

## Understanding the effects of language revitalization workshops using long-format speech environment recordings

Julia Nee\*

**Abstract.** Long-format speech environment (LFSE) recordings are increasingly used to understand language acquisition among young children (Casillas & Cristia 2019). But in language revitalization, OLDER children are sometimes the largest demographic acquiring a language. In Teotitlán del Valle, Mexico, older children have participated in Zapotec language revitalization workshops since 2017. To better understand how these children use language, and to probe whether the language workshops impact language use, I invited learners to collect LFSE recordings. This study addresses two main questions: (1) what methodological challenges emerge when children ages 6-12 collect LFSE data?; and (2) what do the data suggest about the effects of the Zapotec workshops? I argue that, while creating LFSE recordings with older children presents methodological challenges, the results are useful in highlighting the importance of not only teaching language skills, but of creating spaces where learners are comfortable USING the Zapotec language.

**Keywords.** language revitalization; natural language; daylong recordings; Zapotec

**1. Introduction.** One of the key elements for sustainable language revitalization is the (re)establishment of intergenerational language transmission (Fishman 1991), but determining the extent to which intergenerational language use is occurring can be difficult (Dauenhauer & Dauenhauer 1998). While self-reports of language use might be relatively easy to collect, they are often inaccurate (Dauenhauer & Dauenhauer 1998, Bedor et al. 2012), and data from direct observations can be affected by an observer’s paradox (Labov 1972). One potential method for gathering data on naturalistic intergenerational language use is long-format speech environment (LFSE) recordings. In typical LFSE recordings, participants wear a lightweight audio and/or video recording device throughout their day, and thus have the potential to record a rich range of naturalistic data while reducing the impact of the observer’s paradox (Cychosz et al. 2020).

LFSE data is becoming increasingly common in the study of child language acquisition and adult language use in both widely-spoken and less-commonly-spoken languages (Casillas & Cristia 2019), and may help to shed light on language use in revitalization contexts as well. However, many previous LFSE studies focus on language acquisition during early childhood (Casillas & Cristia 2019),<sup>1</sup> but in many language endangerment contexts, intergenerational transmission of a language has been interrupted. As language revitalization projects gain

---

\* Para obtener una copia de este trabajo en español, favor de contactar a la autora por correo. Many thanks to all who made this research possible, especially Licenciada Rosita Jiménez Lorenzo for all of her support with the Zapotec workshops; undergraduate research assistants Raquel Camacho, Nicholas Carrick, Kathia Gonzalez, Isela Peralta, Celine Rezvani, Cassandra Serrano, and Maria Zarate for helping with the design of Zapotec language learning activities; Andrew Garrett, Leanne Hinton, Christine Beier, Beth Piatote, Meg Cychosz, and other colleagues at Berkeley and the LSA who provided feedback on this talk; and the Oswalt Endangered Language Grant which funded this research. Of course, I am extremely grateful to all of the Zapotec learners and families in Teotitlán who are part of this work, but who remain unnamed to protect their privacy. All errors are mine alone. Author: Julia Nee, University of California, Berkeley ([jnee@berkeley.edu](mailto:jnee@berkeley.edu)).

<sup>1</sup> This may be due in part to the fact that the popular LENA software that can be used to automate parts of LFSE data processing, thus making LFSE studies easier to carry out, is designed for use with American English-speaking children 2-36 months of age, though it has been used with other populations (Casillas & Cristia 2019).

traction, it is often the case that older, school-age children are beginning to learn the language as caregivers make a concerted effort to reverse language shift.

In this paper, I report on the methodology, results, and implications of gathering LFSE recordings from Zapotec language learners ages 6-12 in Teotitlán del Valle, Oaxaca, Mexico. The goals of the LFSE study were twofold: (1) to better understand the linguistic environment and language abilities of children in Teotitlán and (2) to assess the efficacy of a series of Zapotec language workshops designed to increase Zapotec use among children. I argue that while LFSE with kids ages 6-12 presented significant methodological challenges and produced data that differs from other LFSE studies, the data children produced are still valuable. Particularly when paired with other data, such as reported language use and researcher observations, LFSE data can help build a more complete picture of how language is being used by participants. It can also assist in assessing children's language proficiency by providing a semi-private, low-stress environment where children can utilize their language abilities. Moreover, LFSE recordings included some data on naturalistic parent-child interactions in Zapotec. This type of data represents an underdocumented genre of Zapotec language use which will be invaluable for future language revitalization efforts in which language learners may wish to bring Zapotec into their families through intergenerational language use.

While the data presented in this study are limited and potentially show sampling biases, I argue that they remain useful for widening our understanding of how some children in Teotitlán are using language. Most notably, children's Zapotec use in the LFSE recordings sometimes showed a wider range of language abilities than was observed in the language revitalization classroom or in more public settings. This finding lends support to the claim that increasing language use requires not only the teaching of language skills, but also the creation of spaces where learners feel comfortable using their language skills. More specifically, LFSE recordings showed that when children spoke to adults using ANY Zapotec (even one Zapotec word within a Spanish sentence), the adults responded in Zapotec. Thus, one key aspect in creating sustainable intergenerational language use may be empowering learners to start conversations with adults using whatever emerging Zapotec skills they have.

1.1. TEOTITLÁN DEL VALLE ZAPOTEC. Teotitlán del Valle Zapotec (TdVZ; zab) is a variety of Western Tlacolula Valley Zapotec within the Otomanguean language family. The 2020 Census reports that 63.51% of the population over the age of three in Teotitlán speaks an Indigenous language, out of a total population of 6,392 people (INEGI 2020). As of 2005, there were 198 reported monolingual Zapotec and 3,835 bilingual Spanish/Zapotec speakers in Teotitlán (Gobierno del estado de Oaxaca 2010).<sup>2</sup> According to UNESCO, Western Tlacolula Valley Zapotec (and thus TdVZ as a subset of Western Tlacolula Valley Zapotec) is “definitely endangered” (Moseley 2010).

While previous research and my own participant observations in Teotitlán suggest that children are increasingly using Spanish over Zapotec, resulting in a shift towards Spanish dominance, Teotitlán continues to be a vibrant center for Zapotec language use and promotion. Zapotec can be heard throughout town, spoken by adults in the marketplace, at community meetings, and at family celebrations. Furthermore, a number of language revitalization activities have been implemented within the community. There have been Zapotec language initia-

---

<sup>2</sup> Spanish and Zapotec are the dominant languages in Teotitlán, though my personal interactions in Teotitlán have shown that some residents also speak additional languages, including other Indigenous languages and English.

tives offered at the preschool, one of the two elementary schools, the middle school, and the high school. There is a cultural center and museum that contain displays written in Zapotec. A community language committee, *Diidxih baa xteh gulaas*, promotes Zapotec. Finally, I have been involved in hosting Zapotec language workshops for kids periodically during summer and winter since 2017. Thus, while Zapotec use continues to be threatened by colonial languages (like Spanish and English), there is also a strong push for Zapotec language revitalization.

1.2. ZAPOTEC LANGUAGE WORKSHOPS FOR KIDS. The participants in this LFSE study were also participants in the Zapotec language workshops for kids held in summer 2019. These workshops have been co-organized by me and (beginning in January 2019) Licenciada Rosita Jiménez Lorenzo, a native Zapotec speaker and resident of Teotitlán. The workshops were held in summer 2017, summer 2018, January 2019, summer 2019, and December 2019 for periods from one to three weeks.<sup>3</sup> Within each workshop, sessions were held for about two hours per day, Monday through Friday,<sup>4</sup> for a total of ~110 hours of instruction over the 5 iterations of the workshops. In each workshop, time was divided between two main activities: (1) classroom instruction during which Rosita and I worked to build students' vocabularies, conversational skills, and confidence in speaking and (2) field trips during which children visited with native Zapotec speakers throughout the community to put their Zapotec skills into practice.

1.3. RESEARCHER POSITIONALITY. I would like to present a brief overview of my background and positioning within this project, following best practices for researcher transparency in linguistics (Montoya 2020). I am a white settler colonial US American who was raised near Chicago, Illinois, and I am currently a PhD student in Linguistics at the University of California, Berkeley. Since 2012, I have been involved in language work in Teotitlán del Valle, first doing linguistic research and documentation of syntax and semantics, and more recently collaborating on language revitalization projects. My language revitalization research has been carried out through a Participatory Action Research (PAR) framework (Whyte et al. 1991) through which stakeholders in language revitalization projects participate throughout the research and implementation process.<sup>5</sup> During the research process, I have taken on various roles: researcher, teacher, language activist, student, colleague, and friend, among others. While it is possible to name these various roles, they are not distinct, but rather simultaneously influence one another. In particular, I wish to acknowledge that my role as TEACHER consistently influences my role as RESEARCHER. While I seek to gather data to enable the research aspect of this project to be successful, when put in a position where the research may get in the way of creating positive learning experiences for students, I chose to prioritize students' well-being. In the case of gathering LFSE data, more specifically, I was not simply a

<sup>3</sup> More workshops are planned for the future, but at the time of writing they are on hold due to COVID-19.

<sup>4</sup> With the exception of the first workshop, during which classes were held Monday, Wednesday, and Friday.

<sup>5</sup> PAR shares key elements with research models proposed by other scholars, including Community-Based Language Research (Czaykowska-Higgins 2009) and collaborative consultation (Leonard & Haynes 2010). While there are also some differences between specific collaborative and participatory models, I focus on their similarities as approaches that oppose the dehumanization of science (Chevalier & Buckles 2019), and which have similar benefits (such as empowering minoritized groups, developing collective support, and strengthening intergenerational relationships) and challenges (such as the difficulty of ensuring that all voices are truly heard and that power and control are equitably shared) (Savin-Baden & Wimpenny 2007, Gaffney 2008, Chevalier & Buckles 2019, Ozer 2017). The goal of PAR as envisioned for this project was to enable all interested stakeholders to participate equitably. While complete equity and inclusion were not always achieved, we nevertheless worked towards that goal.

researcher with no other relationship to the research study participants. Instead, the participants and I knew one another in our roles as teacher and students. This connection was important to ensure that participants felt comfortable asking questions about how their data would be used in the study; however, it also impacted data collection, as will be discussed in section 2.

**2. Methodology and methodological challenges.** In this section, I present the methodology used in collecting LFSE data with children ages 6-12 in Teotitlán, following best practices for ethical LFSE data collection outlined in Casillas & Cristia (2019) and Cychosz et al. (2020).

2.1. PARTICIPANTS. In the pilot phase conducted in January 2019, children were recruited for the study via radio announcement. Additional participants were recruited through their participation in the Zapotec language workshops for kids. I explained that I was interested in learning more about how kids in Teotitlán use all of the languages that they know—Zapotec, Spanish, English, or any other language—and that was why I was asking children to record their speech throughout the day using a small audio recorder. In the main study, conducted in summer 2019, I collected data from 10 participants, ages 6-12 (median age 10.5; 6 male, 4 female). These 10 participants represented 6 distinct households and 5 distinct homesteads.<sup>6</sup>

2.2. PROCEDURE. Participants were given a t-shirt with a special pocket sewn into the center of the shirt that was large enough to fit a Zoom H1 sound recorder (Figure 1). I showed participants how to start and stop the audio recorder, then I verified that they could perform this process on their own. We discussed the possibility that a participant might record something that they didn't want to record (such as an embarrassing conversation or a time they used the bathroom). Participants were asked to wear the special t-shirt with the Zoom H1 sound recorder in the pocket during the entirety of their waking hours, but were encouraged to turn off the audio recorder during any parts of the day that they didn't want recorded.<sup>7</sup> They were invited to participate in up to two days of recording before or during the workshops and two days after the workshops.<sup>8</sup> At the end of each recording session, participants and their guardians were also invited to advise me if they wanted anything to be erased. A few participants wanted to listen back through their recordings to see if anything should be deleted, and I happily obliged, in keeping with recommendations for best practices (Cychosz et al. 2020).

2.3. AFFORDANCES AND CHALLENGES. Unlike previous studies focusing either on early childhood language acquisition or on adult language use, participants in this study were children ages 6-12. This demographic of children is significantly different from both young children and adults in ways that fundamentally altered the type of data participants created. Child participants in early language acquisition studies (who are often under the age of 3) are largely unaware of their participation in this type of study, both in their inability to comprehend the purpose of such study, and in their tendency to forget that they are wearing a recorder. Adults, on the other hand, can comprehend the purpose of the studies in which they participate. If

---

<sup>6</sup> That is, in four cases two participants came from the same household. In one case, participants came from two different households that share a common patio, which I consider to be one homestead but two households.

<sup>7</sup> As we will see in section 2.3, participants did not record the entirety of their waking hours, but rather started and stopped the recorder frequently. Furthermore, participants did not always use the t-shirt to hold the recorder, but sometimes carried the recorder in their hand or used a different pocket or bag to hold it.

<sup>8</sup> Ideally, I would have recorded two days before the workshops and two days after, but because of difficulties in participant recruitment, some participants recorded their first two days after they had begun attending the workshops, as they found out about the study through participation in the Zapotec workshops.



Figure 1. Child (at left) wearing a special t-shirt with a yellow pocket for holding a Zoom H1 recorder to collect LFSE recordings while on a Zapotec class field trip. Participants' faces have been covered to protect their privacy.

they feel comfortable participating, they too often pay little attention to the recorder and will go about their day as normal. The children who participated in this study fall between those two groups: they are old enough to understand the basic premise of the study in which they are participating, but they are, for the most part, not mature enough to understand the details of how their data will be used or the motivations behind doing this type of research. For example, I explained to participants that only I or other researchers who are trained in handling this type of data would listen to their recordings, and that we would not react negatively to anything that they might consider embarrassing if a parent were to hear it (such as telling an inappropriate joke or making a rude comment). However, some participants found a great deal of their daily activities to be private or potentially embarrassing, and thus recorded only small segments of their daily activities. Furthermore, participants tended not to forget about the recorders, as the special t-shirts (although perceived as 'cool' by at least some participants, who proudly wore them to field trips and other outings) were often commented on by others, reminding participants of the recorders. For this reason, some participants preferred not to wear the shirts, carrying their recorder by hand or mounting it elsewhere, such as in a purse or other pocket. This, of course, negatively affected the audibility of the recordings. It also made it difficult or impossible in many instances to identify which voice corresponded to the study participant and which voices belonged to other children in the environment. However, it was essential for me to allow participants to feel as comfortable as possible while participating in the study, so participants were encouraged—but not required—to wear the shirts.

A further result of children not fully understanding the purpose of the study is that some participants seem to have focused their recordings on cases where they were speaking Zapotec. In some instances, this includes children apparently performing for the recorder, doing activities such as monologuing, rehearsing a dialogue with a sibling, or singing in Zapotec. Some

participants, when reviewing their recordings, suggested that recordings containing Spanish should be erased since they understood the purpose of the project to be related only to their Zapotec language use. This clearly indicates that my explanation to participants was insufficient for their understanding of the project. This could be, at least in part, because of my role as a teacher in the Zapotec language workshops and the fact that I am generally known by both participants and other community members to be someone who promotes the Zapotec language. In each case where a misunderstanding seemed to have taken place, I attempted to further explain the purpose of the study to the participants. Nevertheless, the recordings that participants gave permission for me to analyze often contain instances of apparent performances rather than natural speech, even after children had participated over multiple days and had multiple conversations with me and their parents about the study. This suggests that either I failed to fully explain the study, or that participants weren't interested in simply behaving "naturalistically" around the recorder, but rather preferred to continue recording performance-style speech.<sup>9</sup> Finally, my participation could have affected participants' willingness to record their full days out of fear of embarrassing themselves in front of me in my capacity as their teacher.

These potential complications had a significant effect on the amount of data that was collected through attempted day-long recordings, as most participants recorded only a small fraction of their day. Despite these methodological challenges, the benefits of LFSE remain. Although this method resulted in a smaller quantity of recorded time than expected (which could have been ~400 hours of recordings), it nevertheless resulted in a significant quantity of data (over 70 hours) which provides insights both into the instances of naturalistic language use that ARE captured, as well as the children's overall Zapotec language abilities as displayed in shorter, more performative recordings. Furthermore, LFSE recordings document key aspects of language socialization and use that are not typically part of documentation projects which often focus on adult speech in formal settings rather than the child-directed speech that is crucial for intergenerational language transmission (Meek 2011, Vallejos 2016). In order to bring languages into the home for use between parents and children, it is useful to have a robust record of how parents and children use the language that future learners can use as a resource, and this study shows that LFSE can provide at least some such data.

**3. Results and analysis.** One significant consequence of the methodological challenges presented above is that the data collected were non-uniform. The recordings can be roughly categorized into two main types, which I will call LONG and SHORT RECORDINGS. The long recordings are more similar in nature to typical LFSE recordings, featuring uninterrupted recordings as long as several hours, while the short recordings are much shorter (in some cases, only a few seconds). Moreover, the long recordings feature large stretches of silence and ambient noise, as well as conversations between various individuals in which participants may or may not participate. The short recordings, on the other hand, are almost exclusively either participant speech or conversations between the study participants. Finally, these short recordings, as will be discussed in subsection 3.2 below, are qualitatively performative in nature.

For the purposes of this paper, I consider recordings longer than 10 minutes to be LONG

---

<sup>9</sup> It seems that the latter may be true, as in some cases, participants recorded mostly performative Zapotec speech during the first set of days recording, then recorded a combination of performative and naturalistic speech in the second set of days recording. This suggests that my explanation of the purpose of the study between the first and second set of recordings did help to clarify things, but participants continued wanting to record their Zapotec performances.

RECORDINGS and those shorter than 10 minutes to be SHORT RECORDINGS. I have chosen 10 minutes as the cutoff as most recordings shorter than 10 minutes appear to contain performative speech in which the participant speaks to the recorder. While there are some recordings longer than 10 minutes that feature some such performative speech, there are many recordings of 10 or more minutes that do not and which are qualitatively similar to the longest recordings in that they contain long stretches of silence or speech between people other than the participants as well as environmental noise like TV and radio.

While it would have been ideal to separate the recordings based on whether they were or were not performative, creating criteria to separate the recordings based on qualitative standards alone proved difficult. Determining whether or not language was performative was particularly difficult without access to visual aspects of the scene, such as whether or not the participant was focused on the audio recorder or was simply speaking aloud while playing on their own. For example, while some performative recordings were easy to identify (such as when participants repeatedly recorded similar recitations, or uttered things like, “take two”), other recordings featuring monologues could have been either performative, or naturalistic recordings of a participant playing independently and talking to themselves. For that reason, I rely on a quantitative cutoff of 10 minutes to distinguish short and long recordings. In the following subsections, I analyze the data from both of these categories of recordings.

3.1. LONG RECORDINGS. I sampled a five second segment beginning at the 15-second mark<sup>10</sup> of each minute for all recordings over 10 minutes for which I was not present.<sup>11</sup> Each clip was then labeled as containing or not containing speech. In clips containing speech, I annotated the clip with who was speaking and in what language. I distinguished whether speakers were the participant, adults, or other children. If I could not determine the quantity or identities (i.e. participant vs. adult vs. other child) of the speakers, or the languages they were using, I discarded the clip. In this paper, I present data from six of the ten participants, providing one participant’s data from each household that participated in the study.<sup>12</sup> This procedure resulted in 1676 total clips, of which 784 contained speech that was tagged for language and participants.

Table 1 provides a quantitative overview of the data, in which several points stand out. First, participants varied greatly in how much time they chose to record, with Daniel<sup>13</sup> creating 16 recordings with a combined length of over 16 hours and Leo creating only one 28-minute recording.<sup>14</sup> Given that the data are quite limited, it is difficult to form any conclusive generalizations, though we can make some limited observations. First, participants’ language input from adults is in Spanish a minority of the time for four out of the six participants. Furthermore, Yessica, who had the lowest amount of exposure to Zapotec from adults in her environment, may be an outlier; she had only recently returned to Teotitlán after having been raised in

<sup>10</sup> Given that the participants in this study also participated in sociolinguistic interviews I conducted, they had heard me give spoken metadata for those recordings. Some copied that format, making statements along the lines of, “Today is August 10. I am (Name) and I am here with (Name) to record.” As a result, the first 15 seconds of some recordings contain very similar statements in Spanish.

<sup>11</sup> Some participants also turned on their recorder during my interviews with them or their family members, and I have not included that data in my analysis.

<sup>12</sup> I have included data from one participant per household given that in many cases, participants from the same household produced nearly identical recordings. Siblings often turned on both recorders simultaneously in the same place.

<sup>13</sup> All participant names are pseudonyms.

<sup>14</sup> Leo created other recordings, but all were less than 10 minutes long and will be discussed in the next section.

	Participant			Adult			Child			Spch clips	Clips
	Z	S	Z%	Z	S	Z%	Z	S	Z%		
Daniel	49	104	32%	53	57	48%	0	6	0%	408	993
Giselle	2	7	22%	121	59	67%	0	20	0%	128	242
Yessica	0.5	90.5	0.5%	14.5	44.5	25%	0	56	0%	131	195
Osvaldo	24	18	57%	16	3	84%	2	8	20%	64	153
Elena	1	20	5%	14	9	61%	0	11	0%	36	65
Leo	1.5	4.5	25%	14	3	82%	1	1	50%	17	28

Table 1. Quantitative analysis of clips from long recordings. Z and S represent the count of instances of Zapotec and Spanish spoken in a clip respectively; Z% represents the percentage of total instances that Zapotec was spoken for that individual type. For example, the “Adult Z” of 53 reported for Daniel means that there were 53 instances of an adult speaking Zapotec in a clip. “Adult Z%” of 48% indicates that those 53 instances made up 48% of all adult speech in Daniel’s recordings. To calculate percentages, I counted each instance of an individual speaking in a given language in a given clip as a quantity of 1. If an individual used both Spanish and Zapotec, I counted that as 0.5 for each language. The total counts of speech used by participants, adults, and other children in the recordings are the sum total, and percentages are calculated from that. Clips are the total number of clips from recordings over 10 minutes created by each participant. Speech clips (“Spch clips”) are the count of how many clips contained taggable speech. “Participant” means the participant was speaking; “Child” means that a child other than the participant was speaking.

another town where Zapotec was not widely spoken. Her history outside of Teotitlán may have impacted her Zapotec exposure and patterns of language use within her family. Nevertheless, all participants recorded instances of Zapotec being spoken by adults in their environment, indicating that all participants had access to at least some Zapotec language input.

In contrast, participants’ language output is heavily skewed towards Spanish. Five of the six participants speak Spanish in a majority of clips where they are speaking. In the case of Osvaldo, the one participant who speaks Zapotec in a majority of his speech clips (57% of the time), one of his five recordings (representing 21 of his 64 speech clips) while over 10 minutes in length is qualitatively similar to the short recordings, as in it he recites items from semantic categories like body parts, plants, numbers, colors, and verbs out loud to himself in Zapotec. In addition, children’s language use with one another takes place overwhelmingly in Spanish. For four of the six participants, none of the speech from other children is in Zapotec. These quantitative results have implications for future language revitalization work, as they suggest that despite exposure to Zapotec, children are not using the language. Thus, future revitalization efforts may wish to consider not just teaching or increasing children’s exposure to the language, but also promoting and supporting language USE, particularly between children.

Qualitatively, we can observe participants in the recordings participating as active overhearers of speech between adults, as they agree and disagree with adults using phrases like “uh-huh” and “uh-uh” that signal agreement and disagreement respectively in both Spanish and Zapotec. The conversations to which participants are exposed include both monolingual Zapotec and bilingual Spanish-Zapotec conversations between adults. Participants’ active par-

ticipation (even as over-hearers) suggests that they have at least some level of Zapotec comprehension. There is also some monolingual Spanish adult conversation in the recordings, but it is largely at an event outside of the home. This fits with the findings from reported language use that Zapotec may be used in familiar and home settings more often than in public settings.

Parents use both Zapotec and Spanish with their children in the recordings. However, children respond overwhelmingly in Spanish. There are also instances of parents actively teaching Zapotec to their children as well as passing down their linguistic ideologies, engaging in activities like explaining to their children why Zapotec is important. In one case, this involved the parents asking the participant to repeat (in Zapotec) the phrase “I like to learn Zapotec because I’m from Teotitlán,” and thus both teaching Zapotec language and socializing their child into beliefs about the links between being from Teotitlán and speaking Zapotec.

While most participant responses were in Spanish, children did incorporate some of the emergent Zapotec skills that they do have in conversations with adult Zapotec speakers. For example, participants produced Spanish words that had been phonologized into Zapotec:

(1) Spanish and Zapotec (SCOIL 2016-02)

me costó *dineer*<sup>15</sup>  
‘it cost me *money*’

In the example above, a child was conversing with his parent when he uttered the sentence in (1), in which he took the Spanish ‘dinero’ and pronounced it following typical phonological processes for borrowing words into Zapotec, *dineer*. Interestingly, after uttering this sentence, produced largely in Spanish, the parent followed up by suggesting that the child could use the Zapotec word for ‘money,’ *medih*. It seems that the child’s phonologized borrowing was interpreted as indicating an interest in learning and using Zapotec. Thus, by using the phonologized Spanish loanword, the child was able to engage his parent and further his learning of Zapotec.

Similarly, another participant was being addressed by an adult in Zapotec, and the child used a code-switch, inserting the Zapotec word for ‘green’ into an otherwise Spanish utterance:

(2) Spanish and Zapotec (SCOIL 2016-02)

con uniforme verde, con uniforme *bixui*  
‘with a green uniform, with a *green* uniform’

In this example, the child first responds with a monolingual Spanish phrase, but then repeats it and code-switches the Zapotec word for ‘green’ into the Spanish sentence. I find this example to be illuminating as it shows a case where a child was able to use his emerging Zapotec language skills (even though it involved using only one, uninflected word in Zapotec). Following this utterance, the adult continued addressing the participant in Zapotec.

Crucially, in all of the cases that I was able to hear in the recordings, if a child initiated a conversation using any amount of Zapotec—whether a full Zapotec phrase, a code-switched Zapotec utterance within a Spanish frame, or a Zapotec-phonologized borrowed word from Spanish—they received a response from their parent in Zapotec. On the other hand, if the child responded in Spanish, the adult sometimes responded in Spanish, and sometimes responded in Zapotec. This suggests that one crucial aspect of building sustainable intergenerational language transmission is empowering Zapotec learners to begin conversations with

<sup>15</sup> In this paper, I present Zapotec (including borrowed words phonologized to Zapotec and Zapotec elements in Spanish-Zapotec code-switches) in italics and Spanish in regular font.

adults in Zapotec using any Zapotec language skills that they may have. As we will see in the following section, the participants in this study showed notable abilities in their Zapotec language production when they recorded themselves privately, though these abilities were not always showcased in more public spaces like the Zapotec classroom or in conversations with adult Zapotec speakers. Thus, for language revitalization projects to result in more widespread language use among children in Teotitlán, we may need to focus more on not simply building language skills, but rather building learners' investment in USING their language skills.

3.2. SHORT RECORDINGS. Unlike long recordings, which were generally more naturalistic, short recordings were overwhelmingly performative in nature. They featured instances of children listing words from domains covered in the workshops (such as colors, animals, and verbs), as well as songs, monologues, and dialogues between children. The monologues and dialogues seem to be performative in nature, as they were often rehearsed and recorded repeatedly, as highlighted by one participant's introduction in a recording:

(3) Spanish and Zapotec (SCOIL 2016-02)

Segunda toma. Maestra Julia Nee *rului, ruluian, chuulmaate' ruluian*.

'Take two. Teacher Julia Nee *teach, teaches, in a very cool way she teaches.*'

In this example, we see that the participant indicates that this utterance is rehearsed by calling it 'take two.' This example also highlights the complication of my dual role as teacher and researcher, as this monologue is not only performative, but also mentions my status as a teacher.

Similar performances can be found in conversations among children. In example (4), a group of children is having a conversation in Zapotec, but when one participant doesn't know what to say in Zapotec, another assists him by suggesting (in a whispered voice) what to say:

(4) Spanish and Zapotec (SCOIL 2016-02)

Mario: *Xa yu'u?*

*'How are you?'*

Giselle: *Guenka. Lui ga? Xa yu'u?*

*'Good. And you? How are you?'*

Mario: *Guenka.* [Pause]

*'Good.'* [Pause]

Giselle: *Kayaua tu laa manzana. Lui ga? Xi kayau?*

*'I'm eating an apple. And you? What are you eating?'*

Juan: *Sí, es cierto, 'Ketchidi'.*

*'Yes, it's true, 'Nothing.'*

Mario: *Ketchidi.*

*'Nothing.'*

Again, this conversation seems to be performative. For one, the topic is somewhat abnormal. Giselle is narrating what she's eating, despite the fact that she is audibly eating an apple, and there should be no confusion on the part of Mario what she's eating. Furthermore, there is a long pause between the opening 'How are you?' and the discussion of what they're eating, as if Giselle is thinking of what she might ask about in Zapotec. Finally, when Giselle asks Mario what he's eating, he isn't sure how to respond, and falls back on help and encouragement from Juan, who supplies the answer of 'nothing' in Zapotec.

While these short recordings are performative, they are nevertheless helpful in providing insight into learners' language abilities. In both examples (3) and (4), the Zapotec produced by the participants includes forms that were not covered in the workshops. Participants integrated vocabulary acquired elsewhere (such as *chuulmaate*') along with verbal forms we had not yet practiced (such as the progressive *ka-*). In the case of verbal forms, moreover, the participants produced the progressive verb forms with their enclitic person markings as we would expect an adult speaker to, indicating that they have productive control of such forms.

### 3.3. COMPARISON WITH REPORTED LANGUAGE USE AND RESEARCHER OBSERVATIONS.

The comparison of LFSE data with reported language use<sup>16</sup> presented in table 2 shows, generally, an underreporting of exposure to Zapotec, as 5/6 participants and 4/5 parents reported less exposure to Zapotec than was recorded in the LSFE data. Only Giselle and her parent reported Zapotec dominance in both hearing and speaking; however, her observed speech was only 22% Zapotec. This is notable because Giselle demonstrated creative use of fluent Zapotec in the recordings. Interviews with her and her parents suggest that she has been spoken to in Zapotec since birth, and she is capable of conversational fluency in Zapotec. However, she uses more Spanish than Zapotec in her daily interactions. Giselle's case highlights the fact that even children with strong Zapotec skills may not use the language. Thus, increasing Zapotec language use requires more than simply teaching children Zapotec language skills, as some learners may already have those skills. Instead, we must also foster environments where children feel comfortable and invested in actually using the language. In Giselle's case, for example, the private setting of talking to the recorder (both alone and with other children) provided an environment where Giselle was able to showcase a fuller range of her Zapotec abilities than she did either in the classroom or in the long LFSE recordings where adults were present.

I would also like to consider why adults' perceptions of language use by children differ from children's own perceptions of language use. While the differences may be due to reporting error, they may also be related to an idea that some parents have expressed to me that their children do in fact know Zapotec already and will eventually come to speak it as they come of age and take on more responsibilities in community activities that are carried out in the language (such as parties and community meetings). This is a viewpoint that I have not heard expressed by children, who talk about the importance of studying and learning Zapotec in order to speak it. It is possible that adults consider their children to be emergent Zapotec speakers in a way that the children themselves do not, but which may be more in line with the results of this study. This view from some parents suggests that increasing children's language use doesn't necessarily require increasing children's language skills (which parents may perceive them to already have), but rather requires access to contexts in which the language can be felicitously used (such as community events where adolescents and adults, but not children, may be expected to speak publicly in Zapotec). While the focus of some parents on contexts for Zapotec use is consistent with the suggestions provided in this study, it is nevertheless important to note that some children do seem to need (or at least want) additional support in their acquisition of Zapotec language skills as well as contexts for use. Overall, these mixed reports

---

<sup>16</sup> Data on reported language use was collected through one-on-one interviews in summer 2019. Following the procedure outlined in Bedor et al. (2012), participants and their parents were asked to say what the child participant does during each hour of a typical day as well as what language they use more during that hour, who they speak to at that time, what language that person uses, and what language the child participant responds in.

	Parent report		Participant report		LFSE observation	
	Spoken	Heard	Spoken	Heard	Spoken	Heard
Daniel	0%	39%	7%	17%	32%	45%
Giselle	54%	82%	85%	85%	22%	61%
Yessica	NA	NA	0%	0%	1%	13%
Oswaldo	19%	21%	56%	36%	57%	62%
Elena	0%	14%	0%	13%	5%	41%
Leo	23%	56%	27%	19%	25%	79%

Table 2. Comparison of parent reported, participant reported, and observed language use, summer 2019. Reports are divided into reports of language spoken and language heard. Percentages represent percentage of time speaking and hearing Zapotec respectively. Yessica’s parent did not provide a report of Yessica’s language use.

from both observational and reported use suggest the complicated nature of reaching a deep understanding of how and why Zapotec is used by each unique participant in the study.

Of course, given the small LFSE sample size, it is possible that the differences between observed and reported language use could be due to sample bias in the LFSE recordings. They could also be due to inaccuracies in the self-reports, which may be particularly unreliable in endangered language contexts where social pressures may encourage both over- and under-reporting of language use depending on the specific social circumstance (Dauenhauer & Dauenhauer 1998). Thus, while we must take these data with a grain of salt, it is important to note the difference between children’s Zapotec ABILITIES when they are in an environment where they feel comfortable speaking versus children’s Zapotec USE in everyday contexts.

**4. Efficacy of Zapotec language workshops.** To evaluate the efficacy of the Zapotec language workshops, I compared recordings created by participants before and after the language workshops. I was able to collect recordings from five participants from before and during the workshop, which can then be compared with the recordings from after the workshop to see if there are any notable changes in behavior that might be the result of the workshop. Again, some participants were members of the same household and thus recorded very similar data; here I present data from one child’s recordings per household. Unfortunately, in one household the participant did not record very much before the workshop, so I present only qualitative observations on that data. In two other cases, there was sufficient evidence to compare the audio clips from before and after the workshop; the results are summarized in table 3. It should be noted that I have condensed the distinction between long and short recordings, and I include data from all recordings greater than 2 minutes collected from both participants. I included shorter recordings given that all of the recordings collected from Giselle before the workshop were short. As I showed above, the short clips tend to be more performative and the longer clips more naturalistic, so by including shorter clips in the post-Zapotec workshop data I hope to make those two sets of recordings more comparable.

Based on the data in table 3, Giselle’s proportion of Zapotec exposure and use appears to have decreased after the workshop. This is likely due to differences in what she recorded during the different days. In the recordings from before the workshop, for example, there are some longer recordings during which adults speak together in Zapotec, apparently at home; in

	Participant			Adult			Child			Rec.	Spch clips	Clips
	Z	S	Z%	Z	S	Z%	Z	S	Z%			
G before	14	1	93%	98.5	2.5	98%	10	2	83%	10	75	88
G after	16.5	8.5	66%	122	59	67%	15	20	43%	11	148	278
E before	0	57	0%	7	18	18%	0	5	0%	9	69	91
E after	2.5	29.5	8%	30.5	13.5	69%	0	18	0%	5	59	90

Table 3. Comparison of observed Spanish language use before and after the Zapotec workshop for Giselle (G) and Elena (E), summer 2019, including both long and short recordings. Each 5-second-long clip was tagged as including speech by the participant in Zapotec or Spanish, speech in Zapotec or Spanish from an adult, or speech in Zapotec or Spanish from another child (who is not the participant). I have also included counts of the number of recordings analyzed, the number of speech clips annotated, and the total number of clips annotated (including clips without speech).

the recordings from after the workshop, there are recordings from an event outside the home in which adults speak both Zapotec and Spanish with one another and children mostly use Spanish. It seems likely that differences in adult language use between the two recording periods are related to the differences in setting, with the home environment seeing more Zapotec use than the public event. This is consistent with the results from the reported language use interviews, in which Zapotec was more often used in home settings than visits to other places. Qualitatively, Giselle’s short recordings from before and after the workshops are very similar. They are performative in nature and include practiced dialogues in Zapotec as well as songs. In the recordings from after the workshop, these songs include those apparently created by the children, as well as some songs by Gario Ángeles, a Zapotec singer and songwriter, that we sang together during the course.

Elena’s language use does seem to have changed slightly over time, with an increase in her Zapotec use from no use to some use (about 8% of her speech).<sup>17</sup> Furthermore, there was an increase in her exposure to Zapotec from adults—from 28% to 69%. As I also suggested in the case of Giselle, it is possible that this difference reflects a change in the genre of recordings collected, as the recordings from after the workshop are on average about twice as long as the recordings from before the workshop. However, Elena’s parent reported in their interview that Elena showed an increased interest in learning Zapotec following the workshops, evidenced by an increase in how often she asked how to say things in Zapotec. The recordings seem to support such a change, as Elena seems to be speaking Zapotec with her parent only after the Zapotec workshop. In the recordings from before the workshop, the parent appears to try teaching Zapotec by repeating the names of objects mentioned by Elena, but Elena continues conversing with her parent in Spanish, asking in Spanish ‘¿dónde?’ (‘where?’). After the workshop, on the other hand, when the parent speaks in Zapotec with Elena (at least in one instance), Elena responds in Zapotec, asking ‘*xuur?*’ (‘when?’) and incorporates words from Zapotec into responses throughout the conversation. Interestingly, there is no difference in the language used by other children around Elena, which is entirely Spanish both before and after the language workshop, despite the fact that some recordings include the voices of other

<sup>17</sup> It is difficult to say whether this represents a meaningful difference or whether it is the result of sample bias.

attendees of the Zapotec workshop.

Finally, one other family had a participant who recorded both before and after the workshop, but the recordings he produced from before the workshop were all under one minute long. These recordings were all performative in nature, as the child and his sibling spoke in Zapotec, repeating common phrases and questions as in a rehearsed dialogue. For example, they stated their names, asked one another what they were doing, and responded. In the recordings from after the language workshop, there was an abundance of similarly performative recordings, but they contained a greater diversity of content. For example, the participant recorded himself counting in Zapotec, naming plants, body parts, colors, verbs, and his favorite things, asking if a movie has finished, narrating his own actions, and asking another child, *kali gueu?* ‘where did you go?’ along with basic phrases such as the ones recorded before the workshop. In addition, one child can be heard helping the other child to learn a new word in Zapotec, showing the type of collaborative learning that was emphasized in the language workshop. Of course, the increase in diversity of Zapotec language content could be an artifact of the greater quantity of time that the child recorded, but it is also possible that it reflects an increase in Zapotec abilities or at least an increase in a willingness to try using Zapotec.

**5. Conclusion.** Overall, while the challenges of working with kids ages 6-12 resulted in atypical LFSE data, those data nevertheless provide a useful look into some aspects of participants’ language environment and language abilities. My analysis of the data suggests that Zapotec learners may be more proficient than they or their parents report, or than they demonstrate in public spaces like the language revitalization classroom. By inviting children to record their language use, at least in this case, we provide an environment where children feel more comfortable speaking in Zapotec. They display a wider range of Zapotec language, which in turn provides greater insight into their overall language abilities. It also provides some documentation of how participants and their families use language in Teotitlán.

While the limited quantity of LFSE recordings participants created do not provide a clear picture of participants’ naturalistic language use generally, they nevertheless contribute to a clearer picture of language use when paired with other data, including reported language use and researcher observations. Furthermore, the recordings show that some aspects of the language workshops (such as songs and strategies for language learning) were being taken up by participants at home in the days following the workshops. They also highlighted that learners’ use of Zapotec resulted in greater Zapotec input from adults, indicating that strengthening learners’ confidence in starting Zapotec conversations may enable them to use the language more frequently. Moreover, by revealing a fuller range of participants’ Zapotec language skills—which were generally greater than what they had demonstrated in the language revitalization classroom—the LFSE recordings reinforce the fact that sustainable language revitalization requires more than simply teaching language skills. Instead, we must also focus on creating spaces where learners feel comfortable using their growing language skills.

## References

- Bedore, Lisa M., Elizabeth D. Peña, Connie L. Summers, Karin M. Boerger, Maria D. Resendiz, Kai Greene, Thomas M. Bohman & Ronald B. Gillam. 2012. The measure matters: Language dominance profiles across measure in Spanish/English bilingual children. *Bilingualism: Language and Cognition* 15(3). 616–629. <https://doi.org/10.1017/S1366728912000090>.
- Casillas, Marisa & Alejandrina Cristia. 2019. A step-by-step guide to collecting and analyzing

- ing long-format speech environment (LFSE) recordings. *Collabra: Psychology* 5(1). 24. <https://doi.org/10.1525/collabra.209>.
- Chevalier, Jacques. M. & Daniel J. Buckles. 2019. *Participatory action research: Theory and methods for engaged inquiry*. New York: Routledge.
- Cychosz, Margaret, Rachel Romeo, Melanie Soderstrom, Camila Scaff, Hillary Ganek, Alejandrina Cristia, Marisa Casillas, Kaya de Barbaro, Janet Y. Bang & Adriana Weisleder. 2020. Longform recordings of everyday life: Ethics for best practices. *Behavior Research Methods* 52. 1951–1969. <https://doi.org/10.3758/s13428-020-01365-9>.
- Czaykowska-Higgins, E. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian indigenous communities. *Language Documentation & Conservation* 3(1). 15–50.
- Dauenhauer, Nora Marks & Richard Dauenhauer. 1998. Technical, emotional, and ideological issues in reversing language shift: Examples from Southeast Alaska. In Lenore A. Grenoble & Lindsay J. Whaley (eds.), *Endangered languages: Language loss and community response*, 57–98. Port Chester, NY: Cambridge University Press.
- Fishman, Joshua. 1991. *Reversing language shift: Theory and practice of assistance to threatened languages*. Clevedon, England: Multilingual Matters.
- Gaffney, Michael. 2008. Participatory action research: An overview, What makes it tick? *Kairaranga* 9. 9–15.
- Gobierno del estado de Oaxaca. 2010. Plan municipal de desarrollo: Teotitlán del Valle. Tech. rep.
- INEGI. 2020. Banco de Indicadores. <https://www.inegi.org.mx/app/indicadores/?ag=20546#divFV6207019034>.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Leonard, Wesley & Erin Haynes. 2010. Making “collaboration” collaborative: An examination of perspectives that frame linguistic field research. *Language Documentation & Conservation* 4. 268–293.
- Meek, Barbara. 2011. Failing American Indian Languages. *American Indian Culture and Research Journal* 35(2). 43–60. <https://doi.org/10.17953/aicr.35.2.m272376nl73v332t>.
- Montoya, Ignacio L. 2020. Enabling excellence and racial justice in universities by addressing structural obstacles to work by and with people from racially minoritized communities: Response to Charity Hudley et al. *Language* 96(4). e236–e246. <https://doi.org/10.1353/lan.2020.0075>.
- Moseley, Christopher (ed.). 2010. *Atlas of the world’s languages in danger*. Paris: UNESCO Publishing, 3rd edn.
- Ozer, Emily J. 2017. Youth-led participatory action research: Overview and potential for enhancing adolescent development. *Child Development Perspectives* 11(3). 173–177. <https://doi.org/10.1111/cdep.12228>.
- Savin-Baden, Maggi & Katherine Wimpenny. 2007. Exploring and implementing participatory action research. *Journal of Geography in Higher Education* 31(1). 331–343. <https://doi.org/10.1080/03098260601065136>.
- Vallejos, Rosa. 2016. Structural outcomes of obsolescence and revitalization: Documenting variation among the Kukama-Kukamirias. In Gabriela Perez-Baez & Jorge Emilio Rosés Labrada (eds.), *Language documentation and revitalization in Latin American contexts*, 143–164. Berlin: Mouton de Gruyter.
- Whyte, William Foote, D. J. Greenwood & Peter Lazes. 1991. Participatory Action Research: Through Practice to Science in Social Research. In William Foote Whyte (ed.), *Participatory action research*. Thousand Oaks, CA: Sage Publications.