Daylong egocentric recordings in small- and large-scale language communities: A practical introduction

Marisa Casillas[a,*] **and Kennedy Casey**[b]
[a]Comparative Human Development Department, University of Chicago
[b]Department of Psychology, Princeton University
*Corresponding author. e-mail address: mcasillas@uchicago.edu

Contents

Abstract

Daylong egocentric (i.e., participant-centered) recordings promise an unprecedented view into the experiences that drive early language learning, impacting both assumptions and theories about how learning happens. Thanks to recent advances in technology, collecting long-form audio, photo, and video recordings with child-worn devices is cheaper and more convenient than ever. These recording methods can be similarly deployed across small- and large-scale language communities around the world, opening up enormous possibilities for comparative research on early language development. However, building new high-quality naturalistic corpora is a massive investment of time and money. In this chapter, we provide a *practical* look into considerations relevant for developing and managing daylong egocentric recording projects: Is it possible to reuse existing data? How much time will manual annotation take? Can automated tools sufficiently tackle the questions at hand? We conclude by outlining two exciting directions for future naturalistic child language research.

Advances in Child Development and Behavior
ISSN 0065-2407, https://doi.org/10.1016/bs.acdb.2024.05.002

1. Introduction

If only we could experience the world as children do; we could much better understand what they learn, when learning happens, and how it happens. Children develop within environments that are structured by their caregivers, institutions, and societies (see, e.g., Bronfenbrenner, 1979; García Coll et al., 1996; Spencer, 2007). Crucially, they navigate these structured environments on the basis of their current experiences, skills, and understanding—children's perspectives on the world are different from adults'—sometimes quite literally (e.g., Fausey, Jayaraman, & Smith, 2016). Naturalistic, especially egocentric (i.e., from the child's point of view), data are as close as we can come to capturing these experiences for the purpose of explaining how learning happens. This chapter gives a practical introduction to the collection, annotation, and archival of naturalistic developmental data, especially data collected with a *daylong* and *egocentric* (i.e., child-centered) perspective.

Long-format (often egocentric) recordings of children's natural environments have ushered several key insights into developmental science. For example, children's home environments offer continuous, multimodal perceptual experiences that become more sensible to the infant through the progression of neurotypical motor development; this perceptual curriculum carves meaning out of the noise, equipping children to learn the names for common objects (Iverson, 2010; Yu & Smith, 2012; Yu, Zhang, Slone, & Smith, 2021). A second example is that, both in learning to walk and learning to talk, children practice their new skills often—and they falter for a long time: Adolph et al. (2012) estimate that early walkers in the U.S. take an average of 2368 steps and 17 falls per hour during free play. Typically-developing 9–12-month-old infants in the U.S. produce more than 350 vocalizations per hour during everyday activities, only around 17% of which are speech-like syllables (Patten et al., 2014). These estimates do not give a full account of babble practice either: infants also engage in so-called "crib monologues" during which they produce large volumes of babble on their own (H. L. Long et al., 2022; McGillion et al., 2017; Nelson, 1989). These examples illustrate the value of naturalistic egocentric data. By observing development in action—in the home, and from the child learner's point of view—we gain a sense of children's moment-by-moment experiences and can then better reason about how they take in and use information from the environment around them.

In this chapter, we balance description of the incredible potential of daylong egocentric data with its known weaknesses. We do so with special attention to *language* and to *smaller-scale* projects. Our hope is to share with others what we have learned in managing several egocentric datasets—the good, the bad, and the ugly. We first discuss the value of building on existing data. We then dive into important considerations for developing new corpora, giving examples of the different issues we have faced in creating new datasets for smaller- and larger-scale language communities. We then briefly discuss the current state of automated annotation tools applicable to daylong egocentric recordings for child language research. Finally, we end the chapter by discussing what we consider to be some of the most exciting future directions for work in this domain.

2. Using existing data

Whenever possible, building on existing data is a smart choice. Most practically, it saves researchers time and money. It also fortifies the existing network of data contributors and re-users in the language sciences, improves the quality of existing datasets, and thus helps ensure that these valuable resources will be available to students and scientists long term. As a field, developmental researchers have called for and demonstrated an enormously successful history of data sharing (Adolph, Gilmore, Freeman, Sanderson, & Millman, 2012; Gennetian, Tamis-LeMonda, & Frank, 2020; Gilmore, 2022; Kosie & Lew-Williams, 2022) thanks to the generosity of participating families, contributing researchers, and the scientific leaders who implement and maintain digital sharing infrastructures (Frank, Braginsky, Yurovsky, & Marchman, 2017; Gilmore, Adolph, & Millman, 2016; MacWhinney, 2000; VanDam et al., 2016; Zettersten et al., 2023).

Whether you are using existing data from smaller- or larger-scale language communities, many of the same key issues hold for planning the project:

- First, you may need to seek out data from multiple sources to get the collection of recordings or transcripts required for your research question. The public repositories cited above—CHILDES, HomeBank, Databrary, etc.—are excellent sources of naturalistic data, but many researchers have private collections that they may be willing to share if you reach out.
- Second, sharing is not always simple—expect that some researchers may need to set up institutionally ratified data use agreements, which can involve the IRB and, sometimes, legal teams associated with each institution.

- Third, try to be mindful of the data contributor—make it clear from the outset whether they have authorial rights, what they might need to contribute in addition to simple data access rights, and what benefit they might expect from sharing their private collections.
- Fourth, have a back-up plan for when the data you need do not exist. For example, in a current project we were able to find existing recordings for all but two age-sex combinations that we needed; we plan to make our own recordings to fill just these two gaps.

Finally, carefully consider how new annotations might be added to existing data structures. If the current annotation is in a time-aligned format (e.g., transcripts in ELAN (https://archive.mpi.nl/tla/elan) or time-stamped CHAT (MacWhinney, (2019), https://talkbank.org/manuals/CHAT.pdf)), it would be most useful to provide *time-aligned* data for new annotations (e.g., not just counts of conversational blocks, but their time-linked onsets and offsets). If you are updating existing annotations or providing alternatives, consider how you can clearly document the differing versions of the data.

3. Building new corpora

High-quality corpora (i.e., datasets; in our case, collections of day-long egocentric recordings and annotations) allow deep insights into naturalistic behavior. These information-rich sources of naturalistic data can be used again and again to inspire and examine many different research questions. For this reason, resources like CHILDES (MacWhinney, 2000), HomeBank (VanDam et al., 2016), and Databrary (Gilmore et al., 2016) have been crucial for the advancement of the child language sciences.[1] New corpora are a gift to the research community. When researchers opt for manual annotation of new recordings (and share their data), then the field can immensely profit from the hours of toil put into each and every transcript, and tool developers can use manually generated data to improve the outlook of future automated annotation approaches. This community-oriented attitude toward new corpus creation is essential because researchers unable (or unmotivated) to record and annotate their own naturalistic

[1] How big is this impact? Based on Google Scholar citations of the CHAT manual (MacWhinney, 2019, which does not count direct citations of individual CHILDES corpora), CHILDES has conservatively been used in the production of around 11,000 papers (December 2023).

corpora can still produce groundbreaking new ideas on the basis of existing information-rich data. But corpus creation is *tough* work. Anyone who has transcribed or otherwise manually annotated naturalistic data, especially long-format and egocentric data, will tell you: it is time-consuming, challenging, and tedious.

Researchers who are looking to build or expand corpora can expect at least three types of investment: time, money, and "planning effort." We briefly outline each of these investments below, with considerations for those collecting their own data (or not) and completing manual annotations (or not).

- The first investment is a **monetary** one. If you are recording your own data, the first costs will be equipment and participant compensation. However, we find that these costs are *vastly* outweighed by the hundreds (sometimes thousands) of hours of paid research assistant time needed to manually annotate the data afterwards. Even if you plan to use automated annotation, you will likely need to budget for manual validation of your automated annotations to ensure that your data are of the quality you require.
- With respect to **time** investment, corpus development can be painfully slow. For those recording their own data, it may take quite some time to recruit the desired sample when the targeted population is difficult to access for some reason (e.g., is relatively small, far away, or requires a rare trait/experience). When it comes to manual data annotation (including validation for automated tools), a long wait may be unavoidable if it is difficult to find appropriate research assistants (e.g., those who have knowledge of some specific language). Further, if there are special considerations around how to ethically conduct the work (e.g., the transcribed data need to be archived in a specific way), these additional steps may further slow down the process of data production.
- The third investment is "**planning effort.**" From the start of the project, there should be a clear vision for the long-term utility of the data that includes: strategies for participant consent (i.e., to maximize future data re-use), documentation of all data collection and data annotation procedures (i.e., open training manuals), and a data format and variable selection that balance current effort (what is easiest and fastest right now?) with future potential (what can I afford to add that will make these data maximally reusable?). Whenever possible, research products (e.g., journal articles) should be planned to be produced along the way, rather than after 2 + years of initial investment. This is especially the case for early-career researchers, for whom a significant time investment (perhaps years) presents significant risk.

No matter how well-planned a project is, unexpected issues that impact each of these three investment types will inevitably arise. To name a few issues that we have faced when dealing with audio and video data (more on the latter below): audio/video misalignment, video encoding and playback trouble, frequent software crashes, delays in data upload/download from cloud storage, and substantial individual differences in research assistants' speed and accuracy of annotation. Some specific issues relating to daylong egocentric audio and video data have been: children pulling at or removing recording devices, limited camera angles, over- and/or under-exposed lighting, noisy or unusable audio due to movement, and occasional motion sickness thanks to a toddler's wobbly run across the room. This is all to say that researchers working on similar naturalistic datasets should leave significant resources for flexible problem solving, re-starts, and re-dos.

3.1 Project-specific considerations

This chapter adds a unique perspective to the existing literature on daylong recording methods (e.g., Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2019; Casillas, 2023; Cristia et al., 2021; Cychosz & Cristia, 2022; Gautheron, Lavechin, Riad, Scaff, & Cristia, 2020; Montag, 2020): corpus building in the context of smaller- vs. larger-scale language communities. In the following two subsections, we illustrate how priorities and anticipated issues differed when we built corpora for two smaller-scale, subsistence communities (Tseltal and Yélî Dnye) compared to a larger-scale, urban/suburban sample (U.S. English).

3.1.1 Ex. smaller-scale language communities: Tseltal and Yélî Dnye

We first describe the primary challenges associated with the development of our two smaller-scale language datasets (HomeBank; Casillas, Brown, & Levinson, 2017). We focus especially on the challenges associated with manual transcription and annotation of these data.

These corpora began with a European-funded project that aimed to comparatively examine the early language environments of children in two communities: a Tseltal Mayan community in Southern Mexico and a collection of villages on Rossel Island, which is at the far end of Papua New Guinea's Louisiade Archipelago. A major goal of the project was to document young children's at-home language experiences, motivated by prior ethnographic accounts describing how Tseltal adult-adult talk is prioritized over child bids for social attention, but how Rossel adults are likely to center their shared social attention on children (see also "non-

child-centric" vs. "child-centric" social environments; Brown & Casillas, 2020; Brown, 2011, 2014). Because the purpose of the project was to get a naturalistic view of children's home language experiences, we needed a child-perspective, long-format home recording. But we also knew we would need to plan for a great deal of manual (i.e., non-automated) transcription. Why?

We were compelled to do manual transcription for several important reasons: (1) We wanted to be able to answer questions about speech *content*, which at the time (2015–2016, but still now) was not possible with available automated tools. Our desire was especially strong because we anticipated that automatically derived "quantity" measures of linguistic input would only take us part of the way in understanding children's language environments. (2) Given that we were asking for time and resources in someone else's community and that the work involved some personal risk for the on-site research team (e.g., illness, injury), we wanted to invest in a lasting resource—one that could be a starting point for future research (by us or by others) and that could potentially support long-term language maintenance for the target populations. (3) As visitors to these communities, we wanted to stay open to new ideas and new questions. Our collaborative on-site manual transcription process, in which we produced transcripts side-by-side with local community members, gave us one fruitful method for engaging in thoughtful observation of natural interactions. Last but not least, (4) manual transcription gave us an excellent basis for language learning, establishing relationships with community members, training paid local research assistants with new skills, and giving the broader community context for the focus of our research project (i.e., child language).

In the initial datasets, a 1-min clip of audio data took an average of 50–60 min to transcribe,[2] such that it took just under 1000 h to create ~17.5 h of transcription data (i.e., 0.75–1 audio hour for each of 20 children). Most of that time was spent on site in the recording locations and required two people working simultaneously: a native speaker steering the transcription and a researcher steering the laptop. As such, it took several years of 4–8-week trips to each site to produce corpora that were adequately sized for publication: the Tseltal data were collected in 2015 and published in 2020 (Casillas, Brown, &

[2] Annotations included: transcription and loose translations of child-produced speech, all other hearable speech, and addressee information for all non-child speech. The researcher was efficient and experienced in the transcription program, ELAN (https://archive.mpi.nl/tla/elan; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006), so software was not a bottleneck.

Levinson, 2020); the Yélî Dnye data were collected in 2016 and published in 2021 (Casillas, Brown, & Levinson, 2021). If you are planning a similar project, consider that our clips were randomly selected, so the average time reported above is lowered by the occasional clips in which children were sleeping and there was nothing to transcribe—if we had exclusively selected from waking periods or interactively-engaged periods, the transcription time estimates would be higher.

Our investment in on-site, collaborative transcription showed its benefits in early 2019, when we were first able to employ one of our trained Tseltal transcribers for year-round, independent remote work. We had just shifted to a more efficient workflow in which he transcribed and gave translations of target-child-produced speech and target-child-directed speech. His files were then handed off to trained U.S. research assistants for fine tuning in precise utterance boundary placement and target child vocal maturity annotations. So, even while we could not visit the community due to COVID-19 pandemic measures, the Tseltal transcriber produced draft transcripts for forty-four new recordings (i.e., more than four times the original ten recordings annotated, and in half the time). However, these changes in transcription workflow came with changes in the major issues to be addressed: It has not been easy to track the different versions of each transcribed sub-clip, and this method is much, much more expensive (costs of living are higher for this transcriber, who lives outside the primary village). That said, we are deeply grateful to the local institutional infrastructure that makes remote employment possible in this case.

Looking back, even if validated and high-quality automated transcription were an option from the very start of this project, it would *still* have been wise to invest significant time and resources into doing manual transcription with local, native speakers. Manual transcription gave us higher-quality and more information-rich data, and the time and personnel investment helped us to build partnerships with villages, community members, and local institutions that are now helping us grow our capacity to more efficiently produce high-quality transcription data. If and when automated transcription tools are ready to take on noisy, egocentric recordings in low-resource languages (more on this below), then the manual transcriptions will also provide gold-standard data.

Researchers who are considering working on an under-represented language should thus plan to do significant manual transcription for at least three reasons: (1) to produce data for direct analysis, (2) to validate any data produced by existing automated annotation tools, and (3) to provide

potential training data for future automated annotation tools. Doing so can present community members with increased opportunities for scientific training and higher or more stable earnings (see also Sarvasy, under review) and can serve as a basis for discussions with community members about what types of documentation and automated tools are most useful locally.

3.1.2 Ex. larger-scale language communities: North American English

When manual transcription is needed in a larger-scale language context (e.g., for annotation types that automated programs cannot yet provide), the primary issues look rather different.

One clear difference comes from the division of labor for transcription. The Tseltal and Yélî Dnye transcriptions were made possible by the training and highly dedicated effort of 2–3 research assistants per language (working hundreds of hours each), typically working in close collaboration with the researcher. In the North American English-speaking context, the transcription work is distributed in much smaller portions over a much larger group of research assistants (RAs). Each RA is trained very rigorously with the hope that they can produce high-quality work with minimal supervision. In our group, RAs tend to work 4–8 h per week during the academic year, which means it often takes months for annotators to become fully trained and to begin contributing high-quality, usable annotations. A colleague of ours reports 80–100 h of training time needed per annotator to ensure high-quality transcription outputs (Montag, personal communication, September 21, 2022) using a common ELAN-based annotation scheme (Soderstrom et al., 2021). Therefore, an RA's tenure in the lab may sometimes elapse before they contribute significant (if any) usable data to a manual annotation project. Even after training is complete, RAs may need significant help from researchers in making moment-to-moment decisions about what to transcribe and how—this requires double personnel time, similar to what we use in the smaller-scale datasets. As such, investing in research assistants tends to have less long-term payoff in the large-scale context. Still, having a larger pool of RAs has enabled us to train some specialists in tasks that are a poor use of time for native speakers in the smaller-scale context: precise speech onset/offset timing, vocal maturity tags for children's vocalizations, etc.

Another clear difference is the cost of transcription projects. While the number of available RAs is much greater in our home lab context, funding to train and employ all of those RAs becomes a significant burden. In the past 2.5 years, our lab has had success in recruiting 20 different annotators,

thanks to support from the University of Chicago (faculty grants, RA summer funding programs), the U.S. National Science Foundation, and the U.S. Federal Work-Study Program. Without these resources, we would have likely avoided building a significant transcription corpus in English because of its prohibitive costs. Then again, the smaller-scale contexts have required much more PI involvement (e.g., via field stays), which puts hard limits on the scope of small-scale corpus growth, even with generous grant funding.

A third difference has to do with the detail and quality of speaker data. The acute nature of the field visits, with transcription as a priority and with RAs who know the recorded families personally, ensures consistently high-quality and confident speaker identification. The English-speaking RAs in our lab, in contrast, lack familiarity with the recorded families and sometimes struggle in distinguishing individual speakers and interpreting some individuals' speech patterns, especially in the noisy, multi-speaker environments that are often captured in our egocentric recordings.

Finally, the greater availability of naturalistic recordings from U.S. English-speaking homes (e.g., via CHILDES, HomeBank, Databrary, etc.) gives us the opportunity to start our projects on English with existing corpora. This saves us immense time in collecting our own recordings from scratch.

4. The promise (and shortfalls) of automated solutions

In today's world, it is easy to take for granted the availability of accurate, automated speech recognition: systems linked through our phones, in our homes, digital workspaces, and even in public appear rather good at capturing and auto-transcribing spoken language. These truly remarkable technological advances in speech processing, along with advances in computer vision technology, give many child language researchers high expectations about what is possible with automated tools (e.g., with the LENA system). The idea of scaling up child language science to "big data" analysis is very appealing, especially considering our field's history of small and convenient participant samples (Doebel & Frank, 2023; Nielsen, Huan, Kärtner, & Legare, 2017; Singh, Cristia, Karasik, Rajendra, & Oakes, 2023). But when it comes to daylong egocentric data, the utility of automated tools is unfortunately highly limited. While there are many excellent examples of work capitalizing on the best and most reliable automated annotations

available for daylong egocentric data (e.g., Bergelson et al., 2024; among many others), researchers' expectations about what automated tools can do beyond these basic annotation types are often unrealistic. In what follows, we give a quick and current overview of what types of annotations are (and are not) reliably available for daylong egocentric audio recordings from natural child language environments.

While significant and important advances have been made in basic speech signal processing (Lavechin, Bousbib, Bredin, Dupoux, & Cristia, 2020; Radford et al., 2023; Xu, Yapanel, & Gray, 2009; Cao et al., 2018), it remains the case that other historically essential data types—most notably, transcriptions—are not typically available from automated tools (but see Lavechin et al., 2023). Lack of available and reliable tools for any linguistic task is further exacerbated for low-resource language communities (for whom there is relatively little gold-standard manual transcription data with which to train automated systems), not to mention sign language communities (for whom speech processing tools are not applicable). Further, even when automated transcription software and other linguistic analysis tools are available, they typically focus on adult-produced (not child-produced) speech.

The most important question for a researcher embarking on an egocentric, naturalistic recording project is: what types of information are relevant for the research questions? If the research project can be completed on the basis of one or more of the following sources of information, existing automated tools can probably get you there:

- **Speech/non-speech segmentation** (distinguishing recording periods that are likely to contain speech from those that are not). Some example questions answerable based on segmentation alone include: How much speech is present in the recording? How much silence is there? How are speech and silence distributed over time? These segmentations are made more useful in combination with the next type of information…
- **Speaker type diarization** (classifying each segment of detected speech into broad speaker types). For child language recordings, relevant speaker types typically include the key/recorded child, a different child, a female adult, or a male adult. Note that these tools do not yet reliably distinguish between individual speakers within each type (e.g., consistent demarcation of two nearby female adults), even if a human could easily tell them apart. Some example questions one can answer based on speaker-type classified speech segments include: How often is the key child talking? What types of speakers are talking near the key child and how often?

When and for how long does the key child converse with other speakers (inferred from timing data)? These classified segments can be further analyzed for superficial aspects of their linguistic content…

- **Linguistic unit count estimation** (quantifying the number of phones, syllables, words, etc.). For segments classified as adult speech, phonetic evidence from each segment can be used to estimate how many units of linguistic information are present.

If you are a LENA (Xu et al., 2009) or ALICE (Räsänen, Seshadri, Lavechin, Cristia, & Casillas, 2021) user, the software can give you estimates about vocalization counts, word counts, turn-taking counts, and more on the basis of the three information types described above.[3]

If you have listened to any egocentric, naturalistic recordings, you may find these automated outcomes very impressive. A combination of ambient noise, child microphone handling, and movement-related rustling often make these recording data difficult to parse, even for trained human listeners. The above annotation types have been sufficient to answer many research questions, and the number of publications using them has been rapidly growing since the introduction of LENA in the early 2000s.[4] Indeed, egocentric, naturalistic recording data sometimes feel synonymous with automated outcome measures. Yet, the answers to many research questions—especially those emerging from lines of inquiry that were developed around transcription data—are largely unanswerable with current automated tools.

If a project requires the researcher to know about the *content* of the produced utterances, manual annotation is almost certainly going to be necessary. Automated tools, at present, are far from being able to offer reliable transcription of adult speech in children's daylong egocentric recordings, let alone transcription of children's speech, or other more detailed linguistic annotations (e.g., morphosyntactic glosses, phonetic transcription, etc.). What if you do not need all of this transcription? One idea is to start with a list of target words and then create a tool that can simply scan the audio for matches to those words ("keyword spotting"). This is a nice idea, but unfortunately, no such tool to our knowledge has so far been applied and validated with egocentric recordings from naturalistic child language environments.

[3] This description leaves out some important details about LENA's software, such as attempting to separate the key child's language-relevant (e.g., babble) from non-language relevant vocalizations (e.g., crying) using date-of-birth information provided by the researcher.

[4] A Google Scholar search requiring the phrases "LENA," "AWC" and "child language" returns 2 results prior to 2008 and 244 results at time of writing (December 2023).

Thinking now specifically about the speech produced by children, there are few options available. For example, the tools available for counting linguistic units (via LENA and ALICE, described above) are only applied and validated for adult speech. LENA's software classifies child vocalizations as speech-related (words, babbling, and communicative sounds like squeals, growls, and raspberries) or non-speech-related ("fixed": crying, screaming, laughing; "vegetative": burping, breathing). When speaker-type classification is correct, accuracy of this vocalization classification is fairly high: speech-related vocalizations are correctly classified as such 75% of the time and non-speech-related ones 84% of the time (Xu et al., 2009). Efforts to further sub-classify speech-related vocalizations have been less successful; the LENA system uses a two-way classification scheme (speech- vs. non-speech-related) but others have attempted to develop a more fine-grained scheme. For example, Al Futaisi, Zhang, Cristia, Warlaumont, and Schuller (2019) made one attempt with two speech-related sub-classes (canonical babble (including words) vs. non-canonical babble) and two non-speech-related sub-classes (cry vs. laugh). While several of their approaches achieve above-baseline performance (and all above-chance performance), the accuracy scores are still far too low to be used directly in analyses of child language development (unweighted average recall (accuracy) scores of up to 50.1% in a test set where chance is 25%).

Other examples of common measures that would be timesaving if automated tools were available include addressee classification (e.g., child-directed vs. adult-directed speech) and estimates of interactional exchange (e.g., turn-taking rates, conversation onsets/offsets). Both annotation types rely on content, perhaps to a surprising degree. The difference between infant- and adult-directed English appears more subtle in daylong recordings than in shorter, more controlled ones (MacDonald, Räsänen, Casillas, & Warlaumont, 2020). One initial effort to build addressee classifiers failed (Schuller et al., 2017), though further tool development is underway (e.g., Bang, Kachergis, Weisleder, & Marchman, 2023). LENA provides estimates for conversational blocks and the rate of speaker switches (conversational turn count, "CTC"). However, the CTC measure has worse accuracy than the other core LENA measures (child vocalization count, "CVC;" adult word count, "AWC") and has an error pattern affected by both the age of the recorded child (Ferjan Ramírez, Hippe, & Kuhl, 2021) and the number of nearby child and adult speakers (Ferjan Ramirez, Hippe, Braverman, Weiss, & Kuhl, 2023). Thus, studies of CTC across children of different ages or with systematically different household

compositions face serious issues of underlying heteroskedasticity in their analyses. We have used a custom set of R scripts ("chattr"; https://github.com/marisacasillas/chattr-basic) to calculate CTC-like measures adapted to include more principles from Conversation Analysis and to be usable on more types of tabular speech data. The results from chattr correlate with LENA estimates (Casillas & Scaff, 2021) but are likely subject to the same age- and interaction-related errors as LENA. Of course, there are many more research questions that can be asked of daylong egocentric data not addressed here because there simply do not exist automated annotation tools (even preliminary ones) to investigate them at present (e.g., those relating to daylong video data, which we discuss below).

5. Exciting future directions

Now, nearly 20 years after the first LENA system was piloted, egocentric daylong recordings have become a standard approach to studying children's language environments. Scores of researchers rely on LENA, but many others use their own systems, including alternative audio devices (e.g., USB "spy" recorders; Olympus/Sony handheld recorders), photo-linked audio (e.g., pairing an audio recorder with a camera device like the Narrative Clip), and egocentric video (e.g., head- or chest-worn, typically for shorter recordings). LENA-users and non-LENA-users alike have expressed a keen interest in alternative open-source applications, such as ALICE and chattr. Many have also taken the leap into manual transcription and annotation, opening up a whole other line of logistical questions, such as how and how much to sample from long recordings (Casillas, Bergelson, et al., 2017; Cychosz, Villanueva, & Weisleder, 2021; Marasli & Montag, 2023; Micheletti et al., 2020).

Work on daylong egocentric recordings is bringing us rapidly closer to characterizing the typical quantities of different sources of input in children's language environments, and across diverse contexts—a basic, but until now, unanswerable question informing realistic constraints on child language learning. As we build up more transcription data, it will also bring us closer to characterizing the content of input sources, making more solid connections to what we observe when measuring infants' implicit language knowledge in the lab.

Importantly, with new techniques also come new insights—in the remainder of this chapter, we will discuss two clear areas in which long-form, egocentric recordings may help us ask (and answer) questions about language development in new ways.

5.1 Getting out of the house

The vast majority of what we know about children's language experiences is based on recordings made in lab settings or in children's own homes. For logistical and ethical reasons (especially privacy laws and concerns), recordings made in daycares, schools, and other settings outside of more controlled lab and home contexts are much rarer. For many children around the world, these "other" (i.e., non-lab and non-home) settings make up an enormous proportion of their daily language experiences.

In the U.S., approximately 60% of children up to age 5 reportedly receive (as of 2019) some type of non-parental care (daycare, preschool, care in a private home provided by a non-relative, care provided by a relative) at least once per week (Cui & Natzke, 2021). On Rossel Island, the home of our Yélî Dnye corpus, very young children are cared for by a wide network of nearby family members and neighbors, including older children—by age two children typically join large independent play groups for much of the day (Brown & Casillas, 2020; Brown, 2011, 2014). In the Tseltal community where we work, infants up to age 18 months are carried on their mother's back for large stretches of the day, in part due to the village's location on a mountainside—when children and their mothers leave home, they must traverse challenging terrain that is unfit for inexperienced walkers (Brown, 2011, 2014). But later in childhood, Tseltal children tend to participate in small social groups within and around their household grounds (Brown, 2011, 2014). Thus, the physical and social landscapes of children's early language environments are highly variable within and across communities—where, with whom, and exactly how children interact with the world varies widely, and for most children, extends far beyond interaction with their primary adult caregiver(s) and within bounds of their home. There is a great deal to learn from these experiences that are currently under-documented.

One clear topic that merits further investigation is children's language experiences in daycares and preschools, where the number of interactants present and the types of activities children engage in are likely quite different from their home environments. Some studies have quantified features of language input from adults in out-of-home care settings (e.g., Larson, Barrett, & McConnell, 2020; Soderstrom, Grauer, Dufault, & McDivitt, 2018), but much less work has considered the frequency and features of talk from peers—a key aspect of children's language environments that also varies considerably across communities (e.g., Bunce et al., in press). Another setting that intrigues us is children's ***outdoor*** language experiences. Children's outdoor

language input is likely to vary widely in quantity and type for children growing up in different geographical locations and sociocultural contexts, and we can imagine how outdoor time differently influences the types and content of children's early verbal interactions from community to community. Our group's indoor versus outdoor pilot egocentric audio data suggest comparable transcription reliability and background noise levels, indicating the feasibility of using existing recording technology to record children's outdoor language experiences in future research.

5.2 Characterizing multimodal input

While the prominence of daylong *audio* recording technology has privileged investigations of children's speech environments and their own vocal activity (i.e., almost exclusively focusing on the auditory modality), we are looking forward to the increasing use of daylong *video* recording technology. After all, we know that children learn spoken and sign languages from multimodal input. Methodological barriers have so far limited our ability to investigate multimodal features of children's home language input on a daylong scale; however, improvements in the battery life and recording capacity of cameras and video recorders, along with other wearable technologies (e.g., Salo et al., 2022; Wass, Smith, Clackson, & Mirza, 2021), make examining multiple modalities across longer timescales much more feasible. Among the other types of perceptual input (sight, touch, taste, smell), the most accessible input to add to daylong recording studies is sight—by making a video recording instead of an audio recording.

Shorter-form home video recordings have provided great insight into children's experiences beyond the auditory modality. The bulk of this research has focused on children's visual input—how often different objects are in view (Clerkin & Smith, 2022; Long, Kachergis, Bhatt, & Frank, 2021), how often they are the focus of joint attention (Bergelson et al., 2019; Schroer, Peters, Yarbrough, & Yu, 2022), or how often they are interacted with by children and their caregivers (Suarez-Rivera, Linn, & Tamis-LeMonda, 2022; Swirbul, Herzberg, & Tamis-LeMonda, 2022). Some studies have taken advantage of information provided in the visual signal to characterize other aspects of children's home experiences, including their physical proximity to and touch or gesture from adult caregivers (Abu-Zhaya, Seidl, & Cristia, 2017; Kosie & Lew-Williams, 2023; Suarez-Rivera, Pinheiro-Mehta, & Tamis-LeMonda, 2023), along with their physical location in space throughout the home (Custode & Tamis-LeMonda, 2020; Roy, Frank, DeCamp, Miller, & Roy, 2015).

In recent years, we have seen the development of many new systems for capturing at-home egocentric video data, including head-worn cameras, such as BabyView (Long et al., 2023) and EgoActive (Geangu et al., 2023), as well as advancements in head-mounted eye-trackers (Schroer et al., 2022). Personal security cameras (similar to police cameras) open up another off-the-shelf option. Our group has piloted this technology, outfitting U.S. English-speaking 1–5-year-olds with a custom shirt and a chest-worn camera that we have fitted with a fisheye lens (Fig. 1A). One potential concern with this off-the-shelf video technology is the quality of the audio signal. Encouragingly, we have found manual transcription outputs from the raw audio data to be reliable, with background noise interference comparable to other audio recorders.

On the path toward daylong video recordings is daylong photo-linked audio recordings (i.e., using two devices to continuously record audio and intermittently capture static photos). We used this method for the 2015–2016 Tseltal and Yélî Dnye HomeBank corpora mentioned above (Casillas, Brown, et al., 2017). The camera used in creating those corpora—a Narrative Clip 1 (now discontinued) with an attached Photojojo Super Fisheye lens (Fig. 1B)—provided a static 180-degree view of the environment from the child's perspective with snapshots taken every 15–30 s for approximately 8–9 continuous hours on one battery charge. This rich dataset has allowed our group to ask questions about the number of potential interactants (child and adult) nearby each recorded child throughout the day, as well as the frequency with which children engage in object handling episodes and the different types of objects most commonly available to children in these communities of study (Casey et al., in prep). With an annotation program specifically designed for efficient tagging of information in photo streams (https://github.com/marisacasillas/imco), a trained annotator can code at a rate of 5–10 average seconds per photo, depending on the specific task (e.g., identifying number of people present vs. identifying the target of an object handling bout). Generating a full dataset of > 100k manually-annotated photos has been a laborious but worthwhile investment and has positioned our group to be able to conduct comparative analyses of children's object handling behaviors (Casey et al., 2022) and, in ongoing work, to use object-centric input estimates to predict age-of-acquisition for object nouns and real-time word recognition in gaze experiments. However, photo streams are not sufficient to calculate accurate temporal patterns in manual behaviors, to analyze brief visual or manual inputs (e.g., gestures), or to precisely time align children's multimodal input (e.g., to determine the exact co-occurrence of handling an object and hearing its label).

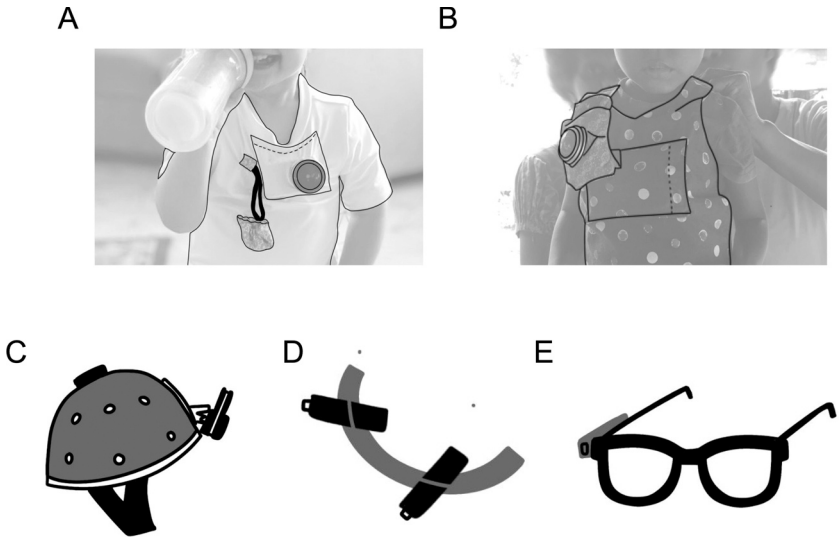

Fig. 1 Examples of different egocentric video-recording devices. (A) Chest-worn video security camera (Boblov 64 GB N9 Mini with attached fisheye lens) in a custom shirt and attached lens cover, (B) chest-worn photo camera (Narrative Clip 1 with attached fisheye lens) and audio recorder (Olympus WS-832) in a custom vest, (C) BabyView head-mounted camera (Long et al., 2023), (D) two Looxcie cameras attached to headband (Bergelson et al., 2019), (E) wearable eye-tracker mounted to glasses (Schroer et al., 2022).

Each of these egocentric recording techniques—daylong audio, photo, or video—has its own strengths and weaknesses, differing on many dimensions, including cost, placement (i.e., head-worn or chest-worn), afforded comfort and mobility for the recorded child, length of recording, participation rate,[5] etc.

Our best advice: Consider the relevant dimensions for your particular project and recording context, with a primary goal of maximizing data re-use potential. Opt for video over photos when feasible (photos can always be subsampled later if continuous data is not necessary to answer a particular question). For camera recordings, *always* add a fisheye lens to the device if it is not already equipped with one (Fig. 2). Minimize data loss by limiting children's and caregivers' ability to accidentally turn on/off the

[5] Anecdotally, our group has found it relatively more difficult to recruit for daylong video recording studies. In an ongoing study with 1–5-year-olds from English-speaking homes in the Chicagoland area, we have only achieved a 14% participation rate for daylong video recordings when parents are offered this opportunity following a successful in-person lab session with their child.

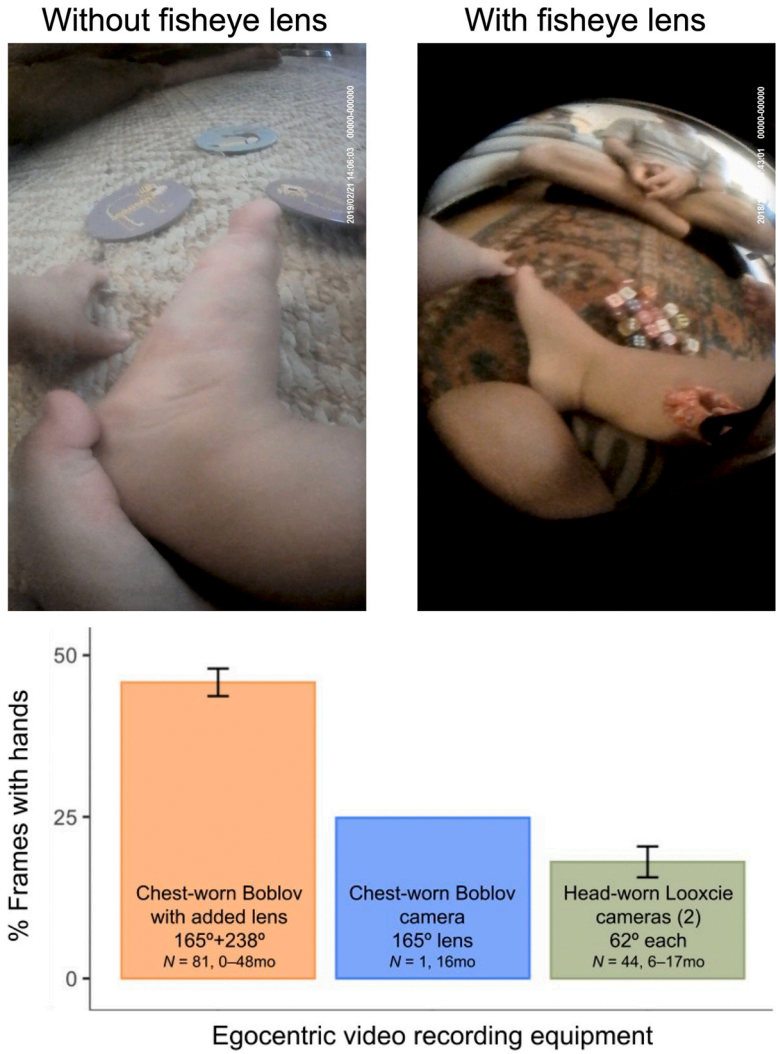

Fig. 2 Example chest-worn security camera (Boblov 64 GB N9 Mini) output without vs. with a supplementary fisheye lens. This off-the-shelf device has a 165-degree lens built in (top left) to which we add a 238-degree mobile phone lens (top right; Efocakiox 7.5 mm Super Fisheye lens). Both images feature object-centric interaction with one adult caregiver and similar recorded child age and body position. More so than camera position (e.g., head vs. chest), we find that adding a very wide-view lens is most effective in revealing children's ongoing manual and social activities; in the plot, we show percentage of frames capturing children's own hands across three egocentric video recording setups. *Looxcie data courtesy of Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. Developmental Science, 22(1), e12715. https://doi.org/10.1111/desc.12715; available on Databrary under volumes named "SEEDLingS."*

recording device during the day; encourage the use of a camera privacy cover and/or placement of the device somewhere outside of earshot if families need a brief break from recording.

Equipped with daylong photo/video recording technology, in addition to audio, we can uncover novel and underexplored features of children's natural, *multimodal* language environments.

6. Conclusion

Daylong egocentric recordings give us the opportunity to observe the world as children do, bringing us closer to understanding how, when, and what they learn. Creating these rich, naturalistic datasets comes with immense challenges—challenges that differ across smaller- and larger-scale language communities. Automated tools give us a first foot in the door to analyzing features of the recorded speech signal, but—especially for low-resource languages and signed languages—fall short of the mark. For now, we must continue to rely on manual transcription for research questions that require knowledge of the *content* of children's language environments. A community-oriented approach, valuing the sharing and augmenting of datasets and open-source tools, is a powerful way to overcome these challenges in the long run. We hope to see future work that goes beyond tool improvement and additional annotations. We would especially like to see work that explores less well-charted territory for daylong recordings: outdoor language use and daylong multimodality.

Acknowledgments

We gratefully acknowledge the Tseltal, Yélî, and North American participants and the many transcribers and annotators who have contributed to this work over the years (Alexander Klerman, Alexander Stern, Anapaula Silva Mandujano, Antun Gusman Osil, Ariana Maisonet, Carla Escalante, Cielke Hendriks, Daphne Jansen, Elizabeth Mickiewicz, Emily Chan, Erica Hsieh, Ghaalyu Yidika, Humbertina Gómez Pérez, Ine Alvarez van Tussenbroek, Isabella di Giovanni, Jenny Bo, Jordyn Martin, Juan Méndez Girón, Kimberly Shorter, Maartje Weenink, Mara Duquette, Mary Elliott, Maryam Mohammed-Norgan, Mia Rimmer, Ndapw:éé Yidika, Rebeca Gúzman López, Ruby Swensen, Sarah Kelso, Sarah Sommer, Solana Guillu, Subin Kim, Taakêmê Ńamono, Will Fisher, Y:aaw:aa Pikuwa, Yuchen Jin). Their experiences are reflected in this chapter. We are grateful, too, to the local research institutions who make the international projects possible (especially CIESAS Sureste and the Papua New Guinea National Research Institute) as well as our collaborators in the ACLEW project (Melanie Soderstrom, Celia Rosemberg, John Bunce, Okko Räsänen, Marvin Lavechin, Elika Bergelson, Alex Cristia). We give special thanks to Benjamin Morris, Marvin Lavechin, and Solana Guillu for their feedback on an early version of this chapter; any remaining errors are our own. This chapter was supported by an NSF CAREER grant to MC (BCS 2238609).

References

Abu-Zhaya, R., Seidl, A., & Cristia, A. (2017). Multimodal infant-directed communication: How caregivers combine tactile and linguistic cues. *Journal of Child Language, 44*(5), 1088–1116. https://doi.org/10.1017/S0305000916000416.

Adolph, K. E., Cole, W. G., Komati, M., Garciaguirre, J. S., Badaly, D., Lingeman, J. M., ... Sotsky, R. B. (2012). How do you learn to walk? Thousands of steps and dozens of falls per day. *Psychological Science, 23*(11), 1387–1394. https://doi.org/10.1177/0956797612446346.

Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson, P., & Millman, D. (2012). Toward open behavioral science. *Psychological Inquiry, 23*(3), 244–247. https://doi.org/10.1080/1047840X.2012.705133.

Al Futaisi, N. D., Zhang, Z., Cristia, A., Warlaumont, A. S., & Schuller, B. W. (2019). VCMNet: Weakly supervised learning for automatic infant vocalisation maturity analysis. *Proceedings of the International Conference on Multimodal Interaction,* 205–209. https://doi.org/10.1145/3340555.3353751.

Bang, J. Y., Kachergis, G., Weisleder, A., & Marchman, V. A. (2023). Evaluating the feasibility of an automated classifier for target-child-directed speech from LENA recordings. *Language Development Research, 3*(1), 211–248. https://doi.org/10.34842/xmrq-er43.

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science, 22*(1), e12715. https://doi.org/10.1111/desc.12715.

Bergelson, E., Soderstrom, M., Schwarz, I., Rowland, C. F., Ramirez-Esparza, N., Hamrick, L. R., ... Cristia, A. (2024). Everyday language input and production in 1001 children from 6 continents. *Proceedings of the National Academy of Sciences, 120*(52), e2300671120. https://doi.org/10.1073/pnas.2300671120.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

Brown, P. (2011). *The cultural organization of attention. The Handbook of Language Socialization.* John Wiley & Sons, Ltd, 29–55. https://doi.org/10.1002/9781444342901.ch2.

Brown, P. (2014). The interactional context of language learning in Tzeltal. In I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (Eds.). *Language in interaction: Studies in Honor of Eve V. Clark* (pp. 51–82). John Benjamins Publishing Company. https://www.torrossa.com/en/resources/an/5000623.

Brown, P., & Casillas, M. (2020). Child rearing through social interaction on Rossel Island, PNG. *PsyArXiv*. https://doi.org/10.31234/osf.io/5rvky.

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein A., Alam, F., ... Casillas, M. (in press). A cross-cultural examination of young children's everyday language experiences, Journal of Child Language.

Cao, X.-N., Dakhlia, C., Del Carmen, P., Jaouani, M.-A., Ould-Arbi, M., & Dupoux, E. (2018, May). Baby Cloud, a technological platform for parents and researchers. In *LREC 2018 - 11th Edition of the Language Resources and Evaluation Conference*. https://hal.science/hal-01948107.

Casey, K., Elliott, M., Mickiewicz, E., Silva Mandujano, A., Shorter, K., Duquette, M., ... Casillas, M. (2022). Sticks, leaves, buckets, and bowls: Distributional patterns of children's at-home object handling in two subsistence societies. *Proceedings of the Annual Meeting of the Cognitive Science Society, 44*(44), https://escholarship.org/uc/item/6wx2x30s.

Casey, K., Elliott, M., Mickiewicz, E., Bergelson, E., & Casillas, M. (in prep). Daylong patterns of object-centric interactions in two subsistence societies.

Casillas, M. (2023). Learning language in vivo. *Child Development Perspectives, 17*(1), 10–17. https://doi.org/10.1111/cdep.12469.

Casillas, M., Bergelson, E., Warlaumont, A. S., Cristia, A., Soderstrom, M., VanDam, M., & Sloetjes, H. (2017). A new workflow for semi-automatized annotations: Tests with long-form naturalistic recordings of children's language environments. *Interspeech 2017,* 2098–2102. https://doi.org/10.21437/Interspeech.2017-1418.

Casillas, M., Brown, P., & Levinson, S.C. (2017). *Casillas HomeBank Corpus* [dataset]. https://doi.org/10.21415/T51X12.

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development, 91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349.

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language, 48*(4), 792–814. https://doi.org/10.1017/S0305000920000549.

Casillas, M., & Scaff, C. (2021). Analyzing contingent interactions in R with `chattr. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*(43), https://escholarship.org/uc/item/4rr848x0.

Clerkin, E. M., & Smith, L. B. (2022). Real-world statistics at two timescales and a mechanism for infant learning of object names. *Proceedings of the National Academy of Sciences, 119*(18), e2123239119. https://doi.org/10.1073/pnas.2123239119.

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., ... Bergelson, E. (2021). A thorough evaluation of the Language Environment Analysis (LENA) system. *Behavior Research Methods, 53*(2), 467–486. https://doi.org/10.3758/s13428-020-01393-5.

Cui, J., & Natzke, L. (2021). Early Childhood Program Participation: 2019. In *National Center for Education Statistics*. U. S. Department of Education, National Center for Education Statistics. https://nces.ed.gov/pubs2020/2020075REV.pdf.

Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy, 25*(6), 809–826. https://doi.org/10.1111/infa.12361.

Cychosz, M., & Cristia, A. (2022). Using big data from long-form recordings to study development and optimize societal impact. R. O. Gilmore, & J. J. Lockman (Eds.). *Advances in Child Development and Behavior, Vol. 62*, 1–36. https://doi.org/10.1016/bs.acdb.2021.12.001.

Cychosz, M., Villanueva, A., & Weisleder, A. (2021). Efficient estimation of children's language exposure in two bilingual communities. *Journal of Speech, Language, and Hearing Research, 64*(10), 3843–3866. https://doi.org/10.1044/2021_JSLHR-20-00755.

Doebel, S., & Frank, M. C. (2023). Broadening convenience samples to advance theoretical progress and avoid bias in developmental science. *Journal of Cognition and Development,* 1–12. https://doi.org/10.1080/15248372.2023.2270055.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition, 152*, 101–107. https://doi.org/10.1016/j.cognition.2016.03.005.

Ferjan Ramirez, N., Hippe, D. S., Braverman, A., Weiss, Y., & Kuhl, P. K. (2023). A comparison of automatic and manual measures of turn-taking in monolingual and bilingual contexts. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02127-z.

Ferjan Ramírez, N., Hippe, D. S., & Kuhl, P. K. (2021). Comparing automatic and manual measures of parent–infant conversational turns: A word of caution. *Child Development, 92*(2), 672–681. https://doi.org/10.1111/cdev.13495.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language, 44*(3), 677–694. https://doi.org/10.1017/S0305000916000209.

García Coll, C., Crnic, K., Lamberty, G., Wasik, B. H., Jenkins, R., García, H. V., & McAdoo, H. P. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development, 67*, 1891–1914. https://doi.org/10.1111/j.1467-8624.1996.tb01834.x.

Gautheron, L., Lavechin, M., Riad, R., Scaff, C., & Cristia, A. (2020). Longform recordings: Opportunities and challenges. In T. Poibeau, Y. Parmentier, & E. Schang (Eds.). *LIFT 2020—2èmes journées scientifiques du Groupement de Recherche "Linguistique informatique, formelle et de terrain"* (pp. 64–71) CNRS. https://hal.science/hal-03047153.

Geangu, E., Smith, W. A. P., Mason, H. T., Martinez-Cedillo, A. P., Hunter, D., Knight, M. I., ... Muller, B. R. (2023). EgoActive: Integrated wireless wearable sensors for capturing infant egocentric auditory–visual statistics and autonomic nervous system function 'in the wild. *Sensors, 23*(18), https://doi.org/10.3390/s23187930.

Gennetian, L. A., Tamis-LeMonda, C. S., & Frank, M. C. (2020). Advancing transparency and openness in child development research: Opportunities. *Child Development Perspectives, 14*(1), 3–8. https://doi.org/10.1111/cdep.12356.

Gilmore, R. O. (2022). Show your work: Tools for open developmental science. R. O. Gilmore, & J. J. Lockman (Eds.). *Advances in child development and behavior, Vol. 62*, 37–59. https://doi.org/10.1016/bs.acdb.2022.01.001.

Gilmore, R. O., Adolph, K. E., & Millman, D. S. (2016). Curating identifiable data for sharing: The databrary project. *2016 New York Scientific Data Summit (NYSDS)*, 1–6. https://doi.org/10.1109/NYSDS.2016.7747817.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language, 37*(2), 229–261. https://doi.org/10.1017/S0305000909990432.

Kosie, J. E., & Lew-Williams, C. (2022). Open science considerations for descriptive research in developmental science. *Infant and Child Development*, e2377. https://doi.org/10.1002/icd.2377.

Kosie, J. E., & Lew-Williams, C. (2023). Infant-directed communication: Examining the many dimensions of everyday caregiver-infant interactions. *Developmental Science*, e13515. https://doi.org/10.1111/desc.13515.

Larson, A. L., Barrett, T. S., & McConnell, S. R. (2020). Exploring early childhood language environments: A comparison of language use, exposure, and interactions in the home and childcare settings. *Language, Speech, and Hearing Services in Schools, 51*(3), 706–719. https://doi.org/10.1044/2019_LSHSS-19-00066.

Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (2020). An open-source voice type classifier for child-centered daylong recordings. *Proceedings of Interspeech 2020,* 3072–3076. https://doi.org/10.21437/Interspeech.2020-1690.

Lavechin, M., Métais, M., Titeux, H., Boissonnet, A., Copet, J., Rivière, M., ... Bredin, H. (2023). Brouhaha: Multi-task training for voice activity detection, speech-to-noise ratio, and C50 room acoustics estimation. *Automatic Speech Recognition and Understanding Workshop*.

Long, B., Goodin, S., Kachergis, G., Marchman, V. A., Radwan, S. F., Sparks, R. Z., ... Frank, M. C. (2023). The BabyView camera: Designing a new head-mounted camera to capture children's early social and visual environments. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02206-1.

Long, B., Kachergis, G., Bhatt, N. S., & Frank, M. C. (2021). Characterizing the object categories two children see and interact with in a dense dataset of naturalistic visual experience. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*(43), https://escholarship.org/uc/item/5t30m4qz.

Long, H. L., Ramsay, G., Griebel, U., Bene, E. R., Bowman, D. D., Burkhardt-Reed, M. M., & Oller, D. K. (2022). Perspectives on the origin of language: Infants vocalize most during independent vocal play but produce their most speech-like vocalizations during turn taking. *PLoS One, 17*(12), e0279395. https://doi.org/10.1371/journal.pone.0279395.

MacDonald, K., Räsänen, O., Casillas, M., & Warlaumont, A. S. (2020). Measuring prosodic predictability in children's home language environments. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 695–701. https://cognitivesciencesociety.org/cogsci20/papers/0126/index.html.

MacWhinney, B. (2000). (3rd ed.). *The CHILDES project: Tools for analyzing talk: Transcription format and programs*. Lawrence Erlbaum Associates Publishers, 366 (xi).

MacWhinney, B. (2019). *CHAT Manual*. https://doi.org/10.21415/3MHN-0Z89.

Marasli, Z., & Montag, J. L. (2023). Optimizing random sampling of daylong audio. *Proceedings of the Annual Meeting of the Cognitive Science Society, 45*(45). https://escholarship.org/uc/item/0478r9pb.

Max Planck Institute for Psycholinguistics, The Language Archive. (2023). *ELAN (Version 6.7)* [Computer software]. https://archive.mpi.nl/tla/elan.

McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development, 88*(1), 156–166. https://doi.org/10.1111/cdev.12671.

Micheletti, M., de Barbaro, K., Fellows, M. D., Hixon, J. G., Slatcher, R. B., & Pennebaker, J. W. (2020). Optimal sampling strategies for characterizing behavior and affect from ambulatory audio recordings. *Journal of Family Psychology, 34*(8), 980–990. https://doi.org/10.1037/fam0000654.

Montag, J. L. (2020). New insights from daylong audio transcripts of children's language environments. *Proceedings of the Annual Meeting of the Cognitive Science Society, 42*(42).

Nelson, K. (1989). *Narratives from the crib*. Cambridge, MA: Harvard University Press.

Nielsen, M., Huan, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology, 162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017.

Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal patterns in infants with Autism Spectrum Disorder: Canonical babbling status and vocalization frequency. *Journal of Autism and Developmental Disorders, 44*(10), 2413–2428. https://doi.org/10.1007/s10803-014-2047-4.

Radford, A., Kim, J. W., Xu, T., Brockman, G., Mcleavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. *Proceedings of the 40th International Conference on Machine Learning,* 28492–28518. https://proceedings.mlr.press/v202/radford23a.html.

Räsänen, O., Seshadri, S., Lavechin, M., Cristia, A., & Casillas, M. (2021). ALICE: An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods, 53*(2), 818–835. https://doi.org/10.3758/s13428-020-01460-x.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences, 112*(41), 12663–12668. https://doi.org/10.1073/pnas.1419773112.

Salo, V. C., Pannuto, P., Hedgecock, W., Biri, A., Russo, D. A., Piersiak, H. A., & Humphreys, K. L. (2022). Measuring naturalistic proximity as a window into caregiver–child interaction patterns. *Behavior Research Methods, 54*(4), 1580–1594. https://doi.org/10.3758/s13428-021-01681-8.

Sarvasy, H. (under review). Ethical budgets in psycholinguistic (and other) fieldwork.

Schroer, S. E., Peters, R. E., Yarbrough, A., & Yu, C. (2022). Visual attention and language exposure during everyday activities: An at-home study of early word learning using wearable eye trackers. *Proceedings of the Annual Meeting of the Cognitive Science Society, 44*(44). https://escholarship.org/uc/item/10g0t3m6.

Schuller, B., Steidl, S., Batliner, A., Bergelson, E., Krajewski, J., Janott, C., … Zafeiriou, S. (2017). The INTERSPEECH 2017 computational paralinguistics challenge. *Addressee, Cold & Snoring,* 3442–3446. https://doi.org/10.21437/Interspeech.2017-43.

Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalized science of infant development. *Infancy, 28*(4), 708–737. https://doi.org/10.1111/infa.12545.

Soderstrom, M., Casillas, M., Bergelson, E., Rosemberg, C., Alam, F., Warlaumont, A. S., & Bunce, J. (2021). Developing a cross-cultural annotation system and metacorpus for studying infants' real world language experience. *Collabra: Psychology,* 7(1), 23445. https://doi.org/10.1525/collabra.23445.

Soderstrom, M., Grauer, E., Dufault, B., & McDivitt, K. (2018). Influences of number of adults and adult:child ratios on the quantity of adult language input across childcare settings. *First Language, 38*(6), 563–581. https://doi.org/10.1177/0142723718785013.

Spencer, M. B. (2007). Phenomenology and ecological systems theory: Development of diverse groups. In W. Damon, & R. M. Lerner (Eds.). *Handbook of Child Psychology*. https://doi.org/10.1002/9780470147658.chpsy0115.

Suarez-Rivera, C., Linn, E., & Tamis-LeMonda, C. S. (2022). From play to language: Infants' actions on objects cascade to word learning. *Language Learning, 72*(4), 1092–1127. https://doi.org/10.1111/lang.12512.

Suarez-Rivera, C., Pinheiro-Mehta, N., & Tamis-LeMonda, C. S. (2023). Within arms reach: Physical proximity shapes mother-infant language exchanges in real-time. *Developmental Cognitive Neuroscience, 64*, 101298. https://doi.org/10.1016/j.dcn.2023.101298.

Swirbul, M. S., Herzberg, O., & Tamis-LeMonda, C. S. (2022). Object play in the everyday home environment generates rich opportunities for infant learning. *Infant Behavior and Development, 67*, 101712. https://doi.org/10.1016/j.infbeh.2022.101712.

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech and Language, 37*(2), 128–142. https://doi.org/10.1055/s-0036-1580745.

Wass, S. V., Smith, C. G., Clackson, K., & Mirza, F. U. (2021). In infancy, it's the extremes of arousal that are 'sticky': Naturalistic data challenge purely homeostatic approaches to studying self-regulation. *Developmental Science, 24*(3), e13059. https://doi.org/10.1111/desc.13059.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: A professional framework for multimodality research,* 1556–1559. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_60436.

Xu, D., Yapanel, U., & Gray, S. (2009). *Reliability of the LENA Language Environment Analysis System in young children's natural home environment*. Lena Foundation.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition, 125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016.

Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences, 118*(52), e2107019118. https://doi.org/10.1073/pnas.2107019118.

Zettersten, M., Yurovsky, D., Xu, T. L., Uner, S., Tsui, A. S. M., Schneider, R. M., ... Frank, M. C. (2023). Peekbank: An open, large-scale repository for developmental eye-tracking data of children's word recognition. *Behavior Research Methods, 55*(5), 2485–2500. https://doi.org/10.3758/s13428-022-01906-4.