LIÈGE université
UR Lilith
Liège · Literature · Linguistics

Data sharing in a language documentation context

An Van linden

ULiège Open Science Day

12 November 2026

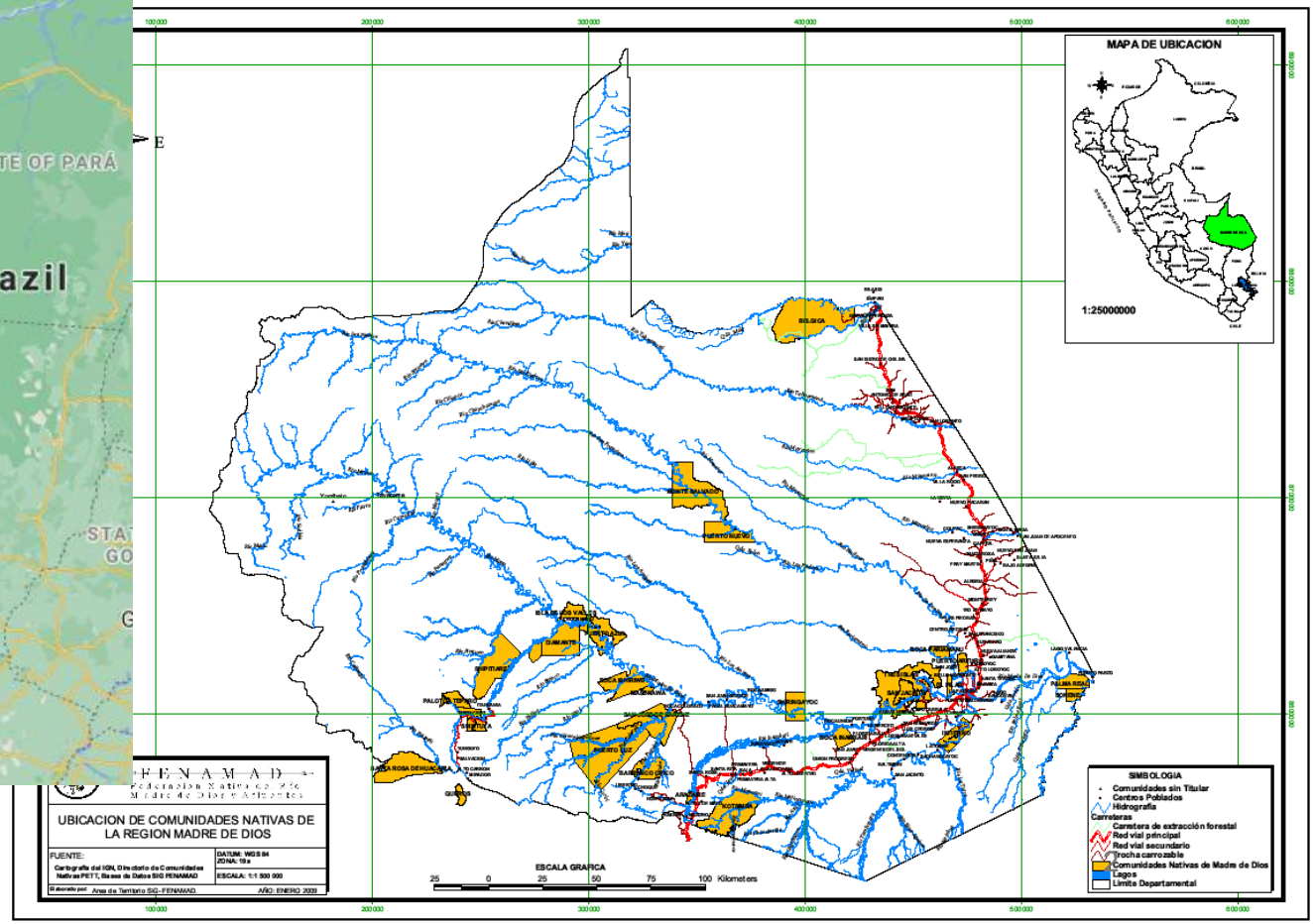
CC BY

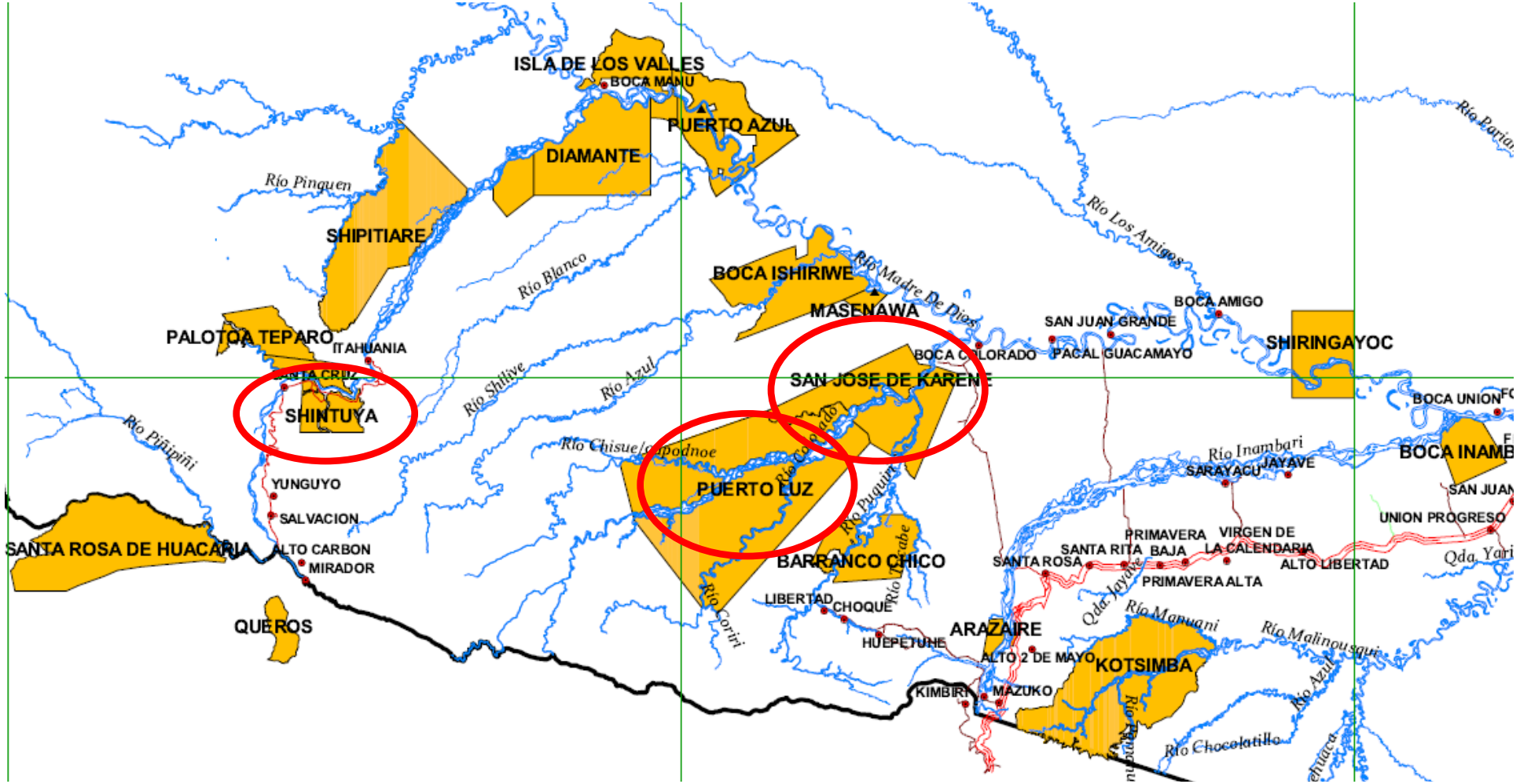


Target language: Harakmbut (unclassified, Peru)



critically endangered





- About 1000 speakers left; distinct dialects
- My focus: Arakmbut/Amarakaeri dialect
- Fieldwork stays in Puerto Luz, San Jose de Karene and Shintuya













Translating Harakmbut recordings with bilingual speaker (Harakmbut – Spanish), community of Puerto Luz,
Madre de Dios, Peru, 2016

Some examples

- Complex verb morphology

(1) *i-ma-niŋ-to-wa-me-te* *wa-knda*
2SG-V.PL-BEN.APPL-CAUS.SOC-go-REC.PST-NFIRSTH NMLZ-egg
'You (SG) took along eggs for them.'

- Verbal classifiers

(2) *men kõsõ ya-**po?**-sak-on?*
which pot 3SG.DUB-CLF:**round**-break-PFV.NVOL
'Which pot is breaking?'

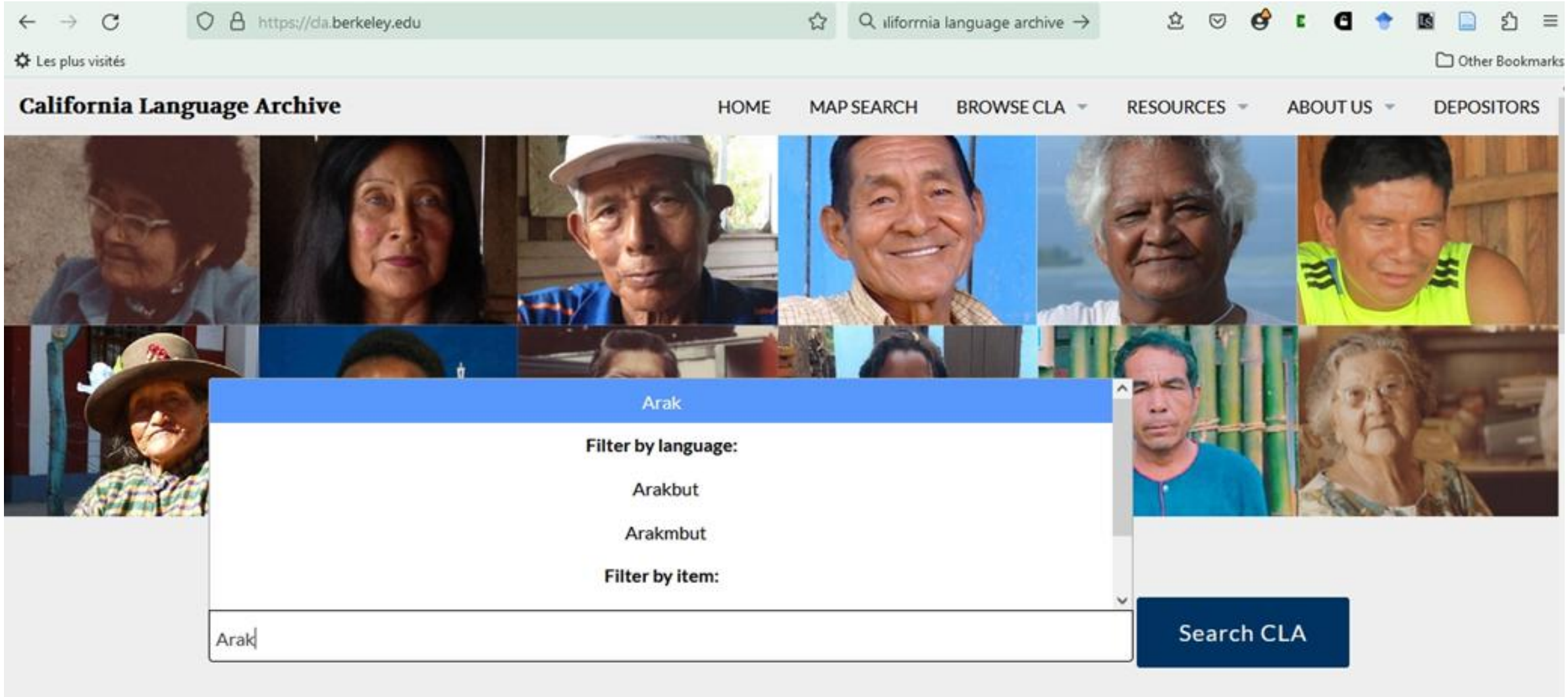
- Bound nouns

(3) *bakoy-oknda*
bird-egg
'bird egg'

DATA SHARING

What? Audio recordings + annotation of Harakmbut oral literature (myths, histories)
→ **intangible cultural heritage**

Where? California Language Archive



Collection accessioned in July 2026

https://cla.berkeley.edu/collection/?collid=11368

34 Arakmbut participants involved



cla.berkeley.edu/collection/?collid=11368

California Language Archive

HOME MAP SEARCH

Collection description | Collection items (119)

Colección de textos arákbut de An Van linden

Collection identifier: 2024-11

Primary contributors: Aquilino Burapa Subia (consultant); Luis Irisanëwa Chinkewa (consultant); Dominga Maca Anïri (consultant); Yésica Patiachi Tayori (consultant); Pablo Yorike Tayori Kä'ari (consultant); An Van linden (researcher, donor)

Additional contributors: Ameriko Aika (consultant); Elías Arbildo Ojeda (consultant); Emilio Ayma Anahija (consultant); Armando Carase Mabirey (consultant); Fermín Chimatani Tayori (consultant, translator); Juan de Dios Chimatani (consultant); Nemecio Huenko Muyuni (consultant); Nilda Huenko (consultant, translator); Äÿöchipo Elías Kodnaro Irinuma (consultant); Juana Machiway (consultant); Lucas Mackavi (consultant); Regino (Chäpo) Manya Sina (consultant); Alfonso Mekeriwa (consultant); Mariano Moque Maca (consultant, translator); Roberto Nayori Machiway (consultant); Lisbeth Patiachi Visse (consultant, translator); Josefina (Yoma) Sewängpua (consultant); Artemio Sikeng (consultant); Bernardo Sonke Dávila (consultant); Óscar Tambet Betera (consultant); Lady Luz Tayori Huenko (consultant, translator); Guadalupe Tayori Keddero (consultant); Luis Miguel Tayori Keddero (consultant); Percy Tayori Keddero (consultant, translator); Gloria Yorike Suki (consultant); Julio Yuri Yagar (consultant); Jonas Arique Moque (translator); Maribel Carase (participant); Efraín Manuqueue Careche (translator)

Language: Arakmbut (amr)

Dates: 2010-

Historical information: La lingüista belga An Van linden obtuvo su doctorado en Lingüística Inglesa en 2009 en la Universidad de Lovaina (Bélgica) y actualmente es profesora de lengua inglesa y lingüística en la Universidad de Lieja (Bélgica). Comenzó a documentar y describir la lengua harákbut en 2010, con una primera estancia de

Search CLA

Collection description | Collection items (119)

1 – 25 of 119 Results Show All/ Collapse All

1. **El tigre que mató al abuelo de Maribel / The jaguar who killed Maribel's grandfather** (07 Aug 2010 to 12 Aug 2011) (5 digital files, with audio)
2. **La esposa que engañó al cazador / The wife who cheated on the hunter** (07 Aug 2010 to 10 Aug 2011) (3 digital files, with audio)
3. **La guerra entre los Tacas y los Arakmbut / The war between the Taca and the Arakmbut** (07 Aug 2010 to 13 Aug 2016) (5 digital files, with audio)
4. **La hija que se perdió en el agua / The daughter who got lost in the river** (07 Aug 2010 to 25 Aug 2011) (4 digital files, with audio)
5. **Los Tacas atacaron a los Keroseris / How the Taca attacked the people of Queros** (07 Aug 2010 to 17 Sep 2011) (4 digital files, with audio)
6. **Los Tacas atacaron a los niños y mujeres Arakmbut / How the Taca attacked the children and women of the Arakmbut** (07 Aug 2010 to 13 Aug 2016) (4 digital files, with audio)
7. **El tigre que orinó al fuego / The jaguar that urinated on the campfire** (07 Aug 2010 to 19 Sep 2011) (4 digital files, with audio)
8. **El tigre que atacó una choza / The jaguar who attacked a hut** (14 Aug 2010 to 12 Aug 2011) (3 digital files, with audio)

Steps

- 5 field trips as of 2010 + continuous annotation (segmentation, transcription, translation)
- 2023: start archiving process with CLA manager Zachary O'Hagan, shared spreadsheet with metadata; file transfer to Berkeley server via Belnet FileSender (4 batches)
- 2026: first collection accessioned + giving back to community, facing language loss in gold-mining context → collection aids **culture preservation**

2010	Monolingual audio recording in 3 communities
2011	Monolingual audio recording + translation in 4 communities
2016	Monolingual audio recording + translation in 2 communities
2023	Monolingual audio recording + translation in 1 community + Asking consultants (or relatives) permission for public access in CLA
2026	Monolingual audio recording + translation in 1 community + Presenting CLA collection → I made relatives listen to recordings of their deceased loved ones

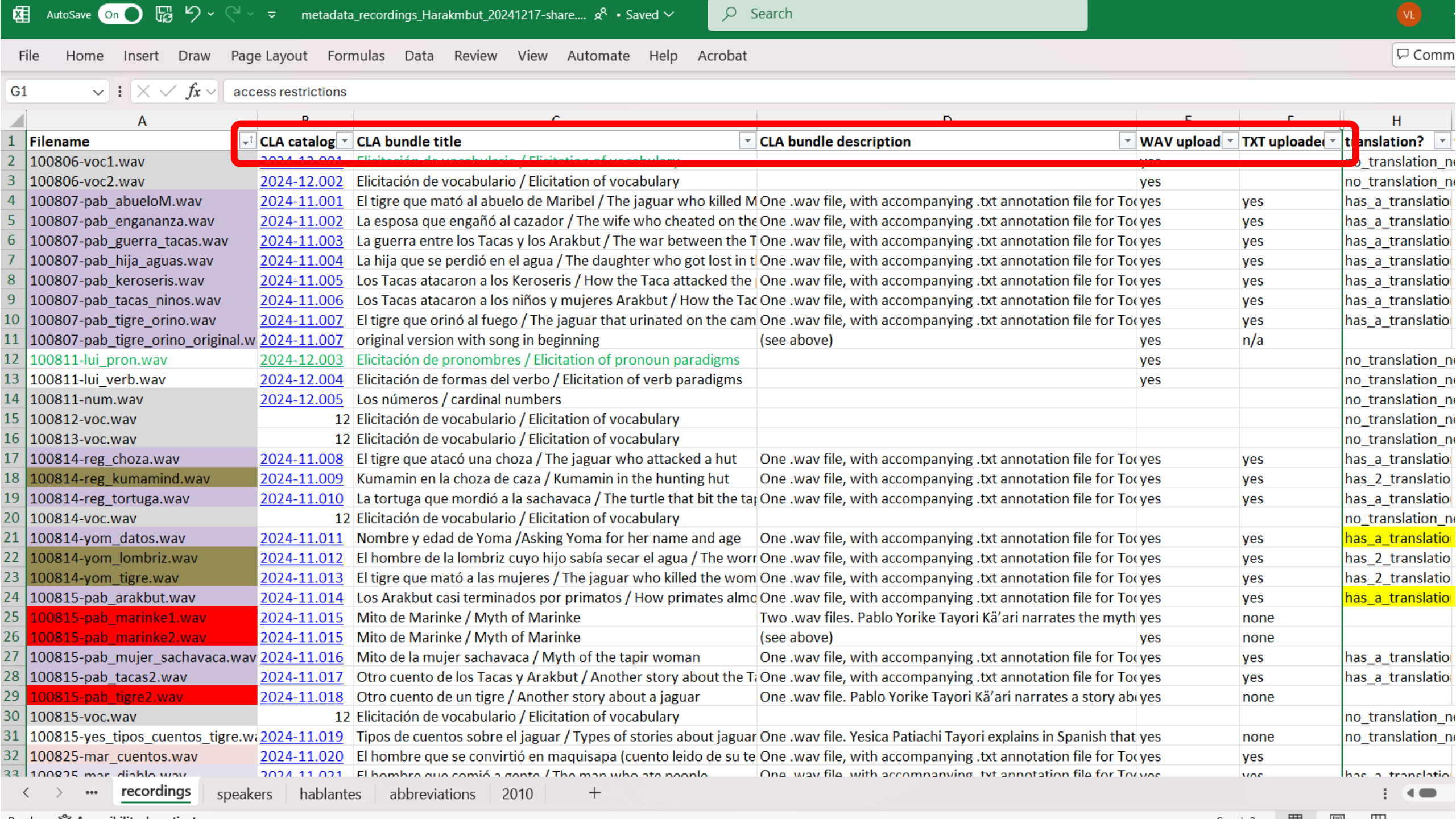
AutoSave Off | metadata_recordings_Harakmbut.xlsx | Search | VL

File Home Insert Draw Page Layout Formulas Data Review View Automate Help Acrobat | Commen

Q6

	A	B	C	D	F	G	H	N	O	P	Q	R	S	T
1	Filename	translation?	files with trans	Title	year	segmen	trans	file format	Genre	Participants	Pa	Length	Topics	Location/C
2	100806_voc1.wav	no_translation_needed			2010	yes	t, f	.wav	voc	Maribel Carase, Fermin Chimanati Tayori			vocabulary	Puerto Luz
3	100806-voc2.wav	no_translation_needed			2010	yes	t, f	.wav	voc				vocabulary	Puerto Luz
4	100807-pab_abueloM.wav	has_a_translatior	110811-nil_pab_abueloM_1.wav		2010	yes	no	.wav	anc	Pablo Yurike Tayori Kä'ari			jaguar killed Maribel'	Puerto Luz
5	100807-pab_engananza.wav	has_a_translatior	110810-nil_pab_engananza.wav		2010	yes	no	.wav	nrr	Pablo Yurike Tayori Kä'ari			La esposa que engañ	Puerto Luz
6	100807-pab_guerra_tacas.wav	has_a_translatior	160808-per_pab_guerra_tacas1.wav		2010	yes	no	.wav	hst	Pablo Yurike Tayori Kä'ari			La guerra entre los T	Puerto Luz
7	100807-pab_hija_aguas.wav	has_a_translatior	110825-jon_pab_hija_aguas1.wav		2010	yes	no	.wav	nrr	Pablo Yurike Tayori Kä'ari			La hija que se perdió	Puerto Luz
8	100807-pab_keroseris.wav	has_a_translatior	110917-lis_pab_keroseris1.wav		2010	yes	no	.wav	hst	Pablo Yurike Tayori Kä'ari			Los Tacas atacaron a	Puerto Luz
9	100807-pab_tacas_ninos.wav	has_a_translatior	160813-per_pab_tacas_ninos1.wav		2010	yes	no	.wav	hst	Pablo Yurike Tayori Kä'ari			los Tacas atacaron a	Puerto Luz
10	100807-pab_tigre_orino.wav	has_a_translatior	110919-lis_pab_tigre_orino.wav		2010	yes	no	.wav	nrr	Pablo Yurike Tayori Kä'ari			El tigre que orinó al f	Puerto Luz
11	100811-lui_pron.wav	no_translation_needed			2010			.wav	eli	Luis Tayori Kendero			pronouns	Puerto Luz
12	100811-lui_verb.wav	no_translation_needed			2010			.wav	eli	Luis Tayori Kendero			verbs	Puerto Luz
13	100811-num.wav	no_translation_needed		numerals	2010	yes	t, f	.wav	eli	Pablo Yurike Tayori Kä'ari			numerals	Puerto Luz
14	100812-voc.wav	no_translation_needed			2010	yes	t, f	.wav	voc					Puerto Luz
15	100813-voc.wav	no_translation_needed			2010	yes	t, f	.wav	voc					Puerto Luz
16	100814-reg_choza.wav	has_a_translatior	110812-nil_reg_choza.wav		2010	yes	no	.wav	nrr	Regino Manya				Puerto Luz
17	100814-reg_kumamind.wav	has_2_translatior	110829-mar_reg_kumamind1.wav; 11		2010	yes	no	.wav	nrr	Regino Manya				Puerto Luz
18	100814-reg_tortuga.wav	has_a_translatior	110809-nil_reg_tortuga.wav		2010	yes	no	.wav	nrr	Regino Manya				Puerto Luz
19	100814-voc.wav	no_translation_needed			2010			.wav						Puerto Luz
20	100814-yom_datos.wav				2010			.wav	cvs	Josefina Yoma Sewängpua + Maribel				Puerto Luz
21	100814-yom_lombriz.wav	has_2_translatior	110913-lis_yom_lombriz.wav; 110809		2010	yes	t, f	.wav	nrr	Josefina Yoma Sewängpua				Puerto Luz
22	100814-yom_tigre.wav	has_2_translatior	110913-lis_yom_tigre1.wav; 110809-		2010	yes	t, f	.wav	nrr	Josefina Yoma Sewängpua				Puerto Luz
23	100815-pab_arakbut.wav				2010	yes	no	.wav	hst	Pablo Yurike Tayori Kä'ari				Puerto Luz
24	100815-pab_marinke1.wav				2010			.wav	nrr	Pablo Yurike Tayori Kä'ari				Puerto Luz
25	100815-pab_marinke2.wav				2010			.wav	nrr	Pablo Yurike Tayori Kä'ari				Puerto Luz
26	100815-pab_mujer_sachavaca.wav	has_a_translatior	110914-lis_pab_mujer_sachavaca.wa		2010	yes	no	.wav	nrr	Pablo Yurike Tayori Kä'ari				Puerto Luz
27	100815-pab_tacas2.wav	has_a_translatior	160806-per_pab_tacas2_1.wav		2010	yes	no	.wav	hst	Pablo Yurike Tayori Kä'ari				Puerto Luz
28	100815-pab_tigre2.wav				2010			.wav	nrr	Pablo Yurike Tayori Kä'ari				Puerto Luz
29	100815-voc.wav	no_translation_needed			2010	yes	t, f	.wav	voc					Puerto Luz
30	100815-yes_tipos_cuentos_tigre.wav	no_translation_needed			2010			.wav	exp	Yesica Patiachi Tayori				Puerto Luz
31	100825-mar_cuentos.wav				2010	yes	no	.wav	nrr	Mariano Moque Maca				San José del K
32	100825-mar_diablo.wav	has_a_translatior	110810-nil_mar_diablo_1.wav		2010	yes	no	.wav	nrr	Mariano Moque Maca				San José del K

recordings | speakers | abbreviations | 2010 | +



AutoSave On metadata_recordings_Harakmbut_20241217-share.... • Saved Search VL

File Home Insert Draw Page Layout Formulas Data Review View Automate Help Acrobat Comm

G1 access restrictions

	A	B	C	D	E	F	H
1	Filename	CLA catalog	CLA bundle title	CLA bundle description	WAV upload	TXT uploade	translation?
2	100806-voc1.wav	2024-12.001	Elicitación de vocabulario / Elicitation of vocabulary		yes		o_translation_ne
3	100806-voc2.wav	2024-12.002	Elicitación de vocabulario / Elicitation of vocabulary		yes		no_translation_ne
4	100807-pab_abueloM.wav	2024-11.001	El tigre que mató al abuelo de Maribel / The jaguar who killed M	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
5	100807-pab_engananza.wav	2024-11.002	La esposa que engañó al cazador / The wife who cheated on the	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
6	100807-pab_guerra_tacas.wav	2024-11.003	La guerra entre los Tacas y los Arakbut / The war between the T	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
7	100807-pab_hija_aguas.wav	2024-11.004	La hija que se perdió en el agua / The daughter who got lost in t	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
8	100807-pab_keroseris.wav	2024-11.005	Los Tacas atacaron a los Keroseris / How the Taca attacked the	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
9	100807-pab_tacas_ninos.wav	2024-11.006	Los Tacas atacaron a los niños y mujeres Arakbut / How the Tac	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
10	100807-pab_tigre_orino.wav	2024-11.007	El tigre que orinó al fuego / The jaguar that urinated on the cam	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
11	100807-pab_tigre_orino_original.w	2024-11.007	original version with song in beginning	(see above)	yes	n/a	
12	100811-lui_pron.wav	2024-12.003	Elicitación de pronombres / Elicitation of pronoun paradigms		yes		no_translation_ne
13	100811-lui_verb.wav	2024-12.004	Elicitación de formas del verbo / Elicitation of verb paradigms		yes		no_translation_ne
14	100811-num.wav	2024-12.005	Los números / cardinal numbers				no_translation_ne
15	100812-voc.wav	12	Elicitación de vocabulario / Elicitation of vocabulary				no_translation_ne
16	100813-voc.wav	12	Elicitación de vocabulario / Elicitation of vocabulary				no_translation_ne
17	100814-reg_choza.wav	2024-11.008	El tigre que atacó una choza / The jaguar who attacked a hut	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
18	100814-reg_kumamind.wav	2024-11.009	Kumamin en la choza de caza / Kumamin in the hunting hut	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_2_translatio
19	100814-reg_tortuga.wav	2024-11.010	La tortuga que mordió a la sachavaca / The turtle that bit the tap	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
20	100814-voc.wav	12	Elicitación de vocabulario / Elicitation of vocabulary				no_translation_ne
21	100814-yom_datos.wav	2024-11.011	Nombre y edad de Yoma /Asking Yoma for her name and age	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
22	100814-yom_lombriz.wav	2024-11.012	El hombre de la lombriz cuyo hijo sabía secar el agua / The worr	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_2_translatio
23	100814-yom_tigre.wav	2024-11.013	El tigre que mató a las mujeres / The jaguar who killed the wom	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_2_translatio
24	100815-pab_arakbut.wav	2024-11.014	Los Arakbut casi terminados por primatos / How primates almo	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
25	100815-pab_marinke1.wav	2024-11.015	Mito de Marinke / Myth of Marinke	Two .wav files. Pablo Yorike Tayori Kä'ari narrates the myth	yes	none	
26	100815-pab_marinke2.wav	2024-11.015	Mito de Marinke / Myth of Marinke	(see above)	yes	none	
27	100815-pab_mujer_sachavaca.wav	2024-11.016	Mito de la mujer sachavaca / Myth of the tapir woman	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
28	100815-pab_tacas2.wav	2024-11.017	Otro cuento de los Tacas y Arakbut / Another story about the T	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio
29	100815-pab_tigre2.wav	2024-11.018	Otro cuento de un tigre / Another story about a jaguar	One .wav file. Pablo Yorike Tayori Kä'ari narrates a story ab	yes	none	
30	100815-voc.wav	12	Elicitación de vocabulario / Elicitation of vocabulary				no_translation_ne
31	100815-yes_tipos_cuentos_tigre.wa	2024-11.019	Tipos de cuentos sobre el jaguar / Types of stories about jaguar	One .wav file. Yesica Patiachi Tayori explains in Spanish that	yes	none	no_translation_ne
32	100825-mar_cuentos.wav	2024-11.020	El hombre que se convirtió en maquisapa (cuento leido de su te	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	
33	100825-mar_diablo.wav	2024-11.021	El hombre que comió a gente / The man who ate people	One .wav file, with accompanying .txt annotation file for Toc	yes	yes	has_a_translatio

recordings speakers hablantes abbreviations 2010 +

Outlook: 2nd collection at CLA

- 16 years of documentation and description: also elicitation data on vocabulary and grammar:
  - Harakmbut translation of Spanish lexemes
  - Harakmbut translation of ‘invented’ sentences, contrastive contexts
  - Harakmbut description of visual stimuli
  - Paradigms of pronouns and verb forms
- useful mainly for scientific community (linguists, anthropologists, biologists, archaeologists), but also to local bilingual teachers
- Might prove useful in later **revitalization** efforts (when language is lost)
- Permission from all speakers (n=41 for 2 collections) to make the recordings publicly available

THANK YOU FOR YOUR ATTENTION - Dakiti



Daybreak at the Reserva Comunal Amarakaeri – July 2026