

A'INGAE LANGUAGE DOCUMENTATION

Justin Bai, Maksymilian Dąbkowski, Kalinda Pride, and Nicholas Tomlin
PI: Scott Anderbois | Brown University, Department of CLPS

Abstract

A'ingae (also known as Cofán) is a severely under-documented language spoken in Ecuador and Colombia, with approximately 1500 native speakers. It is considered a language isolate (that is, it has no known genetic relationship with any other languages). In this project, we're working to collect various A'ingae texts and make them accessible online, while also rigorously studying the language. This work is made possible by Hugo Lucitante (Class of 2019), a Brown University student and native speaker of A'ingae.



Trip to Ecuador

Over the summer, part of our team visited Dureno and Zábaló, two Ecuadorian A'ingae-speaking communities. The work included:

- Recording dialogue, traditional narratives, recipes, and medicine
- Meeting community leaders to discuss future research visits
- Demonstrating the software we created
- Training community members to use FLEX/ELAN



What is Language Documentation?

Language documentation is the process of collecting representative and diverse samples of language data, typically from endangered languages. These samples should include spontaneous conversations, story-telling, traditional knowledge, etc. These data may be used for language revitalization and for studying the language's structure. Often, language documentation involves a "triptych" of works: a grammar, a lexicon, and a collection of stories/works in the language. Language documentation is accomplished through linguistic fieldwork, the act of collecting and interpreting data about language in its natural environment. Field methods include eliciting words and sentences, listening to speakers, and making records of data (whether through paper and pencil or recordings—ideally, both).

Example Linguistic Analysis: Shape Clitics

-'cho round or diminutive

tevaen 'to write'	tevaen'cho 'letter'
can 'to look'	can'cho 'glasses'
tsu'tta 'bone'	tsutta'cho 'knee'
dusu- 'young'	dusu'cho 'egg'

-'cco angular or augmentative

tevaen 'to write'	tevaen'cco 'pen'
can 'to look'	canco 'mirror'
nojan 'thorn'	nojan'cco 'needle'
chho'chho 'to grub'	chhochho'cco 'beater'

-'je flat

tevaen 'to write'	tevaen'jen 'book'
na 'fruit'	najen 'leaf'
ondiccu 'to dress'	ondiccu'je 'male wear'
mesa 'table'	mesa'je 'table'

-'fa lateral

va 'this'	vafa 'this side'
tandan 'to tie'	tandan'fa 'string'
anjampa 'blood'	anjampa'fa 'vein'
sho 'vagina'	sho'fa 'penis'

-'qui linear

na'en 'river'	naiqui 'stream'
tsi 'to walk'	tsaiqui 'path'
re'ri- 'little'	re'ruiqui- 'narrow'
osha 'to spill'	oshe'qui 'river arm'

-'ju surface

tsove'ju 'face'	catifo'cho 'trash'
camba'ju 'face'	jurofo'cho 'soot'
chan- 'shine'	atui'cho 'tiny splinters'
cca'ya 'swim'	anfo'cho 'food debris'

-si thorny or hirsute

fundui 'to sweep'	fundui'si 'broom'
butto 'to run'	buttosi 'fast snake'
cuno- 'hunger'	cuñosi 'skinny'
na 'fruit'	nasi 'wasp'

-jin large

quinijin 'tree'	
na'jin 'trunk'	
shavo 'canoe'	
tse 'that'	

-jion tall

ttuttu 'to fell'	ttuttujion 'log'
tsa'o 'house'	tsa'ojion 'storey house'
chango 'hole'	changongojion 'indented'
ta'va 'ceiba'	tavajion 'large ceiba'

-fi'ndi shard

dofi'ndi 'piece'	
atuifi'ndi 'splinter'	
corifi'ndi 'money'	
chhochhofi'ndi 'oar'	

-fo'cho scattered

cati 'to throw away'	catifo'cho 'trash'
ju'ru 'to burn'	jurofo'cho 'soot'
atui 'to chop'	atui'cho 'tiny splinters'
an 'to eat'	anfo'cho 'food debris'

-ccu delimited space

coye 'plantain'	coyeccu 'plantain field'
cu'i 'to drink'	cui'ccu 'plantain drink'
tssutssian 'to dilute'	tssutssian'ccu 'juice'
si'nge 'fire'	singeccu 'match'

Interlinear Glossing

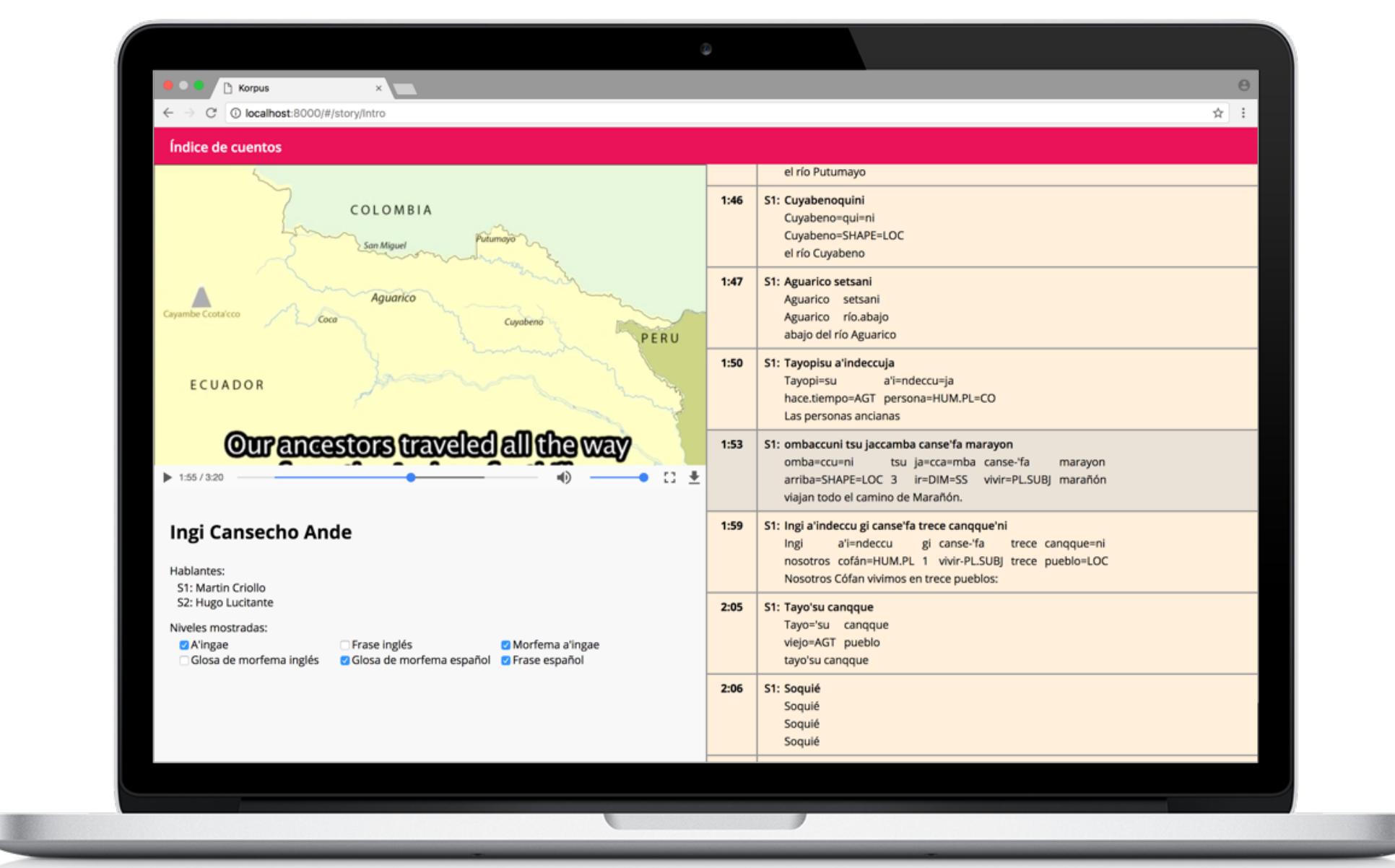
To work with languages they don't speak, linguists use interlinear glosses to associate each word segment with its meaning, along with the meaning of the whole phrase.

Ecuadorningi canse'fa mil a'indeccu
Ecuador=LOC=1 live-PL.SUBJ thousand person-HUM.PL
'A thousand of us A'i live in Ecuador.'

As part of documenting the A'ingae language, we create interlinear glosses for the collected data. We have been working on building a corpus of texts (such as traditional A'ingae folklore) and doing linguistic analyses on them by glossing them. We create and store our annotations using FieldWorks Language Explorer (FLEX) and ELAN, pieces of software which help researchers build lexicons, work on interlinear texts for languages, and segment and annotate recordings such as audio and video.

Software Development

We built a website to flexibly display the interlinear texts, synchronized to corresponding video or audio. Our software will make the interlinear texts accessible to other linguists, to the A'ingae-speaking community, and to the general public.



A fully-featured offline version of the site was given to the A'ingae-speaking community, which has minimal Internet access. We are now integrating a search and concordance feature into the site. We will also create an Android app for offline viewing of the texts and videos, for increased access in the A'ingae-speaking community where phones greatly outnumber laptops and tablets. The software is easily adaptable to other languages, and we hope it will be used by many language documentation projects to democratize access to their data.