

Methods and data management tools for using WhatsApp for language work

Kelsey Neely

ELAR Remote Fieldwork Workshop Series

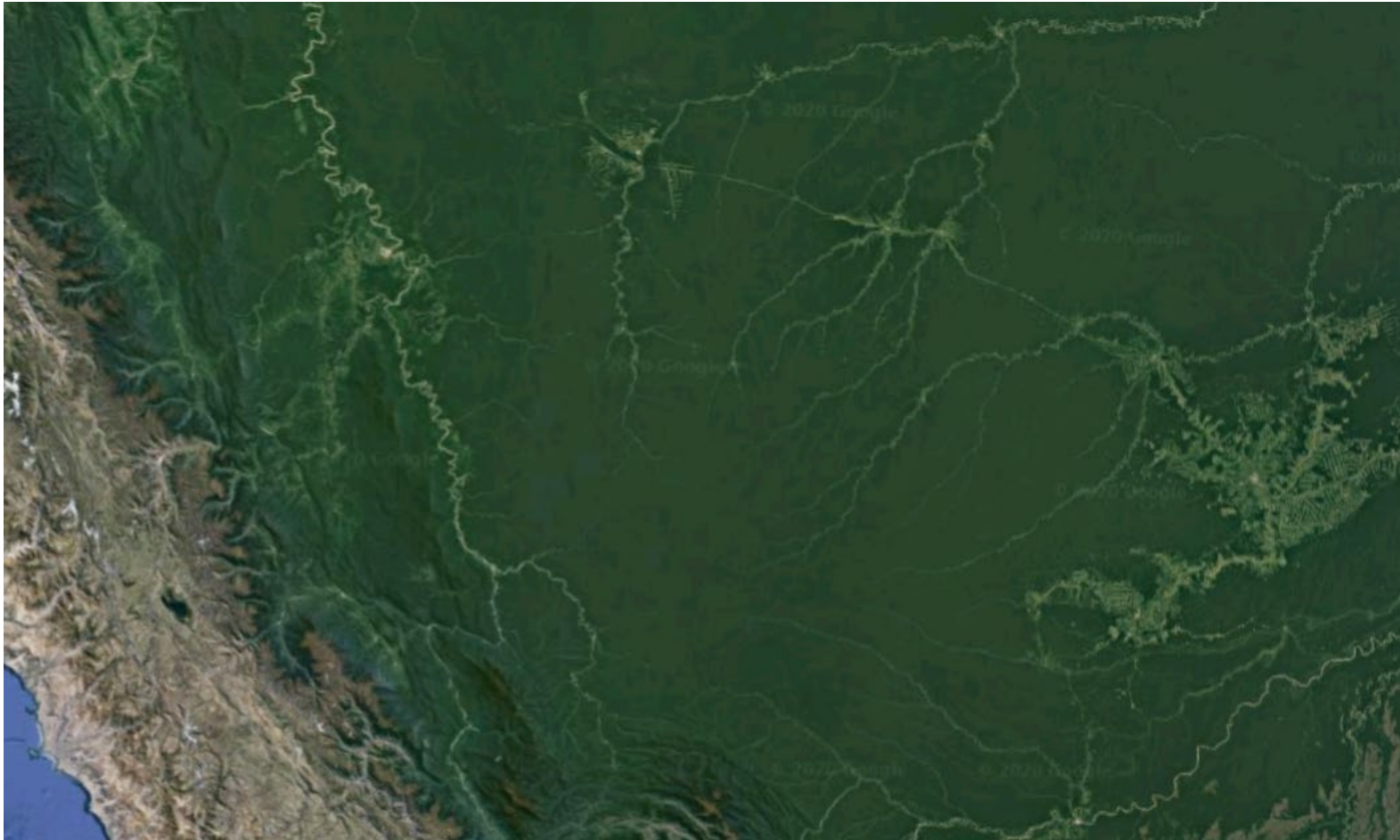
17 September, 2021

Overview

- Background
- Necessary equipment & infrastructure
- Ethical, social, & cultural considerations
- Some example tasks
- Using Python + Ffmpeg to process clips
- Using Python + Ffmpeg to create clips

Background

- Sepahua, Ucayali, Perú



Background

- Sepahua, Ucayali, Perú



Background

- Sepahua, Ucayali, Perú
- Yaminawa Language Documentation Project
- Appx. 12 Yaminahua & Nahua speakers with 7+ years of experience
- Only one or two with smartphone experience

- Two cell towers; one with mobile internet
- Slow, halting satellite-based connection
- Electricity for two periods (8a-12p, 6p-11p)
- Phones available locally; not feasible to send other equipment

Acknowledgements

- The Yaminahua and Nahua speakers who piloted this method with me, most especially Delicia Gomez Ramírez, Rebeca Seido Thaihui, and Rosa Gomez Ramírez
- Generous funding from ELDP grant IPF 0308 “Multidialectal Lexical Documentation of Yaminahua, Nahua, and Sharanahua”
- Abundant flexibility and moral support from the ELDP/ELAR team!

Equipment

- Smartphones with WhatsApp installed & activated
 - 1 for language teacher, 1 for language learner
- Laptop/computer to download clips using WhatsApp Web
 - Only the person managing the data *needs* this
- Wifi connection for data manager to connect phone to computer

- Useful: notebooks & writing utensils, headset and USB mics
- Optional: second monitor, ergonomic mouse for laptop

Software

- Python (3 or higher) installed
- Ffmpeg installed
- ELAN for annotating after you're done!

Infrastructure

- An internet connection at the local (field) site!
 - Mobile is necessary to activate WhatsApp – may need to be reactivated
 - Wifi from a satellite or other connection works for day-to-day process
- Electricity access at the local site for charging phone
 - Or powering modem/router if not using mobile internet
- Internet on a wifi connection at the distant (data manager) site
 - Mobile is also necessary to activate/re-activate WhatsApp
 - Also electricity, obviously
- Human: language work experience, literacy skills, etc.

Infrastructure

- How do you plan to get necessary equipment to speakers?
- How do you plan to compensate participants?
- Is your institution okay with these plans?

Ethical, Social, & Cultural Considerations

- Even though you might not be visiting the community in-person, this is still data collection with human subjects!
- Covid adds an extra layer of concern
- Data will be transmitted over a corporate 3rd party, is this okay?
- Social: If providing equipment, who gets a phone? Does this potentially create or exacerbate social inequality?
- Cultural: What kinds of activities are culturally appropriate over the phone? What kinds of data are people comfortable recording this way?

Some example tasks

- Typical elicitation tasks:
 - lexical
 - morphosyntactic, even morphophonological!
 - Audio from WhatsApp is compressed (opus codec in .ogg container), but you can still make progress even on some phonological topics!
- Creating/improving annotations for existing recordings
 - Adding translations, careful respeakings
 - Checking transcriptions, etc.

Let's get started!