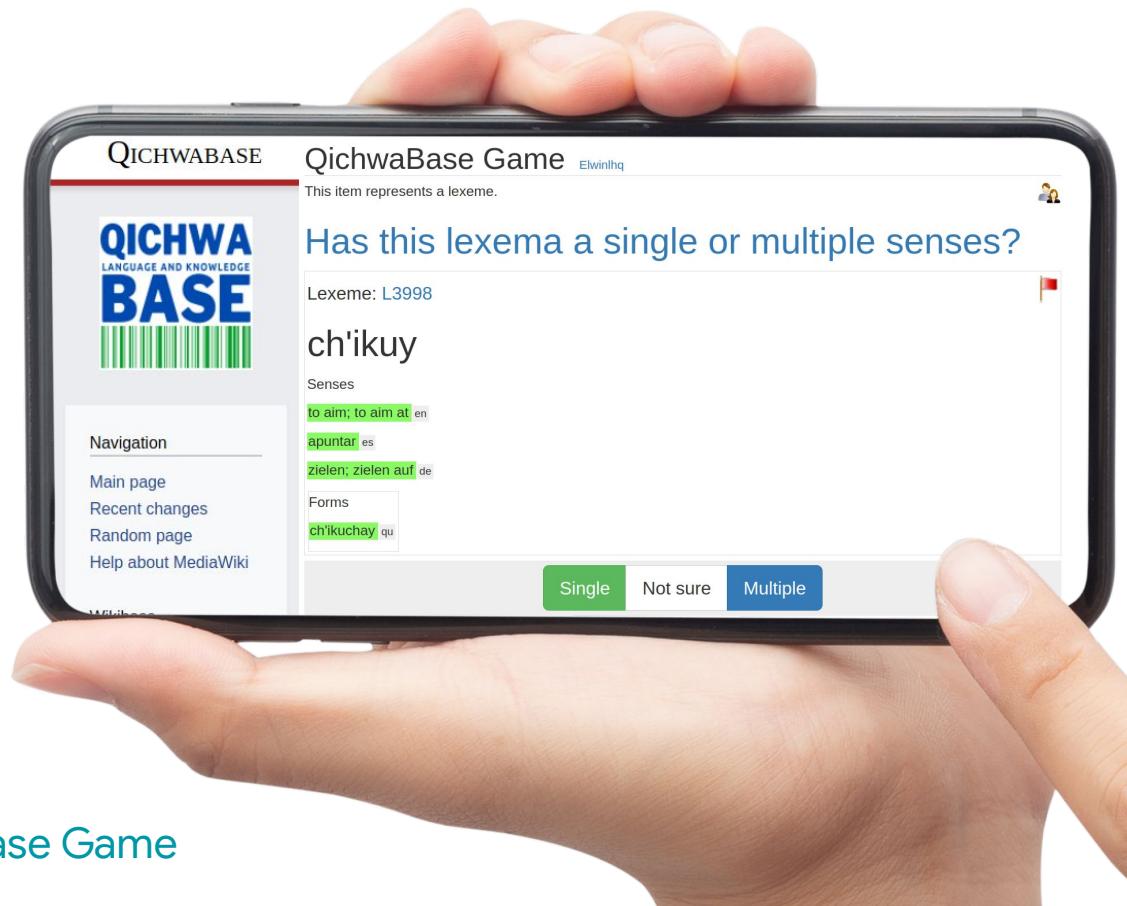
Try it (source code):
[name_processing.py](#)



4. Develop Games : Crowdsourcing quick statements

```
select ?entry ?lemma ?wordForm  
      ?Ses ?Sen ?Sde  
  
where {  
  ?entry ontolex:sense ?sense ;  
        wikibase:lemma ?lemma;  
        ontolex:lexicalForm ?form .  
  ?form ontolex:representation ?wordForm.  
  ?sense a ontolex:LexicalSense ;  
        skos:definition ?Ses ,  
          ?Sen , ?Sde .  
  FILTER(LANG(?Ses) = "es")  
  FILTER(LANG(?Sen) = "en")  
  FILTER(LANG(?Sde) = "de")  
} ORDER BY RAND() LIMIT 1
```

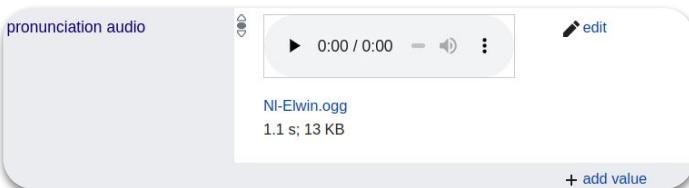
Try it: <https://tinyurl.com/28pf72po>



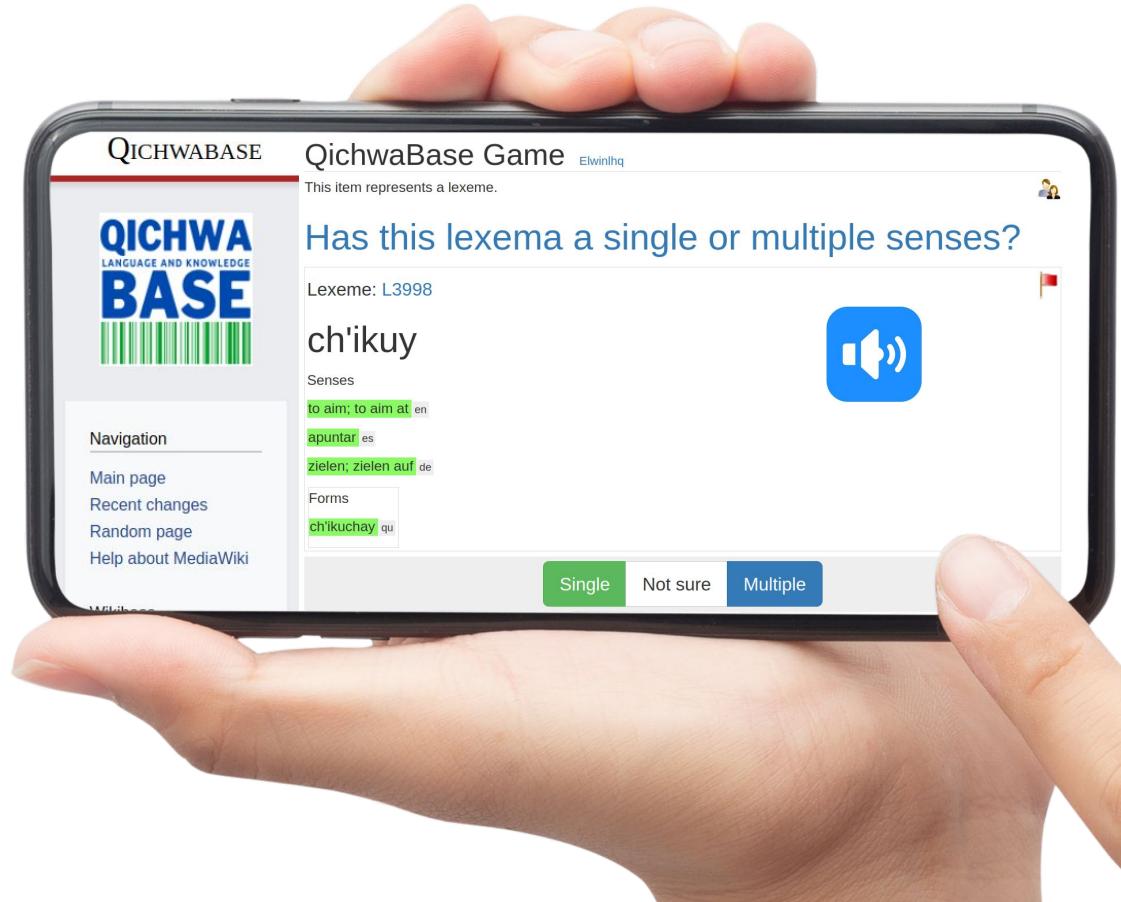
Play it: Qichwabase Game

5. Towards Voices : Recording pronunciation audio

1. Generating a list of Lexemes
 - On Qichwabase ([Try it](#))
2. Recording voices for Lexemes
 - Use LinguaLibre ([Try it](#))
3. Linking the Lexemes
 - from Wikidata
to Qichwabase ([Try it](#))
 - Fill in the property:
 - “pronunciation audio”



4. Add the audio to the Game



Insights - Limitations

- Half a week is not enough
- Modelling is challenging
- Data is dirty
- Get familiar with graph structures
- Reconciliation requires human in the loop
- We need more people

Añay!



Qichwabase

2.0



Natalia Patino & Elwin Huaman
Michael Cochez (Mentor)



June 2023