

Ticuna (tca) language documentation: A guide to materials in the California Language Archive

Amalia Skilton

University of Texas at Austin

Ticuna (ISO: tca) is a language isolate spoken in the northwestern Amazon Basin (Brazil, Colombia, Peru). Ticuna has more speakers than almost all other Indigenous Amazonian languages and – unlike most languages of the area – is still learned by children. Yet academic linguists have given it relatively little research attention. Therefore, to raise the profile of this areally important language, I offer a guide to three collections of Ticuna language materials held in the California Language Archive. These materials are extensive, including over 1,396 hours of recordings – primarily of child language and everyday conversations between adults – and 33 hours of transcriptions. To contextualize the materials, I provide background on the Ticuna language and people; the research projects which produced the materials; the participants who appear in them; and the ethical and permissions issues involved in collecting them. I then discuss the nature and scope of the materials, showing how the content of each collection motivated collection-specific choices about recording, transcription, organization in the archive, and metadata. Last, I outline how other researchers could draw on the collections for comparative analysis.

1. Introduction¹ Language documentation researchers invest enormous amounts of time in creating archival deposits, but we rarely provide descriptions of our collections (Sullivant 2020). Moreover, even when we do describe archival collections in detail, our descriptions are usually not designed to encourage others to use the collections for comparative work. This is surprising, since comparative research is not only a common use of archival collections (e.g., Begay 2017: 2–3) but also a central argument for the funding of language documentation programs.

Thus, on the model of Caballero (2017), this paper describes an archival deposit which was theorized (Woodbury 2011) and structured to facilitate use in comparative research, among other purposes. Beyond describing the deposit, I also outline

¹I thank all of the Ticuna-speaking people who participated in the creation of the materials, especially Angel Bitancourt Serra, Deoclesio Guerrero Gomez, and Lilia Witancort Guerrero. I likewise thank all of the Ticuna-speaking teachers who have taken time to discuss language topics with me, especially Betsabe Guerrero Witancort, Mavis Guerrero Witancort, Sotil Suárez González, and Ling Cándido Serra. I also thank Pattie Epps for comments on this paper; Andrew Garrett and Tony Woodbury for encouraging me to write it; and the staff of the California Language Archive, especially Zach O'Hagan, for help with the archiving process. Moreover, I would like to acknowledge that the University of Texas at Austin, where I completed this work, is located on the traditional land of the Tonakawa, Comanche, Lipan Apache, and other Indigenous peoples, whose relationships with that land continue today. This research was supported by an NSF Graduate Research Fellowship (2014–2019), NSF BCS-1741571 (2017–2020), NSF SMA-1911762 (2019–), and three Oswalt Endangered Language Grants from UC Berkeley (2015–2017).

how other researchers can use it independently for cross-linguistic and cross-dialectal comparisons.

The deposit consists of recordings, transcriptions, and other materials documenting the Ticuna language (ISO: *tca*; isolate) as spoken in the town of Cushillococha, Peru between 2015 and 2019. In subject matter, the deposited materials focus on naturally occurring events (events that would occur independent of research, such as conversations between family members) and on first language acquisition. The materials are held digitally in the California Language Archive at UC Berkeley, in three collections: Collection 2015-06 “Ticuna elicitation and texts”, Collection 2018-19 “Ticuna conversations”, and Collection 2018-20 “Ticuna experiments”.²

I have two goals in describing the collections. First, I want to encourage other researchers to use the materials for comparative projects, per their design. Second, most language documentation corpora currently do not include documentation of informal, everyday forms of talk, such as conversation and child-directed language (Rosenblum & Berez-Kroeker 2018: 347–348; Dingemanse & Floyd 2014: 448). One reason for this is that informal talk is intrinsically challenging to record and transcribe, requiring up to three hours of annotation time for every minute of speech (Rossi et al. 2020: 11–13). Therefore, a second goal of this paper is to present the methods that allowed me to record and analyze high-quality data from several different kinds of informal language use in these collections. In doing this, I hope to stimulate more documentation workers to collect data on conversation and other forms of everyday talk, since this data is crucial for language reclamation (Baldwin et al. 2018: 193; Grenoble 2013: 53–54; Hermes & King 2019) as well as description (Bower 2008: 119–122).

The rest of this article is structured as follows. I begin, in §2, with background information about the Ticuna people, language, and land. In §3, I describe the participants, history, and objectives of the project that produced the materials. §§4–8 are the heart of the paper, providing detailed information about each of the collections. §4 discusses the collection-level and bundle-level organization of the materials, and §5 describes the ethical issues raised by collecting and archiving these forms of data. Following this general discussion, §§6–8 discuss each of Collections 2015-06, 2018-19, and 2018-20 in order. §9 considers possible audiences and uses of the collections, and §10 concludes.

2. Language background

2.1 Speakers and location Ticuna is spoken by 38,680 (Lewis et al. 2014) to 69,000 (Instituto Socio-Ambiental 2017) people in the Amazon Basin. It is the most widely spoken Indigenous language of Brazil and among the most widely spoken languages of the entire Amazon Basin. Genetically, the language is unclassified, though Carvalho (2009) and Seifart & Echeverri (2014) suggest a relationship with Yuri (another iso-

²The collection numbers begin with the (four-digit) year the collection was established, followed by two arbitrary digits. They are not year ranges.

late, which currently has no speakers). All speakers who I have met accept and use the word *Ticuna* as a name for both the people and the language.

Ticuna people currently live along 600km of the Amazonas/Solimoes River in Loreto department, Peru; Amazonas department, Colombia; and Amazonas state, Brazil. The current western extreme of Ticuna-speaking settlement is the town of Cushillococha, Peru; the eastern extreme is in Amazonas state, Brazil, near the mouth of the river Jutai. Figure 1 shows these locations within northern South America. The Amazonas/Solimoes is highlighted in white.

Figure 1. Location of Ticuna territory and of Cushillococha, within northern South America. Base map via Google Earth.



While Ticuna speakers from across the region are readily able to understand one another, there is some dialect variation (Santos Angarita 2005). All of the materials in the collection were collected in Cushillococha, Peru and the adjoining town of Caballococha, Peru.

Cushillococha (Ticuna: *Ki³tʃi³tu¹*) is a titled Indigenous community located in Mariscal Ramón Castilla district and province, Loreto region. As of late 2019, ~5,000 people lived in Cushillococha and its adjacent unincorporated communities. At that time, effectively all residents of the town were ethnically Ticuna and spoke Ticuna as their first and dominant language. Though there were signs of language shift in some age groups, Ticuna was acquired by nearly all children and remained the main language spoken in all domains of life, including church, government, and school (which was taught monolingually in Ticuna up to age eight). Most recordings in the collections were made in Cushillococha.

Caballococha (Ticuna: *Ko³we⁴na²*) is the capital of Mariscal Ramón Castilla district and province. The population of Caballococha was ~15,000 as of late 2019. Most residents were non-Indigenous and spoke Spanish, but the town also included predominantly Indigenous neighborhoods populated by Ticuna and Yagua people

(Yagua is a Peba-Yaguan language). Most recordings of elicitation and some recordings of staged events were made in Caballococha. Additionally, almost all annotation work was conducted in Caballococha, as the participants responsible for that work (§3.2) lived there.

2.2 Literature and corpora Several linguists have conducted research on varieties of Ticuna, including the variety spoken in Cushillococha. Most of this research has focused on phonetics and phonology. Works on phonology include L. Anderson (1959), on the Cushillococha variety; Montes Rodriguez (1995, 2004), on the variety of San Martín de Amacayacu, Colombia; and Carvalho (2010, 2012), on the variety of Filadelfia, Amazonas, Brazil. Outside phonology, research on the language's syntax includes Soares (1992, 2000, 2017), which are based on work with speakers from various towns in Brazil, and L. Anderson (1966), again on the Cushillococha variety. Lexical and pedagogical works on the language include D. Anderson (1962), a textbook for English- and Spanish-speaking missionaries interested in learning Ticuna, and L. Anderson & D. Anderson (2017), a Ticuna-Spanish dictionary intended mainly for Peruvian Ticuna audiences.

None of the authors cited above are ethnically Ticuna, nor am I. By contrast, Abel Santos Angarita – an academically trained anthropologist and linguist – is Ticuna and a native speaker of the language. He completed a master's thesis on the dialectology of the Colombian Ticuna region (Santos Angarita 2005) and another on Ticuna theories of human origins (Santos Angarita 2013). Additionally, Santos Angarita co-authored a book of traditional narratives in Ticuna with facing Spanish translations (Goulard et al. 2016) and another on Ticuna and Kokama ethnobiology (Damaso Yoni et al. 2009).

Wilson de Lima Silva, another non-Ticuna linguist, has deposited several hours of recordings of Ticuna, collected in a diasporic community in Manaus, Brazil, in the California Language Archive (Collection 2019-12). Other than Silva's collection and the collections described here, I am not aware of any other archival deposits documenting Ticuna.

2.3 Typological Profile In terms of phonological typology, the Cushillococha variety of Ticuna combines a relatively small number of segmental phonemes with a very large number of suprasegmental contrasts. Segmentally, there are 14 surface consonants, six vowels, and two diphthongs (which behave as one syllable). Suprasegmentally, the language has eight lexical tones on monosyllables, contrastive creaky voice, and contrastive nasality on vowels. As well as lexical tone, there is grammatical tone, which marks noun class agreement on many constituents of the noun phrase and also marks clause type on verbs (Skilton 2017).

The morphological and syntactic typology displays a mixture of head- and dependent-marking characteristics. Within morphology narrowly defined, the language has an alienable/inalienable possession contrast, noun class, and inflection classes for verbs. Verbs bear proclitics that index the person, number, and noun class of the subject, as well as the clause type of the clause. Other verbal morphology is

polysynthetic, with extensive noun incorporation. However, adjuncts are licensed by oblique case markers, not by applicatives.

In terms of syntax, the language displays nominative/accusative alignment, but with pervasive differential object marking, conditioned by the argument structure of the verb and the animacy and noun class of the object. Information structure has conspicuous effects on syntax, with both topic and focus being marked by dedicated syntactic structures and both topicalization and focus movement ubiquitous in discourse. The unmarked order of constituents depends on the argument structure of the verb. Transitive and unergative verbs prefer S(O)V constituent order; unaccusative verbs usually prefer VS constituent order.

3. Project background

3.1 Project timeline & objectives I met Ticuna speakers for the first time while conducting fieldwork on a Tukanoan language, Máihiki, that is spoken elsewhere in Peru (approximately 500km from Cushillococha). While traveling to a Máihiki field site in 2014, I had a chance encounter with some Ticuna people from Cushillococha. Intrigued by the possibility of research with speakers of a language that was minimally documented, but not severely endangered, I decided to make a research visit to the town in 2015.

The research that resulted in these materials took place over ~13 months in the Cushillococha area. It can be divided into three phases.

First, between June 2015 and July 2017, I spent ~6 months in the Cushillococha region. The goal of this phase was for me to document basic structural phenomena (phonology, morphology, syntax), collect some conversational data, and learn to speak the language myself (a goal identified by participants; see below). Corpus materials from this period are mostly elicitation and texts. This phase was supported by three Oswalt grants from the Survey of California and Other Indigenous Languages.

Second, between July 2017 and August 2018, I conducted ~5 months of fieldwork focused on the language's demonstrative system. This topic was motivated by my own research interests, though participants agreed that it deserved documentation. Corpus materials from this period include psycholinguistic experiments; staged speech events, such as monolingual interviews; and naturally occurring speech events, such as conversations. There is also some elicitation. This phase was supported by a National Science Foundation Documenting Endangered Languages Doctoral Dissertation Research Improvement Grant (NSF/DEL DDRIG).

Third, between July and September 2019, I conducted ~2.5 more months of fieldwork on the acquisition of the demonstrative system by children aged 12 months to 4 years. This topic was motivated by local educators' interest in Ticuna-medium early childhood education, as well as by my continuing research interest in deixis. Corpus materials from this phase consist exclusively of recordings of child language and child-caregiver interaction, mostly from naturally occurring events. This phase was supported by the NSF/DEL DDRIG and an NSF postdoctoral fellowship.

I consulted with local Ticuna education leaders and political leaders about project objectives in all three phases. They consistently supported the project with two conditions: (1) that I should learn to speak Ticuna and communicate exclusively in it as soon as possible, and (2) that I should financially support community projects not funded by the state (for example, improvements of public buildings). I complied with the first expectation by working intensively to learn the language during the first phase, speaking it most of the time in the second phase, and speaking it exclusively in the third phase. I complied with expectations of financial support by earmarking portions of my grants for the relevant projects throughout all three phases.

Ticuna education leaders were very supportive of my interest in studying language acquisition, viewing this research as useful to local schools and early childhood programs. However, education leaders also expressed that I was **not** welcome to participate in creating pedagogical materials, as they wish to keep curriculum under the exclusive control of Indigenous people.

3.2 Project participants Apart from me, at least 163 unique people appear in the recordings in the collection and/or contributed to the annotations in the collection. These participants fall into four categories: the transcriber (1), language consultants (8), experimental/short-term participants (137), and participants in recordings of conversations and other naturally occurring events (at least 17).

Angel Bitancourt Serra was the primary transcriber, active in the project from 2016 to 2019. He collaborated with me in creating transcriptions and Spanish translations for all of the child data currently available; almost all of the data from adult conversation; and most of the data from staged events. Bitancourt Serra was extremely skilled at re-speaking audio and translating it into idiomatic Spanish but was not comfortable writing. Consequently, he contributed to transcription by re-speaking and orally translating audio while I annotated in ELAN. To recognize his central contribution to the annotations, Bitancourt Serra is named as a collection-level contributor to Collection 2018-19.

Language consultants worked with me for several months each on elicitation, as well as transcription of recordings of staged events. There were eight consultants: Deoclesio Guerrero Gómez (2015–2017), Ling Cándido Serra (2015–2016), Katia Lucero Salate Cándido (2015, 2018), Lilia Witancort Guerrero (2016–2018), Marcelo Farías Cándido (2015–2016), Sotil Suárez González (2016–2018), Yaneth Cándido Guerrero (2017–2018), and one consultant who chose not to disclose their name (2017). Before participating as the transcriber, Bitancourt Serra also participated in the project as a language consultant.

Experimental/short-term participants were people who participated on a one-time basis in an experiment, set of recordings of child-caregiver interaction, or staged event. There were 137 unique experimental/short-term participants. These included 20 adults who participated only in experiments; 104 people (57 children and 47 adults) who participated only in recordings of child-caregiver interaction; 6 adults who participated only in staged speech events; and 7 adults who participated in two or more of these activities. Additionally, 6 of the language consultants also partici-

pated in at least one experiment, set of child-caregiver recordings, or staged speech event.

The number of *participants in recordings of conversation* and other naturally occurring events is intrinsically difficult to quantify because of the fluid participation structure of these events. This said, in the 10 recordings of adult conversations and naturally occurring events that have at least five minutes transcribed, there are 36 unique adults who speak. 17 of these participants only appear on conversation recordings, and 19 also appear on at least one other recording type. However, much less than half of the naturally occurring recordings have been transcribed, and those which were transcribed were often specifically chosen because they had a relatively small number of participants. Therefore, the true number of unique adult participants on the conversation recordings is likely closer to 100 than 36.

3.3 Project products The published or publicly available research products derived from the collection are my PhD prospectus (Skilton 2017), my dissertation (Skilton 2019) and two journal articles (Skilton under review, Skilton & Peeters under review). Skilton (2017) describes the phonology and nominal morphology of the language based on elicited data. All of the other publications are about the language’s demonstrative system. Skilton (2019) is a comprehensive study of the demonstrative system based on experimental, elicited, and conversational data. Skilton (under review) argues that three of the language’s demonstratives encode information about the visibility of the referent. Skilton & Peeters (under review) is based on a psycholinguistic production experiment conducted with speakers of Ticuna and speakers of Dutch. Using exclusively quantitative analyses, it compares the effects of referent visibility and referent location on demonstrative use in Ticuna vs. in Dutch.

4. Collection organization

4.1 Collection-level organization The three collections are organized according to the type of event they represent: staged events, semi-staged events, or naturally occurring events. By “staged events”, I mean events that would not occur at all independent of research, such as elicitation, experiments, and interviews. By “naturally occurring events”, I mean events that would occur in very similar form independent of research, such as co-residents talking in their home. “Semi-staged” events are intermediate between these categories. An example of a semi-staged event represented in this collection is object play – an interaction where a caregiver and child play with a toy provided by the researcher but without the researcher’s explicit direction or involvement.

Collection 2015-06 contains data from several types of staged events involving adults only. Collection 2018-19 contains data from naturally occurring events involving adults and children, as well as data from semi-staged events with both age groups. Collection 2018-20 consists exclusively of data from psycholinguistic experiments with adults.

The three collections combined contain a total of 216 document bundles with 2,649 files. 1,760 of these files are recordings. They cover a combined 1,396 hours 15 minutes³ of recording time (1,227 hours 13 minutes of audio; 169 hours 2 minutes of video). Some recordings represent the same event as others: for example, some of the audio recordings were made simultaneously with video recordings. Excluding duplicated time (i.e., counting only audio time that does not overlap with video, and counting only one video stream per event), the recordings cover a combined 1,103 hours 58 minutes⁴ of **unique** recording time (992 hours 48 minutes of audio; 111 hours 9 minutes of video).

Complementing the recordings, the collections also include 209 files with time-aligned transcriptions in ELAN (eaf) and Praat (TextGrid) formats. The transcriptions cover a total of 33 hours 12 minutes of unique recording time (9 hours 57 minutes of audio-only recordings; 23 hours 15 minutes of video recordings) and include roughly 116,000 words of Ticuna.

Table 1 reports the total volume of audio, video, and transcript by collection.

Table 1. Volume of recordings and transcriptions in each collection, by media type.

Collection	Media	Recordings			Transcriptions	
		Count	Time	Unique Time	Count	Time
2015-06 “Ticuna elicitation and texts”	Audio	1017	542 h 4 m	528 h 14 m	108	9 h 37 m
	Video	79	16 h 0 m	16 h 0 m	9	3 h 47 m
2018-19 “Ticuna conversations”	Audio	351	659 h 23 m	461 h 6 m	1	0 h 20 m
	Video	235	137 h 25 m	79 h 32 m	91	19 h 28 m
2018-20 “Ticuna experiments”	Audio	32	25 h 47 m	3 h 28 m	0	n/a
	Video	46	15 h 37 m	15 h 37 m	0	n/a

As metadata, each collection includes a text finding aid accessible from the collection landing page. The finding aids provide detailed file-level metadata about every file in the collection, including the date, location, participants, and topics of every recording. Because the CLA does not have the capacity to include file-level metadata, finding aids were the only way to include metadata about individual files.

Besides the collection-level finding aids, less detailed metadata about each bundle – including location, date, participants, and data type, but excluding recording contents – appears in the California Language Archive’s bundle-level metadata tool. Each bundle also contains a text finding aid which provides file-level metadata about its contents.

4.2 Recording vs. transcription bundles Across all three collections, recordings and their transcriptions are archived in separate bundles. This was necessary for incremental archiving. If recordings and transcripts had been placed in the same bundles, the

³Equivalent to 58 days, 4 hours, 15 minutes.

⁴Equivalent to 45 days, 23 hours, 58 minutes.

CLA – which is unable to edit accessioned bundles – would have been unable to accession any bundles until all of the transcripts were complete. Waiting for completion of transcripts would have delayed the accession of many media bundles by up to two years. To avoid this delay, I organized recordings and transcripts separately.

While the use of separate bundles for recordings and transcriptions made incremental archiving possible, it required the use of extra metadata to relate recordings and their transcriptions. This metadata appears in two places.

First, the ‘Description’ field of the metadata for every recording bundle includes text stating which files in the bundle have been transcribed and where the transcriptions are located. For example, the bundle-level metadata for Bundle 2015-06.021,⁵ a recording bundle, states that transcriptions of three recordings in the bundle (named in the metadata) appear in Bundle 2015-06.027, while the transcription of another recording appears in Bundle 2015-06.065. To determine where transcriptions of a given recording are located, I encourage users to **consult the Description field first**.

Second, the ‘Relations’ field of the metadata for every recording bundle lists the bundles which contain transcriptions of the bundle’s files. This field also provides active links to the transcription bundles. It was not possible to establish links between individual recording files and their transcriptions because, as mentioned, the CLA is unable to store file-level metadata.

4.3 Organization of recording bundles While recordings and transcriptions are archived separately in all three collections, the internal structure of recording bundles varies between the collections. Because Collection 2015-06 is primarily a collection of elicitation sessions, it follows a one **participant**/one year/one bundle structure, rather than the one recording event/one bundle file structure more often used in language archives. Similarly, the child language materials in Collection 2018-19 also follow a one participant/one bundle structure.

The one participant/one bundle structure was chosen to facilitate comparison across recordings of the same individual: for example, it ensures that all of a consultant’s elicitation sessions on a given topic appear in the same bundle. Collection 2019-20 and the adult language materials in Collection 2018-19, in contrast, do not generally have repeated observations of the same participants. Therefore, they follow a one recording event/one bundle file structure.

5. Access and ethics The collections are open access. Currently, the only access restriction is that Bundles 2018-19.86 to 2018-19.124, which consist of transcriptions that have not yet been reviewed by language consultants, are not public. These bundles will become public by fall 2021, at which point the collection will be fully open access.

Because of the large volume of video recordings and recordings of children in these collections, open access to the collections presents some unique ethical concerns. Therefore, I discuss the informed consent and open access processes underlying the

⁵<http://cla.berkeley.edu/item/23755>.

collections in detail. My discussion includes the content of the informed consent process, including informed consent to open-access archiving (§5.1); participants' level of comfort with video recording (§5.2); and issues related to the archiving of child language (§5.3).

At an administrative level of ethics, study procedures were approved by the Institutional Review Boards of the University of California Berkeley and the University of Texas at Austin. Because the research involved recordings of non-public behavior (i.e., family members talking at home) and of children, both IRBs required non-exempt review of the procedures. Researchers planning similar projects should be aware that they will likely require non-exempt review, which involves much more extensive interaction with the IRB than exempt review.

5.1 Informed consent process All adults who participated in the recordings provided informed consent for them to be archived on an open-access basis. During the informed consent process, as well as sharing standard information about the research project and purpose of recordings, I also explained that I was interested in sharing the recordings in an online archive. I stated that the online archive was similar to a library,⁶ and that anyone with an internet connection – no matter their location – would be able to view all archived recordings, unless they were restricted. Per standard ethics, I offered people the option to participate in the project but decline recording; to participate in recordings but decline archiving; and to allow recordings to be archived but restrict them. I also offered participants the choice of whether to disclose their names in the archive.

All of the participants agreed to archiving of recordings, and no one asked for their recordings to be restricted. Several participants did decline to disclose their names in the archive, and others did not express a clear preference about disclosure of their names. In the CLA metadata, these participants are identified as “Anonymous (Ticuna)” followed by a unique identifying number. In the file names and finding aids, they are identified by alphabetical codes. Despite the use of the code “Anonymous”, these participants are technically not “anonymous”, since audio-video recordings are intrinsically identifiable (§5.3). In light of this, when participants declined to disclose their names, I confirmed that they still consented to archiving; understood that others would be able to view their video and audio; and did not want to restrict their recordings.

I typically conducted informed consent at least one day before I recorded a participant or household for the first time. I see the use of separate encounters for consent vs. for recording as a best practice for two reasons. First, it allows participants more time to make their decisions. During the consent encounter, participants can ask questions without the time pressure of beginning a recording and experience less pressure to say yes than if I am already present with my equipment kit, about to record. Second, this practice affords participants a second opportunity to decline recording. If people do not want to be recorded – but feel uncomfortable rejecting my offer during

⁶The Cushillococha/Caballococha region has a public library, several school computer labs, and (limited) internet access. Therefore, participants were familiar with libraries and the internet (cf. Robinson 2010).

the consent encounter – then they can (and sometimes did) simply decide not to be home when I return for the recording encounter.

In sum, participants agreed to open-access archiving with a full understanding of the implications for others' ability to view recordings. This is clear not only because of the informed consent process, but also because, in recordings of conversation, participants frequently displayed their understanding by telling co-participants that I would later show the video to others (cf. Dingemanse 2011: 11–13).

5.2 Responses to video recording In contrast to participants' openness to video in this project, some language documentation researchers working in the Indigenous Americas and Australia have found that speakers do not wish to have video recordings archived (Meakins et al. 2018: 231) or do not want to participate in video recording at all (Bower 2008: 25).

Given this context, I want to acknowledge that – while most prospective participants in this project were open to being video-recorded – I also met with several people who wanted to participate, but did not want to be recorded on video, even if the recordings were restricted. When people declined video recording, I suggested that they could instead contribute to the project through activities which did not require video, such as participating in elicitation, recording texts, and collaborating in transcription. Three people accepted this offer, and two of them – Lilia Witancort Guerrero and a person who chose not to disclose their name – became long-term language consultants.

Beyond recruitment of individual experimental participants and language consultants, I also recruited entire households to participate in recordings of conversation and child-caregiver interaction. Here as well, perhaps 10 of the ~50 households that I approached declined to participate, though it is not clear to me whether they declined specifically because of the inclusion of video. Other households were interested in being recorded doing some activities (e.g., on inter-household social visits) but unwilling to be recorded in other contexts.

Most video-based research on language in interaction encounters a similar pattern of participation to this (Rossi et al. 2020). That is, different people have different levels of comfort with being recorded doing their everyday activities, and people's willingness to be recorded varies over time and between social settings. Moreover, privacy – as relevant to research on language in interaction – is an individual- and culture-specific concept. Activities that are appropriate to record in one social setting might be deeply inappropriate in another. For example, in my research, Ticuna-speaking participants almost never consented to be recorded while eating (cf. D. Anderson 1962: 90 on eating as a private activity in Ticuna culture). In contrast, eating and drinking events are some of the best contexts for recording social interaction in Western societies (Enfield et al. 2010: 2617–2618).

Because privacy is culturally specific, anyone who is doing research away from home needs to invest time and ethnographic effort to learn where people are comfortable being recorded (Enfield 2014: 977–978). During this research, I learned about the settings where recording was vs. was not appropriate through trial and error –

i.e., by approaching people who then declined to be recorded – but also through participant observation. I spent four months living in the Cushillococha area before ever making a video recording, and two more months before I began regularly recording video of interactions in homes. Thus, by the time that I began recording video, I already had relatively long-term relationships with many of the people that I recorded. Rossi et al. (2020: 11) argue that such longitudinal relationships are a prerequisite for successfully collecting data on language use in interaction.

5.3 Open access and recordings of children Because children cannot provide informed consent, research on child language raises unique ethical issues. These issues include whether to archive recordings of children at all; whether to include video in the archive; and how to identify children in archival metadata.

In this project, I archived recordings of children, on an open-access basis, if the children’s caregivers provided informed consent. As with the recordings of adults, all caregivers provided informed consent to open archiving of the data, and no one chose to restrict their recordings.

As well as archiving audio, I also requested and obtained consent to archive video of children. This reflected that video was essential to the research. A major purpose of my child language documentation was to analyze children’s development of deictic visual behaviors, such as pointing and following adults’ gaze. Analyzing visual behavior requires video; likewise, conducting transparent research about the topic requires sharing video with other researchers (e.g., so that they can evaluate the adequacy of coding schemes). These considerations motivated me to request consent to archive video of children and to complete the unique IRB process required for archiving child video data.

While video was essential to the child language research, identifying the children by name was not. Therefore, in order to protect the children’s privacy, I redacted the names of children and caregivers from the child language materials. Instead, children in the 2019 study are identified exclusively by participant numbers, and children in the 2018 pilot study are identified by three-character alphabetical codes. Caregivers are identified by their relationship to the children, since identifying them by name would also identify the children.

Non-disclosure of children and caregivers’ names was required by the IRB. From a practical perspective, of course, redacting names does not fully de-identify participants. Especially in small communities, audio and video recordings are intrinsically identifiable, no matter what efforts we make to redact them (Meakins et al. 2018: 192–193). Since participants are aware of this, when they consent to recording, they also consent to some degree of identifiability.

6. Collection 2015-06: Staged events with adults

6.1 Collection contents Collection 2015-06, “Ticuna elicitation and texts”, contains exclusively recordings of staged events involving adults and annotations of those

recordings. The main participants in this collection are the nine language consultants. To facilitate topic search and incremental archiving, recording bundles in this collection have a one participant/one year/one bundle structure (§4).

In terms of volume, Collection 2015-06 includes 65 bundles. 39 bundles contain recordings; they include 1096 recording files with total time 558 hours 4 minutes⁷ (audio: 542 hours 4 minutes, video: 16 hours 0 minutes). Six bundles contain transcriptions; there are 117 unique transcription files, with total transcribed time 13 hours 24 minutes (audio: 9 hours 37 minutes, video: 3 hours 47 minutes). Table 2 reports the count, total time, and transcribed time of recording files in Collection 2015-06 by data type.

Table 2. Volume of recordings and transcriptions in Collection 2015-06, by data and media type.

Data Type	Media	Recordings			Transcriptions	
		Count	Time	Unique Time	Count	Time
Elicitation	Audio	906	529 h 10 m	519 h 12 m	34	4 h 13 m
	Video	55	11 h 18 m	11 h 18 m	0	n/a
Monolingual Interviews	Audio	14	5 h 51 m	2 h 0 m	0	n/a
	Video	24	4 h 42 m	4 h 42 m	9	3 h 47 m
Text	Audio	97	7 h 3 m	7 h 3 m	74	5 h 24 m
	Video	0	n/a		0	n/a

Event types represented in Collection 2015-06 include elicitation sessions, monologues by participants (texts), and monolingual interviews between me and a participant. Below, I discuss each of these event types in turn.

The **elicitation sessions** in Collection 2015-06 cover a range of phonological, morphosyntactic, and semantic/pragmatic topics. Elicitation sessions were conducted bilingually in Spanish and Ticuna. While I often posed questions in Ticuna, especially in later phases of research, there is Spanish on every recording. Many elicitation sessions involved batteries of stimuli produced by the former Language & Cognition department of the Max Planck Institute for Psycholinguistics. Finding aids for recordings of these sessions include citations of the stimuli.

Texts in Collection 2015-06 include procedural texts, traditional narratives/folktales, personal and oral history narratives, responses to targeted construction storyboards, responses to questions designed to elicit epistemic or moral reasoning (e.g. ‘What will happen this rainy season?’ or ‘How should young men behave?’), and vernacular definitions of content words. Texts were usually recorded in Ticuna only.

(1) provides an excerpt from one of the texts. The author of this text is Ling Cándido Serra, a 41-year-old high school teacher and language activist; he died of COVID-19 in May 2020. In (1), Ling, who was heavily involved in the defense of Cushillococha’s natural resources against incursions by non-Indigenous people, de-

⁷Equivalent to 25 days, 14 hours, 8 minutes.

scribes actions that political leaders are taking to conserve fish stocks in the town's waters. In lines 1–4, he states that fish stocks have fallen, naming three economically important species which are no longer present in Cushillococha's waters. In lines 5–7, Ling describes the first of several actions which Cushillocochanos have taken to conserve fish – banning outsiders from fishing in the territory.

(1) Excerpt from a text.

(Media reference: Bundle 2015-06.012,⁸
tca_20160630_lcs_ahs_lts.wav, 0:31–0:52)

(Transcript reference: Bundle 2015-06.027,⁹
tca_20160630_lcs_ahs_lts_ahs.TextGrid)

<example1_audio.wav>

- (1) *ma³ na⁴ ta² lu² ma³ i⁴ tfo⁴³ ni⁵ gi⁴ i⁴ ta³¹ fi⁴,*
ma³ ri³ na⁴ = ta² lu² ma³ i⁴ tfo⁴³ ni⁵ = gi⁴ i⁴ ta⁴³ = fi⁴
 PERF 3SBJ= NEG.EXIST LNK(IV) fish(IV) =PL LNK(IV) big =NMLZ(IV)
 There aren't large fish anymore,

- (2) *i⁴ de⁴ ti⁴ fi⁴, to³ ma³ ka³ ti⁴ fi¹*
i⁴ de⁴ ti⁴ fi⁴ to³ ma³ ka³ ti⁴ fi¹
 LNK(IV) A.gigas C.macropomum
 such as de⁴ ti⁴ fi⁴ (*Arapaima gigas*) and to³ ma³ ka³ ti⁴ fi¹ (*Colossoma macropomum*).

- (3) *ta³¹ fi⁴ i⁴ ba⁴ we⁴ gi⁴ ma³ na⁴ ta² lu² ma³*
ta⁴³ = fi⁴ i⁴ ba⁴ we⁴ = gi⁴ ma³ ri³ na⁴ = ta² lu² ma³
 be.big =NMLZ(IV) LNK(IV) P.expansa =PL PERF 3SBJ= NEG.EXIST
 There aren't big ba⁴ we⁴ turtles (*Podocnemis expansa*) anymore.

- (4) *ta⁴ ma³ ai² ki⁴ ma⁴ ni³¹ fi³ ta⁴ dau².*
ta⁴ ma³ ai² ki⁴ ma⁴ ni³¹ = fi³ ta⁴ = dau²
 NEG truly 3 =ACC 1EXCL.SBJ= see
 We really don't see them.

⁸<http://cla.berkeley.edu/item/23746>.

⁹<http://cla.berkeley.edu/item/23761>.

- (5) *ŋe³ma²kā¹ ɲu¹ma⁵ ri¹ to⁵ri³ ā³ē¹ga³ki³ ri¹ na⁴na³tʃu³¹ʔu³*
ŋe³ma²kā¹ ɲu¹ma⁵ ri¹ to⁵ri³ ā³ē¹ga³ki³ ri¹ na⁴= na³=
 CONN now TOP 1EXCL.AL.POSS political.leader TOP 3SBJ= 3OBJ=
tʃu³¹ʔ
 withhold
 For that reason, currently our (exclusive) leaders prohibit
- (6) *na⁴ʔ pi³tʃa¹ma⁴ ā² na¹tʃo⁴³ku²ʔi⁴ i⁴ to¹ā⁴ma⁴ i⁴ i³¹a¹ne¹wa⁵ i¹ʔi³¹ʔi⁴ i⁴ du¹ʔi³ʔi⁴gi⁴,*
na⁴ʔ pi³tʃa¹ =ma⁴ā² na¹= tʃo⁴³ku² =ʔi⁴ i⁴ to¹ā⁴ma⁴ i⁴
 COMP net =COM 3SBJ.SC= enter:PLSBJ =SUB LNK(IV) other(IV) LNK(IV)
i³¹a¹ne¹ =wa⁵ i¹= i⁴³ =ʔi⁴ i⁴ du¹ʔi³ʔi⁴ =gi⁴
 town(IV) =ALL 3SBJ.SC= go:PLSBJ =NMLZ(IV) LNK(IV) person(IV) =PL
 (They prohibit) people who come from other towns from entering (our waters) with
 nets.
- (7) *e³ga⁴ mu²ki³gu² ri¹ ɲu¹ʔma⁵tʃi⁵ ɲu¹ʔma⁵ri⁵ʔi¹ i⁴ a¹ʔe¹a¹ne¹gu².*
e³ga⁴ mu²ki³ =gu² ri¹ ɲu¹ʔma⁵tʃi⁵ ɲu¹ʔma⁵ =ri⁵ʔi¹ i⁴ a¹=
 COND wet.season =LOC and also now =like LNK(IV) 3SBJ.SC=
e¹a¹ne¹ =gu²
 fall(water.level) =COND
 (They prohibit it) in the wet season, and also (at times) like now, when the water level
 is falling (i.e., in the dry season).

Texts like the excerpt in (1) are familiar to many documentary linguists. By contrast, the **monolingual interviews** represent a genre less familiar in the field.

These interviews were conducted to collect data on spatial language and gesture. All focus on description of a locality or built space. In every monolingual interview, I am the interviewer. I ask the interviewee questions about how their locality (the Cushillococha region in general, and the neighborhood where they live) has changed over time. I also ask them to describe the current and historical locations of landmarks. In the built space description interviews, I visit a built space, such as a construction site, with one of the space's owners. I ask the owner questions about the layout of the space and plans for changing it through construction. The monolingual interviews differ from elicitation sessions in that they do not contain any metalinguistic discussion.

(2) is an excerpt from a monolingual interview. The interviewee is Lucinda Gómez Cordero, a 58-year-old woman with a subsistence farming/gathering lifestyle. We have been discussing the location of her agricultural fields and wild-gathering places. At the beginning of the excerpt, I ask where her husband goes to fish (normatively, Ticuna women do not go fishing). She mentions two locations, then – at my request – names some fish species found at each one.

(2) Excerpt from a monolingual interview.

(Media reference: Bundle 2015-06.069,¹⁰
tca_20180705_lgc_ahs_haldi_archive.mp4, 15:45–16:22)

(Transcript reference: Bundle 2015-06.066,¹¹
tca_20180705_lgc_ahs_haldi.eaf)

<example2_video.mp4>

- (1) AHS: $\eta e^1 \eta ta^5 \quad ni^{41} \eta \tilde{r}^4 \quad tfo^{43} \eta ni^5 \quad ka^1 \quad ta^4 \eta \tilde{r}^{43}?$
 $\eta e^1 \eta ta^5 \quad ni^{41} \eta \tilde{r}^4 \quad tfo^{43} \eta ni^5 = ka^1 \quad ta^4 = \quad \tilde{r}^{43}$
 where FOC fish(IV) =PURP 3(I)SBJ= go
 Where does he (= your husband) go to fish?
- (2) LGC: $nu^5 a^2 ta^2 \tilde{a}^4, \quad pa^2 \quad tfa u^1 e^3 \eta \tilde{r}^4, \quad Ki^3 \eta \tilde{r}^3 tu^1 wa^5 ta^2 \tilde{a}^4, \quad nu^5 a^2 ta^2 \tilde{a}^4.$
 $nu^5 a^2 = ta^2 \tilde{a}^4, \quad pa^2 \quad tfa u^1 = e^3 \eta \tilde{r}^4 \quad Ki^3 \eta \tilde{r}^3 tu^1 = wa^5 = ta^2 \tilde{a}^4 \quad nu^5 a^2 = ta^2 \tilde{a}^4$
 here =only VOC 1SG =sister Cushillococha =all =only here =only
 Right here, “sister,” right here in (the lake of) Cushillococha.
- (3) LGC: $\eta u^1 \eta gu^2 \tilde{a}^1 ki^2 \quad i^4 \quad je^5 a^2 \tilde{a}^4 ma^4 \quad ta^1 \eta \tilde{r}^2 \eta \tilde{r}^4.$
 $\eta u^1 \eta gu^2 = \tilde{a}^1 ki^2 \quad i^4 \quad je^5 a^2 = \tilde{a}^4 ma^4 \quad ta^1 = \quad \tilde{a}^1 = \eta \tilde{r}^4$
 when? =ADVBZ LNK(IV) there = $\tilde{A}^4 MA^4$ 3(I)SBJ.SC= row =SUB
 Once in a while he rows wa-a-ay over there.
- (4) LGC: $to^1 ku^3 ti^3 a^1 ne^1 wa^5 \quad ta^4 \eta \tilde{r}^2 \eta \tilde{r}^4 \quad i^4 \quad \eta u^1 \eta gu^2 \tilde{a}^1 ki^2.$
 $to^1 ku^3 ti^3 = a^1 ne^1 = wa^5 \quad ta^4 = \quad \tilde{a}^1 = \eta \tilde{r}^4 \quad i^4 \quad \eta u^1 \eta gu^2 = \tilde{a}^1 ki^2$
 other.bank =land =all 3(I)SBJ.SC= row =SUB LNK(IV) when? =ADVBZ
 He rows to the other side (of the lake), once in a while.
- (5) AHS: $ri^1, \quad ri^1 \quad \eta e^3 ma^2 \quad tfo^{43} \eta ni^5, \quad na^4 ta^4,$
 $ri^1 \quad \eta e^3 ma^2 \quad tfo^{43} \eta ni^5 \quad na^4 = \quad ta^4$
 and that.ANA fish(IV) 3SBJ= big
 And the fish, are they big,

¹⁰<http://cla.berkeley.edu/item/25708>.

¹¹<http://cla.berkeley.edu/item/25436>.

- (6) AHS: $ri^3\tilde{e}^1\eta na^5 na^4 mu^4, ri^3\tilde{e}^1\eta na^5 na^4 no^{31} re^5 ta^2 ma^4?$
 $ri^3\tilde{e}^1\eta na^5 na^4 = mu^4 ri^3\tilde{e}^1\eta na^5 na^4 = no^{31} re^5 = ta^2 ma^4$
 ALT 3SBJ= many ALT 3SBJ= few =only
 Are there many of them, or are there just a few?
- (7) LGC: $na^4 - na^4 \eta e^3 ma^2 i^4 tfo^{43} \eta ni^5$.
 $na^4 = \eta e^3 ma^2 i^4 tfo^{43} \eta ni^5$
 3SBJ= exist(IV) LNK(IV) fish(IV)
 Th- there's fish.
- (8) LGC: $tu^3 ku^3 na^4 ri^1, tu^3 ku^3 na^4 ri^1 wa^5 ta^1 \tilde{u}^{43} gu^2 ri^1$,
 $tu^3 ku^3 na^4 ri^1 tu^3 ku^3 na^4 ri^1 = wa^5 ta^1 = \tilde{u}^{43} = gu^2 ri^1$
 Cichla.sp Cichla.sp =ALL 3(I)SBJ.SC= go =SUB TOP
 There's $tu^3 ku^3 na^4 ri^1$ (Cichla sp.), when he goes to get $tu^3 ku^3 na^4 ri^1$,
- (9) LGC: $tu^3 ku^3 na^4 ri^1 ma^4 \tilde{a}^2 to^{31} \eta ka^1 i^5 ta^4 \eta u^3$.
 $tu^3 ku^3 na^4 ri^1 = ma^4 \tilde{a}^2 to^{31} = \eta ka^1 i^5 = ta^4 = \eta u^3$
 Cichla.sp =COM 1EXCL =PURP ASP= 3(I)SBJ= arrive
 He comes back with $tu^3 ku^3 na^4 ri^1$ for us (exclusive).
- (10) LGC: $ju^3 ta^3 wa^5 ta^1 \tilde{u}^{43} gu^2$,
 $ju^3 ta^3 = wa^5 ta^1 = \tilde{u}^{43} = gu^2$
 Z.zungaro =ALL 3(I)SBJ.SC= go =SUB
 When he goes to get $ju^3 ta^3$ (Zungaro zungaro),
- (11) LGC: $na^4 ju^3 ta^3 ma^4 \tilde{a}^2 to^{31} \eta ka^1 i^5 ta^4 \eta u^3$.
 $na^4 ju^3 ta^3 = ma^4 \tilde{a}^2 to^{31} = \eta ka^1 i^5 = ta^4 = \eta u^3$
 COMP Z.zungaro =COM 1EXCL =PURP ASP= 3(I)SBJ= arrive
 He comes back with $ju^3 ta^3$ for us (i.e. he's successful).
- (12) LGC: $ta^1 tfu^{31} \eta ne^3 tfa^1 \tilde{r}^1 ma^3 re^3 gu^2, na^4 tfo^{43} \eta ni^5 i^4 i^5 ra^1 gi^5 \eta$,
 $ta^4 = tfu^{31} \eta ne^3 = tfa^1 \tilde{r}^1 = ma^3 re^3 = gu^2 na^4 tfo^{43} \eta ni^5 i^4 i^5 ra^1$
 3(I)SBJ= bow.fish =PROSP =just =SUB COMP fish(IV) LNK(IV) small
 $= gi^4 = \eta$
 =PL =NMLZ(IV)
 When he just goes bow-fishing, (then it's) small fish,

- (13) LGC: $t\hat{f}i^3ri^3ma^4ta^1ri^1, a^3ra^3wi^3ri^5ri^1, pa^3ku^1gi^4ma^4\tilde{a}^2i^5ta^4\eta u^3$.
 $t\hat{f}i^3ri^3ma^4ta^1ri^1 a^3ra^3wi^3ri^5ri^1 pa^3ku^1 =gi^4 =ma^4\tilde{a}^2 i^5 =$
P.nigricans and Triportheus.sp and Mylossoma.sp =PL =COM ASP=
 $ta^4 = \eta u^3$
3(I)SBJ= arrive
He comes back with $t\hat{f}i^3ri^3ma^4ta^1$ (*Prochilodus nigricans*), $a^3ra^3wi^3ri^5$ (*Triportheus* sp.), and pa^3ku^1 (*Mylossoma* spp.).

As well as the 39 recording bundles, Collection 2015-06 also includes 26 other bundles. 18 of the 26 non-recording bundles contain primary data other than recordings. This includes transcriptions; copies of published texts (from D. Anderson 1962 and pedagogical materials) annotated with language consultants' comments; scanned and born-digital fieldnotes on elicitation sessions; and copies of elicitation stimuli and plans for elicitation sessions. An additional 7 bundles contain secondary data, such as spreadsheets created to track data collection and field reports analyzing the data.

While handwritten transcriptions in Collection 2015-06 use IPA, born-digital files, such as TextGrids, use a practical orthography which I developed to type the language's many suprasegmental features without using diacritics. The final of the 26 non-recording bundles (2015-06.85) contains a key to the diacritic-free orthography.

6.2 Recording Methods Audio recordings of staged events were made using a Zoom H4N audio recorder sampling at 44.1kHz and writing to WAV format. Depending on the level of environmental noise and the purpose of the recording, the Zoom H4N was attached to either an ATH-M30 lapel microphone or a Shure SM10 headworn microphone. Video recordings of staged events were made using a Sony PJR540 video camera (with no external microphone) in 2017 and using a Canon XA30 video camera connected to a Rode NT4 stereo microphone in 2018. In a few video recordings, I used an Olympus VP10 audio recorder worn on the consultant's body to provide a backup audio track to the camera audio. Olympus VP10 recorders sampled at their maximum rate of 22.05kHz and wrote to WAV format. Video was included for staged events when visual aspects of the event were directly relevant to research questions. For example, because deictic demonstratives are tightly coupled with deictic gestures, elicitation sessions about deixis were video-recorded. Otherwise, staged events were recorded with audio only.

Recordings of staged events were annotated in Praat in 2015, 2016, and 2017, and in ELAN in 2017 and 2018. The annotations are all time-aligned and include transcriptions, translations into Spanish, and notes written in English and Spanish. Notes annotations include both explanatory notes by consultants and notes on linguistic phenomena by me.

7. Collection 2018-19: Naturally occurring and semi-staged events with adults and children

7.1 Collection contents Collection 2018-19, “Ticuna conversations”, consists of recordings of naturally occurring events, as well as semi-staged events. Recall from §4.1 that “semi-staged events” are events that would not have occurred without my intervention, but also do not involve me as a central participant. “Naturally occurring events”, by contrast, are events that would have occurred in similar form independent of research.

Collection 2018-19 contains a total of 125 file bundles. Of these, 29 file bundles document adult language and 96 bundles document child language. I describe the adult and child components of the collection separately.

7.2 Adult language recordings and transcripts 29 of the 125 bundles in Collection 2018-19 document primarily adult language use. This set is comprised of 27 recording bundles and 2 non-recording bundles.

The 27 adult recording bundles consist of 24 file bundles with recordings of naturally occurring events and 3 file bundles with recordings of staged conversations. Each recording file bundle contains audio and video recordings from one speech event; thus, this part of the collection has a one recording event/one bundle organization.

In terms of volume, the adult recording bundles in Collection 2018-19 contain 123 recordings, with total recording time 37 hours 45 minutes (audio: 20 hours 1 minute, video: 17 hours 44 minutes). The adult bundles also contain 18 transcription files, with total time 3 hours 41 minutes (audio only: 0 hours 20 minutes, audio-video: 3 hours 21 minutes). Table 3 shows the count, total time, and transcribed time of adult files in Collection 2018-19 by data type.

Table 3. Volume of recordings and transcriptions of adult language in Collection 2018-19, by data and media type.

Data Type	Media	Recordings			Transcriptions	
		Count	Time	Unique Time	Count	Time
Naturally Occurring	Audio	38	15 h 42 m	5 h 18 m	1	0 h 20 m
	Video	70	15 h 15 m	15 h 15 m	15	2 h 25 m
Staged Conversation	Audio	7	4 h 18 m	0 h 0 m	0	n/a
	Video	8	2 h 30 m	2 h 30 m	2	0 h 56 m

Most of the **naturally occurring events** that I recorded (81 of 108 files) were maximally informal conversations – that is, conversations where all participants knew one another well, taking place in the participants’ homes. Examples of maximally informal conversations in the collection include: a father and son chatting while they process wild cane to make arrows, briefly joined by female relatives (2018-19.5); a group of 6–10 closely related adults chatting while waiting for a meal to be ready (2018-19.8); and a group of 3–5 adults with parent-child and in-law relations chatting on a social visit (bundle 2018-19.24).

(3) is an excerpt from one recording of a maximally informal conversation. Teodor Guerrero Coello Sr., left, and his son Teodor Guerrero Coello Jr., right, are processing cane stalks into arrow shafts. Lesli Guerrero Coello (Teodor Sr.'s daughter and Teodor Jr.'s sister) enters from the left and asks Teodor Sr. a series of questions.

(3) Excerpt from a recording of maximally informal conversation.

(Media reference: 2018-19.005,¹²

tca_20180622_disc_video_001_archive.mp4, 2:50–3:22)

(Transcript reference: 2018-19.020,¹³

tca_20180622_disc_video_001.eaf)

<example3_video.mp4>

- (1) Lesli: $t\dot{a}^2\dot{\lambda}^4k\dot{i}^4$ $ni^{41}\dot{\lambda}^4$ $pa^4e^2\dot{\lambda}^2\dot{\lambda}^4$, pa^2 $\dot{o}^1\dot{?}$
 $t\dot{a}^2\dot{\lambda}^4k\dot{i}^4$ $ni^{41}\dot{\lambda}^4$ $pa^4e^2\dot{\lambda}^2$ = $\dot{\lambda}^4$ pa^2 $\dot{o}^1\dot{?}$
 what? FOC play.music =SUB VOC grandfather.VOC
 Who's playing music, grandpa?
- (2) Teodor Jr.: $no^5\dot{\lambda}^3$ jo^3ra^5
 $no^5\dot{\lambda}^3$ jo^3ra^5
 3.AL.POSS owner
 Its owner.
- (3) Teodor Sr.: $no^5\dot{\lambda}^3$ jo^3ra^5 $g\dot{i}^4$, $da^{31}a^1$ $\dot{\lambda}^4\dot{?}$, $\dot{\lambda}^3\dot{\epsilon}^1\dot{\lambda}^5$ $t\dot{a}^2\dot{\lambda}^4k\dot{i}^4\dot{?}$
 $no^5\dot{\lambda}^3$ jo^3ra^5 = $g\dot{i}^4$ $da^{31}a^1$ $\dot{\lambda}^4\dot{?}$ $\dot{\lambda}^3\dot{\epsilon}^1\dot{\lambda}^5$ $t\dot{a}^2\dot{\lambda}^4k\dot{i}^4\dot{?}$
 3.AL.POSS owner =PL this(III) house(III) ALT what?
 The owners of this house, who else (could it be)?
- (4) Lesli: $\eta e^1\dot{\lambda}^1$ ja^1 $Ni^3\dot{\lambda}^5ma^5a^1\dot{\lambda}^3$, pa^2 $\dot{o}^1\dot{?}$
 $\eta e^1\dot{\lambda}^1$ ja^1 $Ni^3\dot{\lambda}^5ma^5$ = $a^1\dot{\lambda}^3$ pa^2 $\dot{o}^1\dot{?}$
 where.is(III) LNK(III) name =AL.POSS VOC grandfather.VOC
 Where's $Ni^3\dot{\lambda}^5ma^5$'s (arrow cane), grandpa?

¹²<http://cla.berkeley.edu/item/25466>.

¹³<http://cla.berkeley.edu/item/25710>.

- (5) Teodor Sr.: $\tilde{e}^1\eta na^5 \eta e^3ma^2ka^1 ni^{41}\eta^4$, “ $\eta e^1\eta^4, \eta e^1\eta^4 a^4 Ni^3\eta ma^5?$ ”
 $\tilde{e}^1\eta na^5 \eta e^3ma^2ka^1 ni^{41}\eta^4 \eta e^1\eta^4 a^4 Ni^3\eta ma^5$
 ALT CONN FOC where.is(IV) LNK(IV) NAME
 Actually, that’s why (I was saying,) “Where is he? Where’s Ni³ηma⁵?”
- (6) Teodor Jr.: *Galilea=gu²*
 Galilea =gu²
 place.name =LOC
 (He’s) in Galilea (another town).
- (7) Teodor Sr.: “*tau⁴ ni³¹gi¹η⁵ ne⁵ ja³na³gu²?*” *na³tfa³?*
 tau⁴ ni³¹ =gi¹ =η⁵ ne⁵ ja³= na³= gu² na³tfa³?
 NEG.FUT 3 =REFL =IBEN some(III) AM= 3OBJ= process.in.fire 1SG.QUOT
 I said, “Why doesn’t he come and process some for himself?”
- (8) Lesli: *ni³¹η⁵ ne⁵i¹gu²*.
 ni³¹ =η⁵ ne⁵= i¹= gu²
 3 =IBEN some(III) IMP= process.in.fire
 Process it for him.
- (9) Lesli: *na⁴³a³ri¹ ta⁴ ku³¹ri³ tfo⁴³ηni⁵ ma⁴ā² i⁵ na⁴ηu³*.
 na⁴³a³ri¹ ta⁴ ku³¹ri³ tfo⁴³ηni⁵ =ma⁴ā² i⁵= na⁴= ηu³
 CONN FUT 2SG.AL.POSS fish =COM ASP= 3SBJ= arrive
 Because he’s going to come back with fish for you.
- (10) Teodor: [Sp] *cuidao*
 [Spanish] Watch out (= don’t say that).
- (11) Lesli: *ηe³ma²gu² nu⁵a² ti³¹η³ ki³ηū⁴³ηē⁴ē³*.
 ηe³ma² =gu² nu⁵a² ti³¹ =η³ ki³= ū⁴³ -ηē⁴ē³
 that.ANA =LOC here 3(I) =ACC 2SG.SBJ= run.away -CAUS
 That’s how you made him run off from here (i.e. by not being helpful).¹⁴

¹⁴The verb root η^{43} can mean ‘come/go’, ‘flee/run away’, or ‘walk’, depending on the form of its subject and aspectual proclitics. This token has a subject proclitic which gives the reading ‘flee’, while the tokens of the verb in (1) and (2) had proclitics which give the reading ‘come/go’.

- (12) Teodor Jr.:
- Galilea=gu*
- ²

Galilea =gu²
 place.name =LOC
 (He's) in Galilea.

- (13) Teodor Sr.:
- Galilea. Ki*
- ³
- ,
- ma*
- ³
- ri*
- ³
- na*
- ³¹
- tʃi*
- ¹
- ga*
- ³
- je*
- ⁵
- a*
- ²
- tʃi*
- ¹
- de*
- ⁴³
- ʔa*
- ²
- ʔ*
- ,

Galilea ki³ ma³ri³ na³¹ =tʃi¹ga³ je⁵a² tʃi¹= de⁴³ʔa² =ʔ
 place.name INTJ PERF 3 =about there.DIST 1SG.SBJ.SC= talk =SUB
 Galilea. Well, there I was talking about it,

- (14) Teodor Sr.:
- na*
- ⁴
- ʔe*
- ³
- ma*
- ²
- ɲu*
- ¹
- ʔma*
- ⁵
- ,
- de*
- ⁴
- ʔne*
- ¹
- wa*
- ⁵
- na*
- ⁴
- ta*
- ⁴
- ʔi*
- ³¹
- ʔ*
- .

na⁴ ʔe³ma² ɲu¹ʔma⁵ de⁴ʔne¹ =wa⁵ na⁴ ta⁴= i³¹
 COMP that.ANA now cane =ALL COMP 1EXCL.SBJ.SC= go:PLSBJ
 =ʔ
 =SUB

Like now, about how we were going to gather cane,

- (15) Teodor Sr.:
- wi*
- ⁴³
- ʔi*
- ⁴
- tʃi*
- ¹
- gi*
- ¹
- ja*
- ¹
- de*
- ⁴
- ʔne*
- ¹
- ,
- na*
- ⁴
- ʔa*
- ³
- ta*
- ³
- ni*
- ³
- ,

wi⁴³ʔi⁴ =tʃi¹gi¹ ja¹ de⁴ʔne¹ na⁴ʔ= a³ =ta³ni³
 one =DISTRIB LNK(III) cane(III) 3SBJ= have =price
 Each cane (stalk) is worth,

- (16) Teodor Sr.:
- dos, dos cincuenta=tʃi*
- ¹
- gi*
- ¹
- wi*
- ⁴³
- ʔi*
- ⁴
- ja*
- ¹
- de*
- ⁴
- ʔne*
- ¹
- .

dos cincuenta =tʃi¹gi¹ wi⁴³ʔi⁴ ja¹ de⁴ʔne¹
 Sp:two Sp:fifty =DISTRIB one LNK(III) cane(III)
 2, 2.50, for each cane.

I also recorded some naturally occurring events other than maximally informal conversations, such as congregations conducting Christian church services (5 bundles with 22 files) and adults coaching a teenagers' soccer game (1 bundle with 3 files). All of the recordings of naturally occurring events include both audio and video, except that church services were generally recorded with audio only.

Recordings of naturally occurring events were made in order to document general properties of participants' everyday communicative behavior. By contrast, the 3 bundles with recordings of **staged conversation** were collected for a specific purpose – to obtain data about spatial language and gesture, for comparison with the monolingual interviews. The recordings in these bundles all involve Angel Bitancourt Serra and one other participant. In the staged conversations, Bitancourt Serra asks the other participant (at my request) to describe a construction site, a locality, or a route. His participation is limited to asking questions and backchanneling.

(4) is an excerpt from one of the staged conversations. Magdalena Moreno Guerrero, left, is in the home of two relatives, who are currently traveling to work as teachers. She is living in the home full-time to care for the relatives' children. Bitancourt Serra, right, is Magdalena's classificatory nephew. Magdalena is telling him about her relatives' plans to renovate the house.

(4) An excerpt from a recording of staged conversation.

(Media reference: Bundle 2018-19.029,¹⁵

tca_20170707_disc_video_004_archive.mp4, 0:15-00:47)

(Transcript reference: Bundle 2015-06.065,¹⁶

tca_20170707_disc_video_004.eaf)

<example4_video.mp4>

(1) MMG: *nu⁵a² ta⁴ ni⁴¹ʔi⁴ ta¹na³tʃa³ʔu²ʔi⁴, nu⁵a², ti³¹ma²a¹ri³ bañoka¹,*
nu⁵a² ta⁴ ni⁴¹ʔi⁴ ta¹= na³= tʃau³ʔ =ʔi⁴ nu⁵a² ti³¹ma² =a¹ri³
 here FUT FOC 3(I)SBJ.SC= 3OBJ= smear =SUB here 3(I) =AL.POSS
baño =ka¹
 Sp:bath =PURP
 He (= owner of the house) is going to pour it (a concrete floor) HERE, for his bathroom,

(2) MMG: *nu¹ʔi¹ka⁵, ti³¹ma²a¹ri³ jau²ʔi⁵ru¹tq¹ʔmi¹ʔi³.*
nu¹ʔi¹ka⁵ ti³¹ma² =a¹ri³ jau² =ʔi⁵ru² =tq¹ʔmi¹ʔi³
 also 3(I) =AL.POSS wash =NI:clothes =rack
 And also, (for) his clothes-washing rack,

(3) MMG: *ti³¹ma²a¹ri³ du³tʃa¹ ki³ ta¹ʔi²ʔi⁴ka¹.*
ti³¹ma² =a¹ri³ du³tʃa¹ ki³ ta¹= i² =ʔi⁴ =ka¹
 3(I) =AL.POSS Sp:shower INTJ 3(I)SBJ.SC= make =SUB =PURP
 And so that he can build himself a shower.

(4) MMG: *pa⁴a² i¹, pa⁴a² pedazo i⁴ ta¹na³gi²ʔi⁴ i⁴ nu¹ʔma⁵,*
pa⁴a² i¹- pa⁴a² pedazo i⁴ ta¹= na³= gi¹ =ʔi⁴
 this(IV) HESIT this(IV) Sp:piece LNK(IV) 3(I)SBJ.SC= 3OBJ= put.grains =SUB
i⁴ nu¹ʔma⁵
 LNK(IV) now
 This, this part is what they're filling in,

¹⁵<http://cla.berkeley.edu/item/25733>.

¹⁶<http://cla.berkeley.edu/item/25435>.

- (5) MMG: $\eta u^1 \tilde{a}^4 t \tilde{f} i^4$, $na^4{}^3 gu^2 ta^4 ta^1 po^4 gi^1 \tilde{e}^4 \tilde{e}^2 \tilde{f} i^4 ka^1$, $da^2 a^2$,
 $\eta u^1 \tilde{a}^4 t \tilde{f} i^4 na^4{}^3 = gu^2 ta^4 ta^1 = po^4 gi^1 - \tilde{f} e^4 \tilde{e}^3 = \tilde{f} i^4 = ka^1$
 also 3 =LOC FUT 3(I)SBJ.SC= disassemble -CAUS =SUB =PURP
 $da^2 a^2$
 this(II)
 So that later he can have them take it down, (I mean) this,
- (6) MMG: $\eta e^3 a^2 ko^5 tu^4$, $na^4 nu^5 a^2 ma^4 ta^4 ta^1 na^3 na^1 \tilde{f} i^4 \tilde{a}^3$.
 $\eta e^3 a^2 ko^5 tu^4 na^4 nu^5 a^2 = \tilde{a}^4 ma^4 ta^4 ta^1 = na^3 =$
 this.MED(IV) thatch(IV) comp here = $\tilde{A}^4 MA^4$ FUT 3(I)SBJ.SC= 3OBJ=
 $na^1 = \tilde{f} i^4 = \tilde{a}^3$
 take:PLOBJ =SUB =REP
 This thatch, he says he's going to bring it down here.
- (7) MMG: $ta^4 na^3 wo^1 \tilde{f} i^3 \tilde{f} i^4 ta^4 \eta e^3 a^2 t \tilde{f} i^3 \tilde{f} i^4 e^2 \tilde{f}$.
 $ta^4 = na^3 = wo^1 - \tilde{f} i^1 = \tilde{f} ta^4 \eta e^3 a^2$
 1EXCL.SBJ.SC= 3OBJ= discard:PLOBJ -out:PLOBJ =SUB FUT this.MED(IV)
 $t \tilde{f} i^3 \tilde{f} i^4 e^2 = \tilde{f}$
 bad =NMLZ(IV)
 We're going to throw away this old stuff (= thatch).
- (8) MMG: ((*ingressive*)) $\eta e^5 ma^2 ta^1 na^3$ -,
 ((*ingressive*)) $\eta e^5 ma^2 ta^1 = na^3 =$
 yes there.ANA 3(I)SBJ.SC= 3OBJ=
 Yeah, and there he um-
- (9) MMG: $\eta e^3 ma^2 ri^1 wa^{31} \tilde{f} i^5 je^5 a^2 i^5 ta^1 pu^1 \tilde{f} ma^3 \tilde{f} i^3 \tilde{f} i^4 gu^2$, $ta^4 ni^4 \tilde{f} i^4$.
 $\eta e^3 ma^2 ri^1 wa^{31} \tilde{f} i^5 je^5 a^2 i^5 = ta^1 = pu^1 = \tilde{f} ma^3 \tilde{f} i^1 = \tilde{f} i^4$
 that.ANA TOP INTJ there.DIST ASP= 3(I)SBJ.SC= insert =NI:hole =NMLZ(IV)
 $= gu^2 ta^4 ni^4 \tilde{f} i^4$
 =LOC FUT FOC
 It's going to be there, where he's stood those (wall supports) up in the holes,
- (10) MMG: $ta^2 \tilde{f} u^2 \tilde{f} i^4 na^{31} \tilde{f} ka^1$, $i^4 \tilde{a}^4 ki^5 \tilde{f} i^4$,
 $tau^2 \tilde{f} = \tilde{f} i^4 na^{31} = \tilde{f} ka^1 i^4 \tilde{a}^4 ki^5 \tilde{f} i^4$
 NEG.EXIST =NMLZ(IV) 3 =PURP LNK(IV) HESIT(IV)
 What's missing now, um,

- (11) MMG: *ladrillo=ma⁴ã² ta⁴ ni⁴¹ʔĩ⁴ ta⁴na³ʔi²ʔta¹pĩ¹ʔi³ʔĩ⁴ ta⁴.*
ladrillo =ma⁴ã² ta⁴ ni⁴¹ʔĩ⁴ ta⁴= na³= i² =ʔta¹pĩ¹ʔi¹ =ʔĩ⁴
 Sp:brick =COM FUT FOC 1EXCL.SBJ.SC= 3OBJ= make =NI:wall =SUB
ta⁴
 FUT
 Is bricks to build the wall.
- (12) MMG: *ʔu¹ʔma⁵ ta⁴ ta¹je⁵a²ta²ʔã⁴ na¹da³ʔ.*
ʔu¹ʔma⁵ ta⁴ ta¹ je⁵a² =ta²ʔã⁴ na¹= da³¹ =ʔ
 now FUT also there.DIST =only 3SBJ.SC= lie.1D =SUB
 So that it reaches to over there.
- (13) ABS: [Sp] *ah ya, nu⁵a² cocina=wa⁵?*
ah.ya nu⁵a² cocina =wa⁵
 Sp:INTJ here Sp:kitchen =ALL
 Oh okay, [Tca] to here in the kitchen?
- (14) MMG: *ĩ²ʔĩ⁴*
ĩ²ʔĩ⁴
 logical.yes
 Uh-huh.

The two non-recording bundles documenting adult conversation consist of ELAN transcriptions. The transcriptions have the same structure as the transcriptions of staged events.

7.3 Adult language recording methods Recordings of naturally occurring events and staged conversations all include both audio and video, with the exceptions of some church services recorded with audio only.

Video recordings were made with the same equipment used for staged events – a Sony PJR540 video camera in 2018, and a Canon XA30 camera in 2018. Simultaneous audio tracks were recorded with a variety of equipment, including Zoom H1N and Zoom H4N recorders recording via the external microphone; Olympus VP10 audio recorders worn on the bodies of the participants; and a Rode NT4 stereo microphone attached to the Canon XA30 camera.

Informal conversations were recorded using unattended video cameras, according to the following procedure. I made an appointment with a household, arrived at their house at the agreed time, set up the video recording equipment in the kitchen (or another place where people were engaged in an extended activity), and then left. I stayed away for 30 to 60 minutes, returning for a minute or two every 15 minutes to check that the equipment was working correctly. Therefore, I am not present in

most of the video recordings. Speakers sometimes comment on the presence of the video camera, but not often – in transcriptions of the conversational data, 1–3 % of all turns concern the camera or the recording activity.

As well as recording conversations with unattended cameras, I also made some handheld video recordings of conversations taking place around me while I was acting as a participant observer. Though I am present in these recordings, they are distinct from monolingual interviews in that speakers do not usually talk to me.

7.4 Child language recordings and transcripts 96 of the 125 bundles in Collection 2018-19 document primarily child language and child-caregiver interaction. This set is comprised of 56 recording bundles and 40 non-recording bundles. The recording bundles include 463 recording files with total time 759 hours 2 minutes¹⁷ (audio time: 639 hours 22 minutes, video time: 119 hours 40 minutes). The non-recording bundles include 74 transcription files, with 16 hours 7 minutes total transcribed time (all of video).

The 96 child-language-focused bundles can be divided into two groups: materials from a larger study conducted in 2019 (86 bundles) and materials from a smaller pilot study conducted in 2018 (10 bundles). Since the 2019 materials are more extensive, I describe them first.

7.4.1 2019 Child language study The 86 child-language bundles from 2019 research are from a cross-sectional study with 46 child participants, aged 12 months to 4 years 11 months. Per §4, these materials have a one participant/one bundle organization and separate recording and transcription bundles.

Each child in the 2019 study participated in three recordings: a daylong recording, a recording of object play, and a recording of free play. All three recording sessions were conducted in the children's homes. Each participant's recordings are archived in a single bundle; thus, there are 46 recording bundles.

The **daylong recordings** include audio only. They were made with audio recorders worn on children's bodies for a full 9-hour day, 8am to 5pm. They document the child participants' own language use, as well as the language that participants heard from adults and other children, over the course of the day.

Recordings of object play include audio and video. They feature one caregiver and either one or two children. Six sibling pairs participated in the task together; the other 34 participants were recorded alone. At the beginning of the recording, I provided the caregiver and child with a locally acquired set of 50 glass marbles. I asked them to play with the marbles for 30 minutes. I was present throughout the recordings in order to operate the video cameras and discourage non-target household members from participating.

(5) is an excerpt from a video recording of object play. The child participant in this recording is a 12-month-old girl. The caregiver's body-worn audio recorder can be seen clipped to her clothing near her right shoulder; the child's is inside the blue pocket

¹⁷Equivalent to 31 days, 15 hours, 2 minutes.

on her t-shirt. Audio on the video comes from the caregiver's body-worn recorder (not the video camera, which suffers from echo in this poured concrete room).

(5) Excerpt from a video recording of an object play session, with a 12-month-old girl and her caregiver as participants.

(Media reference: Bundle 2018-19.083,¹⁸
tca_201909_child_child44_tot_video_pjr.mp4, 3:55–4:24)

(Transcript reference: Bundle 2018-19.123,¹⁹
tca_201909_child_child44_tot_video_ahs.eaf)

<example5_video.mp4>

(1) Caregiver: *ku³¹ʔi⁵ra¹*
ku³¹ = ʔi⁵ra¹
2SG = first
You first,

(2) Caregiver: *ma³je⁵a², gu³¹ʔe²ã⁴ma⁴ di¹ʔ, gu³¹ʔe²ã⁴ma⁴, ku³¹we⁵ã⁴ma⁴*
ma³ri¹ je⁵a² gu³¹ʔe² =ã⁴ma⁴ di¹ʔ gu³¹ʔe² =ã⁴ma⁴ ku³¹
PERF there.DIST that.DIST(I) =ã⁴MA⁴ INTJ:look! that.DIST(I) =ã⁴MA⁴ 2SG
=we⁵ =ã⁴ma⁴
=behind =ã⁴MA⁴
Now, there, there's the other one, look – there's the other one behind you,

(3) Caregiver: *gu³¹ʔe²ã⁴ma⁴, ku³¹we⁵ã⁴ma⁴*
gu³¹ʔe² =ã⁴ma⁴ ku³¹ =we⁵ =ã⁴ma⁴
that.DIST(I) =ã⁴MA⁴ 2SG =behind =ã⁴MA⁴
There's the other one, behind you,

(4) Caregiver: *gu³¹ʔe²ã⁴ma⁴ ʔa⁵tʃa³ʔ,*
gu³¹ʔe² =ã⁴ma⁴ ʔa⁵tʃa³ʔ
that.DIST(I) =ã⁴MA⁴ 1SG.QUOT
There it is, I said,

¹⁸<http://dx.doi.org/doi:10.7297/X2V4oSF1>.

¹⁹No URL; not yet accessioned.

- (5) Caregiver: $gu^{31}\gamma e^2 \tilde{a}^4 ma^4 \gamma^5 t\tilde{f}i^2 ma^3, ku^{31} we^5 \tilde{a}^4 ma^4$
 $gu^{31}\gamma e^2 = \tilde{a}^4 ma^4 = \gamma^5 t\tilde{f}i^2 = ma^3 \quad ku^{31} = we^5 = \tilde{a}^4 ma^4$
 that.DIST(I) = $\tilde{A}^4 MA^4$ =really =INFO 2SG =behind = $\tilde{A}^4 MA^4$
 There it is over there, behind you.
- (6) Caregiver: $t\tilde{i}^{31} \gamma^3 na^1 ja^1 \gamma u^2$
 $t\tilde{i}^{31} = \gamma^3 \quad na^1 = ja^1 \gamma u^2$
 3(I) =ACC IMP= grab
 Grab it!
- (7) Caregiver: $d\tilde{i}^1 \gamma$
 $d\tilde{i}^1 \gamma$
 INTJ:look!
 Look!
- (8) Caregiver: $ta^4 ma^3 ni^{41} \gamma^4 ku^1 na^3 ga^3 na^1 g\tilde{i}^5 \gamma, m^1 m^3$
 $ta^4 ma^3 \quad ni^{41} \gamma^4 \quad ku^1 = \quad na^3 = \quad ga^3 na^1 = g\tilde{i}^5 \gamma = ? \quad m^1 m^3$
 NEG FOC 2SG.SBJ.SC= 3OBJ= win =PL =SUB INTJ
 You're NOT winning them, right?
- (9) Caregiver: $ma^3 r\tilde{i}^3$
 $ma^3 r\tilde{i}^3$
 PERF
 Now!

The object play recordings were collected to document how children and caregivers manage one another's attention, and coordinate actions with one another, through language and visual behavior (e.g. by producing demonstrative words, imperatives, and pointing gestures; and by following others' pointing gestures and gaze). The sessions were organized around the task of playing marbles because it is a common activity for children the participants' age in Cushillococha. Additionally, games of marbles successfully elicited demonstratives, pointing gestures, and other attention-managing behaviors from children in the 2018 pilot study.

Recordings of free play are also audio-video. They feature one or two caregivers and one or two children. 11 pairs of co-resident children – 6 sibling pairs and 5 pairs of children who lived in the same household but were not siblings – participated in the task together; each pair was recorded with one or two caregivers. The other 23 participants were recorded alone.

At the beginning of the free play sessions, I instructed the child and caregiver participants that they could do anything they liked during the recording time, provided that they stayed in the same room. Then, after turning on the recording devices, I left

the scene for ~60 minutes. As in unattended camera recordings of adults, I returned once per 15–20 minutes to check the recording equipment.

(6) is an excerpt from a recording of object play. The child participant in this recording is a boy aged 3 years, 3 months. He and his caregiver, the adult woman, are playing with a commercially manufactured toy. The second video camera is visible at left.

(6) Excerpt from a video of a free play session, with a boy aged 3 years, 3 months and his caregiver as participants.

(Media reference: 2019-19.081,²⁰

tca_201909_child_child43_cci_xa30.mp4, 34:48–35:15)

(Transcript reference: 2019-19.121,²¹

tca_201909_child_child43_cci_ahs.eaf)

<example6_video.mp4>

- (1) Caregiver: *ki²a⁴na⁴ ta⁴na³tʃi⁴na⁴gi⁴, di¹ʔ, ooo³*
ki²a⁴na⁴ ta⁴= na³= tʃi⁴ -na⁴gi⁴ di¹ʔ o³
 INTJ:tag 1INCL.SBJ= 3OBJ= stand -upward INTJ:look! INTJ:oh
 Right, let's make him stand on top, look, ooh!
- (2) Child: *ta⁴ã³*
 [nonword: depicting crash]
- (3) Caregiver: *nu⁵a²ta²ã⁴ ni⁴ʔi⁴ na⁴a² no⁵i¹ri³ e³ne³mi⁵go¹gi⁴ i²ʔki⁴*
nu⁵a² =ta²ã⁴ ni⁴ʔi⁴ na⁴a² no⁵i¹ri³ e³ne³mi⁵go¹ =gi⁴ i²ʔki⁴
 here =only FOC this(IV) 3.AL.POSS Sp:enemy =PL INFO
 They're right here, these, his enemies, right?
- (4) Child: *du³ du³i¹ du³ʔ, du³ du⁴? du¹*
 [nonwords: voicing toys]
- (5) Caregiver: *m³ʔ, a⁴ na⁴a² ri³ã³, na⁴nu⁴e¹gu¹pa³ra¹,*
m³ʔ a⁴ na⁴a² ri³ã³ na⁴= nu⁴ -e¹gu¹ =pa³ra¹
 INTJ:yes LNK(IV) this(IV) TOP 3SBJ= put:INAMPLOBJ -in.circle =NI:foot/leg
 Mm, and this one, it had its feet put back on,

²⁰<http://dx.doi.org/doi:10.7297/X23J3B6Z>.

²¹No URL; not yet accessioned.

- (6) Caregiver: $nu^5 a^2 ma^4 t\tilde{f}i^4 r\tilde{e}^1 me^{31} \tilde{a}^2 nu^4 pa^3 ra^1$.
 $nu^5 a^2 = \tilde{a}^4 ma^4 = t\tilde{f}i^4 r\tilde{e}^1 me^{31} \tilde{a}^2 nu^4 = pa^3 ra^1$
 here = $\tilde{A}^4 MA^4$ = actually well put:INAMPLOBJ =NI:foot/leg
 So now its legs have got put on properly here.
- (7) Child: $d\tilde{i}^3 \tilde{r}^1 d\tilde{i}^3 \tilde{r}^1 d\tilde{i}^3 \tilde{r}^1, d\tilde{i}^3 \tilde{r}^1 ta^4$
 [nonwords: voicing toys]
- (8) Child: $d\tilde{i}^1 \eta ka^4 a^3 a^1, ta^4 \tilde{a}^3 ni^4 \eta \tilde{r}^4$
 $d\tilde{i}^1 \eta ka^4 a^3 a^1 ta^4 ma^3 ni^4 \eta \tilde{r}^4$
 INTJ:look INTJ:ah NEG FOC
 Look, ah, no!
- (9) Caregiver: $t\tilde{f}i^3 h t\tilde{f}i^3 h t\tilde{f}i^3 h t\tilde{f}i^3 h, t\tilde{f}i^3 h t\tilde{f}i^3 h t\tilde{f}i^3 h t\tilde{f}i^3 \eta, t\tilde{f}i^3 h t\tilde{f}i^3 h$
 [nonwords: voicing toys]
- (10) Caregiver: mba^{31}
 mba^{31}
 INTJ:uh-oh
 Uh-oh
- (11) CHI: ((laughter)) $a^4 \eta a^4 a^2 wa^{31} \eta i^5$
 $a^4 \eta a^4 a^2 wa^{31} \eta i^5$
 LNK(IV) this(IV) INTJ
 Ah this one though!

Recordings of free play were collected to document children's language use, as well as child-caregiver interaction, in an everyday setting. This is much the same purpose as the daylong audio recordings. However, the free play recordings are much richer per time than the daylong audio recordings, since they include dual-camera video. The free play sessions also featured a large quantity of adult conversation, as adult household members often wandered into the recording scene and talked to the caregivers.

In addition to these three types of recordings, the child data from 2019 also includes 40 non-recording bundles: 39 transcript bundles and one bundle with a text explanation of the file structure used in the transcripts.

Each of the 39 transcript bundles corresponds either to one participant or to a sibling pair of participants (siblings were always recorded together, except in the

daylong audio recordings). Each transcript bundle contains two ELAN files: one with annotations of 10 continuous minutes of the participant's object play recording, and one with annotations of 10 continuous minutes from the free play recording. In bundles that correspond to a sibling pair of participants, each ELAN file has 20 rather than 10 minutes transcribed. Thus, each of the 45 complete participants in the study has at least 20 minutes of transcription in the archive. Participants in sibling pairs have 40 minutes each.²²

The ELAN annotations consist of phonetic transcriptions of all hearable speech – by child participants, caregivers, and others – in the sampled time, accompanied by translations. Additionally, ELAN annotations on recordings of undirected child-caregiver interaction include a tier which codes the addressee type of each turn, distinguishing child-directed vs. adult-directed speech. The final non-recording file bundle for the 2019 child materials, 2018-19.124, explains the tier structure, transcription conventions, and addressee coding abbreviations used in the ELAN annotations of child data.

7.4.2 2018 Pilot study As well as the child language materials from 2019 just described, Collection 2018-19 also contains 10 recording bundles with materials from a 2018 pilot study. They include 5 bundles of long-form audio recordings of children (2–3 hours each; created to pilot the daylong audio recording protocol); 3 bundles of audio-video recordings of object play between children (created to pilot the object play protocol); and 2 bundles of audio recordings of free play. Bundles of long-form audio recordings from the pilot study are organized on a one age group/one bundle structure. Other recordings from the pilot study are organized on a one recording event/one bundle structure.

Some data from the pilot study has also been transcribed. It is archived together with transcriptions of adult conversation made in 2018, in bundle 2018-19.020. The sole non-recording bundle specific to the 2018 child pilot study is a bundle with scans of a field notebook providing data about the participants, such as their age.

7.4.3 Collection volume Now that all data types represented in the child bundles have been introduced, Table 4 reports the count, total time, and total transcribed time of the child recordings in Collection 2018-19.

7.5 Child language recording methods Daylong and long-form audio recordings of children were made with Olympus VP10 audio recorders sampling at 22.05 kHz and writing to WAV format. Following a method developed by Casillas et al. (2020), children participating in these recordings wore t-shirts fitted with a custom pocket. The pocket was precisely sized to accommodate the Olympus VP10 and was sealed with Velcro. The recorder was inserted into the pocket and worn on the child's body for a full day.

²²There are 45 rather than 46 complete participants because one child did not complete all three recording sessions. Her data was archived but not transcribed.

Table 4. Volume of recordings and transcriptions of child language in Collection 2018-19, by data and media type.

Data Type	Media	Recordings			Transcriptions	
		Count	Time	Unique Time	Count	Time
Daylong/ Long-Form	Audio	116	455 h 48 m	455 h 48 m	0	n/a
	Video	0	n/a	n/a	n/a	n/a
Object Play	Audio	91	57 h 3 m	0 h 0 m	0	n/a
	Video	84	48 h 8 m	25 h 1 m	39	7 h 48 m
Free Play	Audio	99	126 h 31 m	0 h 0 m	0	n/a
	Video	73	71 h 32 m	36 h 47 m	35	8 h 18 m

Video recordings of child-caregiver interaction were made with a dual-camera setup. Participants were simultaneously recorded on two cameras (Sony PJR-540 and Canon XA30) placed at opposing angles – one facing the child participant(s), the other facing the caregiver. Audio data from child-caregiver interaction included four separate audio tracks per event. The first audio track was recorded via an Olympus VP10 recorder worn on the body of the child participant (using the same equipment as the daylong audio recording). The second audio track was recorded with a second Olympus VP10 clipped to the caregiver's clothing. The third and fourth audio tracks were recorded via a Rode NT4 microphone attached to the XA30 video camera and via the internal microphone of the Sony PJR-540 camera.

I chose to use dual-camera video for the child-caregiver interaction recordings because participants were highly mobile. Therefore, it would have been impossible to keep all participants consistently in frame with a single camera. I used the multiple redundant audio channels in order to ensure that the data would be suitable for transcription, despite the significant environmental noise and noise generated by the participants' physical activity.

8. Collection 2018-20: Psycholinguistic experiments with adults Recordings and other data from psycholinguistic experiments make up all of Collection 2018-20. This collection consists of video and audio recordings of 24 adults participating in an experiment on the production of demonstratives (Skilton & Peeters under review). Each participant's recordings appear in a separate bundle. Therefore, the collection has 24 recording bundles, with total recording time 40 hours 17 minutes (total audio time: 24 hours 40 minutes; total video time: 15 hours 37 minutes). Recordings were made with the same methods used for staged events. As well as the 24 participant bundles in Collection 2018-20, there are 3 bundles of written data documenting participants' responses in the experiment, as well as results from a pretest of the experimental materials. There are no transcriptions.

9. Future use of the collections I am currently using the collections as a source of data for studies of first language acquisition; the acquisition of deictic gesture; adults' use of demonstratives and deictic gestures; and adult word prosody. I also plan to write a reference grammar of Ticuna based on the collections.

I actively encourage other linguists to use the collections as a data source. One purpose for which they will definitely prove useful is the study of Ticuna dialectology. The Cushillococha materials can be compared with data from other varieties of Ticuna, whether collected in new fieldwork or collated from published sources. At least one researcher, Denis Bertet (DDL Lyon), has already used the collections for this purpose.

As well as inter-dialectal comparison, I have structured the materials to also facilitate cross-linguistic comparison (§1). In particular, the 2019 child language data set was designed to be as comparable as possible to language acquisition data sets for other languages. For example, the object play data included in this dataset is extremely similar to object play data collected by researchers studying the acquisition of colonial languages like English and French. Therefore, the object play dataset could easily be used to compare properties of child language and/or child-directed speech in Ticuna vs. in other languages. Likewise, the data on conversation between adults was collected with the same method – unattended camera recordings – that Conversation Analysts use to collect data on global languages. Thus, as with the child data, it could be used for projects in pragmatic typology (Rossi et al. 2020) with minimal adaptation.

Even though the collections are open-access and highly accessible in the developed world, they are – unfortunately – still inaccessible to Ticuna speakers. The collections are accessible in the developed world and inaccessible to speakers for exactly the same reason: they are digital. No Ticuna community has internet access that reliably allows users to view audio and video, meaning that any digital archive would be inaccessible to most speakers. Of course, the collections can be **converted** into formats which are accessible to speakers – for example, I have returned copies of many of the text transcriptions to their authors as printed documents. But the collections themselves, because they are digital, will necessarily remain inaccessible until high-speed internet access becomes widely available in the Amazon Basin.

10. Conclusion I have described the collections in detail here for two reasons (§1). First, I want to stimulate other researchers to use the archive for comparative work. In order to facilitate comparison, I have been as explicit as possible about how others might use the materials (§9) and how the materials were collected (§3, §§5–7).

Second, I have also aimed to encourage other documentation workers to consider collecting less-common forms of data – like conversation and child language – by describing the methods that have allowed me to collect these data types. Others have already made the case that conversational data is useful for linguistic analysis (Bowerman 2008: 119–121; Dingemanse & Floyd 2014: 458–466) and valuable for language reclamation (Baldwin et al. 2018: 193; Grenoble 2013: 53–54; Hermes & King 2019). Despite these motivations, documentation researchers still often do not

collect conversational data, believing that it is too fragmented or difficult to work with. To counter this perception, I have provided a transparent example workflow for naturally occurring data, demonstrating how it can be collected, processed, and displayed using common documentation tools (§7).

References

Archival Collections

- [Collection 2015-06] Skilton, Amalia. Ticuna Elicitation and Texts, 2015-06, Survey of California and Other Indian Languages, University of California, Berkeley, <http://dx.doi.org/doi:10.7297/X29P2ZPJ>.
- [Collection 2018-19] Skilton, Amalia and Angel Bitancourt Serra. Ticuna conversations, 2018-19, Survey of California and Other Indian Languages, University of California, Berkeley, <http://dx.doi.org/doi:10.7297/X2F769QD>.
- [Collection 2018-20] Skilton, Amalia. Ticuna experiments, 2018-20, Survey of California and Other Indian Languages, University of California, Berkeley, <http://dx.doi.org/doi:10.7297/X29G5K03>.
- [Collection 2019-12] Félix, Aldenor Basques, Rosa Diques Ponciano, Domingos Ricardo Florentino, Felipe Florentino, Marta Nicanor, Bernardino Pereira, Moisés Pereira, Márcia Pereira, Eucilene Ponciano, Tobias da Silva Pereira, and Wilson de Lima Silva. Ticuna Field Materials, 2019-12, Survey of California and Other Indian Languages, University of California, Berkeley, <http://dx.doi.org/doi:10.7297/X2R20ZMQ>.

Other References

- Anderson, Doris. 1962. *Conversational Ticuna*. Yarinacocha, Peru: SIL.
- Anderson, Doris & Lambert Anderson (eds.). 2017. *Diccionario Ticuna-Castellano*. Lima: Instituto Lingüístico de Verano.
- Anderson, Lambert. 1959. Tikuna vowels with special regard to the system of five tonemes. In *Serie lingüística especial* 1. Rio de Janeiro: Museu Nacional.
- Anderson, Lambert. 1966. The structure and distribution of Ticuna independent clauses. *Linguistics* 4(20). 5–30.
- Baldwin, Daryl, Leanne Hinton, & Gabriela Pérez-Báez. 2018. The Breath of Life workshops and institutes. In Hinton, Leanne, Leena Marjatta Huss, & Gerald Roche (eds.), *The Routledge handbook of language revitalization*, 188–96. New York: Routledge/Taylor & Francis Group.
- Begay, Kayla. 2017. *Wailaki grammar*. University of California Berkeley. (Ph.D. dissertation.)
- Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide*. New York: Palgrave Macmillan.

- Caballero, Gabriela. 2017. Choguita Rarámuri (Tarahumara) language description and documentation: A guide to the deposited collection and associated materials. *Language Documentation & Conservation* 11. 224–55. (<https://scholarspace.manoa.hawaii.edu/handle/10125/24734>)
- Carvalho, Fernando de. 2009. On the genetic kinship of the languages Tikúna and Yuri. *Revista Brasileira de Lingüística Antropológica* 1(2). 247–268.
- Carvalho, Fernando de. 2010. *Estruturas fonéticas da língua Tikúna: Um estudo acústico preliminar*. Universidade de Brasília. (Master's thesis.)
- Carvalho, Fernando de. 2012. Variação de VOT, duração e equações de locus em função do contexto prosódico nas oclusivas do Tikúna. *Revista Diadorim* 12. 129–147.
- Casillas, Marisa, Penelope Brown, & Stephen C. Levinson. 2020. Early language experience in a Tzeltal Mayan village. *Child Development* 91(5). 1819–35.
- Damaso Yoni, Jesús, Abraham Ipuchima Rivera, & Abel Antonio Santos Angarita. 2009. *Conocimiento local indígena sobre los peces de la Amazonia, lagos de Yahuar-caca*, ed. by Santiago R. Duque. Leticia, Colombia: Universidad Nacional de Colombia, Sede Amazonia.
- Dingemanse, Mark. 2011. *The meaning and use of ideophones in Siwu*. Radboud University Nijmegen. (Ph.D. dissertation.)
- Dingemanse, Mark & Simeon Floyd. 2014. Conversation across cultures. In Enfield, N. J., Paul Kockelman, & Jack Sidnell (eds.), *The Cambridge handbook of linguistic anthropology*, 447–80. Cambridge: Cambridge University Press.
- Enfield, N. J. 2014. Doing fieldwork on the body, language and communication. In Müller, C., A. J. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Tessendorf (eds.), *Body – language – communication: An international handbook on multimodality in human interaction*, 974–81. Berlin: De Gruyter.
- Enfield, N. J., Tanya Stivers, & Stephen C. Levinson. 2010. Question-response sequences in conversation across ten languages: An introduction. *Journal of Pragmatics* 42. 2615–2619.
- Goulard, Jean-Pierre, María Emilia Montes Rodríguez, Loida Ángel Ruiz, Abel Antonio Santos Angarita, & Valentina Nieto Fernández (eds.). 2016. *Relato de Chetanükü del Loretoyacu: Origen del mundo y de los Tikuna: Patrimonio oral inmaterial del pueblo Tikuna de la Amazonia*. Bogotá: Universidad Nacional de Colombia, Facultad de Ciencias Humanas.
- Grenoble, Lenore A. 2013. Unanswered questions in language documentation and revitalization: New directions for research and action. In Mihás, Elena, Bernard Perley, Gabriel Rei-Doval, & Kathleen Wheatley (eds.), *Responses to language endangerment: In honor of Mickey Noonan*, 43–58. Amsterdam: John Benjamins.
- Hermes, Mary & Kendall A. King. 2019. Task-based language learning for Ojibwe: A case study of two intermediate adult language learners. In McCarty, Teresa L., Sheilah E. Nicholas, & Gillian Wigglesworth (eds.), *A world of Indigenous languages: Politics, pedagogies and prospects for language reclamation*, 134–52. Bristol: Multilingual Matters.
- Instituto Socio-Ambiental. 2017. *Introdução: Ticuna*.

- Lewis, M. Paul, Gary F. Simons, & Charles D. Fennig (eds.). 2014. *Ethnologue: Languages of the world*. 17th edn. Dallas: SIL International.
- Meakins, Felicity, Jennifer Green, & Myfany Turpin. 2018. *Understanding linguistic fieldwork*. London: Routledge.
- Montes Rodríguez, M. E. 1995. *Tonología de la lengua Ticuna (Amacayacu)*. Bogotá: Universidad de los Andes.
- Montes Rodríguez, M. E. 2004. *Morfosintaxis de la lengua Ticuna*. Bogotá: Universidad de los Andes.
- Robinson, Laura C. 2010. Informed consent among analog people in a digital world. *Language & Communication* 30(3). 186–191.
- Rosenblum, Daisy & Andrea L. Berez-Kroeker. 2018. Reflections on language documentation in North America. In McDonnell, Bradley, Andrea L. Berez-Kroeker, & Gary Holton (eds.), *Reflections on language documentation 20 years after Himmelmann 1998* (Language Documentation & Conservation SP15), 340–353. Honolulu: University of Hawai'i Press.
- Rossi, Giovanni, Simeon Floyd, & N. J. Enfield. 2020. Recruitments and pragmatic typology. In Floyd, Simeon, Giovanni Rossi, & N. J. Enfield (eds.), *Getting others to do things: A pragmatic typology of recruitments*, 1–24. Berlin: Language Sciences Press.
- Santos Angarita, Abel Antonio. 2005. *Hacia una dialectología Ticuna del Trapecio Amazónico Colombiano*. Universidad Nacional de Colombia, Sede Amazonía. (Master's thesis.)
- Santos Angarita, Abel Antonio. 2013. *Percepcion Tikuna de Naane y Naune: Territorio y cuerpo*. Universidad Nacional de Colombia, Sede Amazonía. (Master's thesis.)
- Seifart, Frank & Juan Alvaro Echeverri. 2014. Evidence for the identification of Carabayo, the language of an uncontacted people of the Colombian Amazon, as belonging to the Tikuna-Yurí linguistic family. *PLoS ONE* 9(4). e94814.
- Skilton, Amalia. 2017. *Phonology and nominal morphology of Cushillococha Ticuna*. University of California Berkeley. (Ph.D. dissertation prospectus.)
- Skilton, Amalia. 2019. *Spatial and non-spatial deixis in Cushillococha Ticuna*. University of California Berkeley. (Ph.D. dissertation.)
- Skilton, Amalia. under review. Demonstratives and visibility: Data from Ticuna and implications for theories of deixis.
- Skilton, Amalia & David Peeters. under review. Cross-linguistic differences in demonstrative systems: Comparing spatial and non-spatial influences on demonstrative use in Ticuna and Dutch.
- Soares, Marília Facó. 1992. Ordem de palavra: Primeiros passos para uma relação entre som, forma e estrutura em Tikuna. *Ameríndia* 17. 89–119.
- Soares, Marília Facó. 2000. *O suprasegmental em Tikuna e a teoria Fonológica*. Campinas: UNICAMP.
- Soares, Marília Facó. 2017. A análise de tempo em Ticuna (Tikuna) revisitada: Questões sobre anáfora temporal e sequenciamento temporal. *Revista Linguística* 13(2). 263–285.

- Sullivant, Ryan. 2020. Archival description for language documentation collections. *Language Documentation & Conservation* 14. 520–78. (<https://scholarpace.manoa.hawaii.edu/handle/10125/24949>)
- Woodbury, Anthony. 2011. Language documentation. In Austin, Peter & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 159–186. Cambridge: Cambridge University Press.

Amalia Skilton
amalia.skilton@austin.utexas.edu
 orcid.org/0000-0002-9848-5688