INTRODUCTION

How do children learn language? In the last half-century, data, methods, and theory attempting to answer this question have grown significantly (Slobin, 2014). Developmental psycholinguistics has arrived at a common, three-piece framework for conceptualizing the core puzzle: input, processor, and output (e.g., Kachergis et al., 2022). In one piece of this framework, children are exposed to language patterns in their ambient environment. In another piece, this linguistic input is processed via a variety of cognitive mechanisms supporting learning. The third and final piece recapitulates input patterns in the child’s own linguistic behaviors (i.e., the output).

Undergirding each piece of this framework are ways of thinking that are taken for granted, but that reflect disciplinary and historical artifacts relevant to psychological inquiry. For example, hypothesis-driven design purposefully constrains the knowledge we create. We strip down and operationalize complex phenomena to reliably quantify, justifiably compare, and adequately control their presence in our datasets. These disciplinary traditions reflect both our field’s origins and where the field is headed, and have resulted in an impressive slate of findings about how language development proceeds in some contexts. We have forged an epistemic path that, by its very trajectory, implies sensible ways forward.

Among these ways are approaches to further illuminate the nature of the linguistic input and processor, including a continued focus on child-directed language from adults, automated measures of language behavior, and (relatedly) indices of linguistic behavior that scale to whole days or longer. As I will argue, these approaches have been fruitful in recent years but present a very...
limited view of the learner and therefore limit the depth of the theories that arise from them. Understanding the individual phenomena that add up to the wonder of childhood language learning requires a complementary approach. In this article, I move away from the beaten path toward other possible disciplinary traditions to identify shortcomings among these ways forward that may otherwise be considered sensible.

I propose to explicitly engage with context rather than attempting to average it away. I argue that analytical approaches to understanding linguistic input and knowledge need to center around learners' interactional and cultural contexts to make meaningful connections between mainstream theoretical frameworks and families' everyday experiences. In this article, I review how we have typically measured input, sketch an alternative approach—the language in vivo approach—and then close with what we will gain from engaging with context.

Past and current approaches to linguistic input

Now, I address three facets of the current approach to analyzing input: (1) which sources of input are included, (2) what are captured in recordings, and (3) how input is measured. All three need further consideration in the language in vivo approach.

Sources of input

There is general consensus that child-directed language from adults (CDLA) is the linguistic input par excellence. CDLA is associated with earlier and cumulatively larger gains in vocabulary and speed of word recognition, and has implications for aspects of syntactic development (e.g., Bates & Goodman, 1999; Frank et al., 2021). That said, other sources of input (e.g., observable talk between others) likely contribute significantly beyond vocabulary (e.g., Benigno et al., 2007; Foushee et al., 2021; Oshima-Takane, 1988; Schick et al., 2022).

Child-directed language itself varies immensely between communities in form and function. As linguistic anthropologists have warned, in many communities, language directed to young children is infrequent, involves multiple parties, or focuses more on social practices than on referential communication. The mere presence of CDLA in and of itself also does not guarantee the scaffolding features presumed to make it helpful in the first place (e.g., McClay et al., 2022). For example, the pitch variation prototypically associated with CDLA is especially salient in North American English, compared both to other English varieties and to other languages (including Lebanese Arabic, K'iche, and Tsetsal; Farran et al., 2016; Pye, 1986; Soderstrom et al., 2021). Even within North American English contexts, CDL use is moderated by the age, gender, and affect of its producer (e.g., Bergelson, Casillas, et al., 2019; Kitamura et al., 2001).

Singular emphasis on CDL misses a critical insight: How others talk to children provides essential information about their expected roles in society, including how they are expected to use language. For example, Euro-American caregivers tend to linguistically interpret and respond to infants' expressions of negative and positive inner experience. Such behavior positions infants as individuals with mature psychological states and creates a foundation for children to continue sharing their inner experiences with others (Gaskins, 2006). These caregivers use CDLA to socially center children and pull their inner experiences outward. In contrast, Yucatec Mayan children are encouraged to be more conservative in verbally expressing inner experiences. A lack of CDLA in response to infant discomfort is an important socialization tool; caregivers rely on bodily contact to swiftly, effectively, and (often) nonverbally attend to infants' negative inner experiences (Gaskins, 2006). To chart the contours of the learner's linguistic environment, we must go beyond a focus on CDLA and engage with a variety of sources of input (e.g., different people who interact with the child), diverse interaction types (e.g., dealing with distress, social routines), and multiple modalities (see, e.g., Abu-Zhaya et al., 2017; Capirci et al., 2022; West et al., 2022).

Capturing language in recordings

Methods for recording children's language environments have changed significantly over the last half century. The foundation of our field was built on careful analysis of naturalistic interactions, recorded for relatively short periods at the child's home or in a laboratory playroom, and then painstakingly annotated before analysis. Then, the Language ENvironment Analysis (LENA) system (Greenwood et al., 2011) launched the field into the domain of big-data research with its 16-h audio-recording capacity and software for automated acoustic analysis. While LENA has been an enormous boon to the field, the associated cost of going big is a loss of depth. Specifically, the automated system produces very limited information about the language used and the data are recorded only in audio (e.g., sign language, gaze, and gesture cannot be captured). This tradeoff between scale and depth should make researchers wary of how feasibility and convenience versus theoretical ideals factor into study design.

Some heroic efforts to annotate daylong data in more detail have demonstrated both the richness of what could be studied if it was manually annotated (e.g., Bergelson, Amatuni, et al., 2019; Montag, 2020) and the potential drawbacks of relying exclusively on automated output (e.g., Ferjan Ramirez et al., 2021). Encouragingly, protocols and guides for manually annotating daylong
data have started to emerge, some with an eye toward maximizing efficiency (Casillas, Bergelson, et al., 2017; Cychosz et al., 2021; Mendoza & Fausey, 2021). Researchers choosing this path can expect large investments of time and money (see Figure 1). However, the payoff is worthwhile: Languages, embedded in their social and material contexts, offer an inspiring array of interrelated phenomena whose learning processes must be explained. We need to inspect the input (and output) closely to capture most of them.

Measures of input

Transcribed data are unbeatable in demonstrating how children encounter linguistic units (e.g., words and signs, gestures, syntactic structures, communicative acts), making resources such as Child Language Data Exchange System indispensable (MacWhinney, 2000). However, rather than looking at individual phenomena (e.g., frequency of passive syntactic constructions, phonetic properties of spontaneously produced vowels), many researchers choose to examine input as an aggregated bundle of such features, as in “CDL” (e.g., Bergelson, Amatuni, et al., 2019; Bunce et al., 2022; Casillas et al., 2020, 2021; Cristia et al., 2019; Greenwood et al., 2011; Sperry et al., 2019). Although useful for general models of language development, such high-level measures guarantee little in explaining how children acquire the specific phenomena relevant for their language communities (e.g., how to change verb inflections to appropriately express specific meanings like “I do” vs. “she does” and “she does” vs. “she did”) in minimally versus highly verb-inflecting languages, like English/Mandarin versus Inuktitut/Murrinhpatha. While we can characterize children’s language development as a general process, theory must also cover the specific and interacting subprocesses of this general process. Precisely, because these subprocesses may be realized differently across diverse contexts, they provide useful theoretical levers into the nature of the learner’s cognitive toolkit. Diversifying the way we measure the input is essential to understanding how these general and specific processes of learning interrelate.

Researchers using aggregate input rates have recently focused on a different issue: how to examine quality over quantity input (e.g., Masek et al., 2021). However, the apparent tension between quality and quantity may ultimately be a red herring. Some measures of quantity (e.g., CDL-A) capture numerous input characteristics that are canonically considered to be of high quality. Furthermore, many of these same characteristics (e.g., contingent turn taking with the key child) are likely culturally specific in what they do for the child’s budding competence as a member of society. Returning to the U.S. versus Mayan contrast mentioned earlier, different caregiver responses and encouragement of children’s contributions predict that turn-taking behaviors will invoke different linguistic and social productions in the two groups—thus, the underlying mechanisms relating turn-taking behavior in the input to the child’s learned behavior in the output will likely differ across populations.

The biggest issue at stake seems to be how we can focus our research on meaningful linguistic practices in the child’s environment while also maximizing the naturalness, reliability, and generalizability of our findings. Failure to take this question seriously can lead to limited or even misleading findings, as I will demonstrate.
Universals in the macro level and potential dead ends

The wide application of aggregate- or indicator-style measures of language can reveal cross-linguistic and cross-cultural patterns in development (i.e., potential universals). For example, similar factors predict parent-reported early vocabulary across a diverse group of languages. This suggests that a handful of biases affect learning across these groups similarly (Frank et al., 2021). However, in general, when aggregate measures are exported from their original context (e.g., from urban industrial societies to rural subsistence ones), researchers risk discovering consistencies that overestimate similarity and underestimate diversity.

In recent work using manual measures of CDL input rates across diverse cultural contexts, my colleagues and I stumbled on one such possible universal—children under age 3 encounter infrequent CDL, an average of ~2.5–5 min/h, across their daytime hours (Bunce et al., 2022; Casillas et al., 2020, 2021). This macro-level consistency obscures tremendous cross-cultural variation in how caregivers use CDL. For example, ethnographic work characterizes caregiver-child interaction in two subsistence farming communities as child-centric on Rossel Island (in Papua New Guinea) and non-child-centric in a Tseltal Mayan community (in Mexico; Brown, 2014; see also Shneidman & Goldin-Meadow, 2012). Yet, these communities appear nearly identical in CDL baseline input rates (Casillas et al., 2020, 2021). Where have the differences gone? I argue that the differences will be visible primarily at the micro scale of moment-to-moment interaction. Across individual interactions, we should see the differences via the people children engage with and the things they do with language. But at the macro scale, when averaging a single measure over a wide variety of contexts and scenarios, we cannot see these differences.

Both macro and micro scales provide informative perspectives on language development. The macro scale tends to highlight the processes that relate to robust, high-level phenomena occurring across groups. The micro scale is equally important in its ability to reveal how the content and style of the input leads children to resemble other language users in their community. Without centering both perspectives, models of children's language cannot reach their full theoretical potential. In this vein, I propose what I term the language in vivo approach to studying children's linguistic experiences.

THE LANGUAGE IN VIVO APPROACH

The core concept behind the language in vivo approach is that we cannot extract language acquisition from the social context of everyday life (see Adolph & Sternberg, 2019; de Barbaro, 2019, for related appeals.) The way children learn to attend to and produce language is entwined with what they do with and around others. Even from the bottom-up perspective of internalizing the statistical properties of the surrounding language, linguistic representations and processes are not fenced off from social experiences (e.g., when to use “blankie” vs. “blanket”). Children do not just learn a language, they learn it to communicate. What they talk about and to whom is shaped by local social and cultural context (e.g., Bruner, 1985; Ochs & Schieffelin, 1984). The theoretical impact of these ideas is that the language learning system must be adaptive and highly flexible. While core mechanisms (e.g., some types of statistical learning) may work similarly across most children, other mechanisms likely vary depending on context (e.g., given the propensity to learn from others’ interactions; Rogoff et al., 2003) or may be specialized to certain domains (e.g., words vs. facts vs. phonology; Foushee et al., 2021; Knightly et al., 2003).

How can we socially center our approach to studying early language experience? Research on language socialization achieves this aim by engaging ethnographically with the community being studied (Ochs & Schieffelin, 1984). However, such an approach is infeasible for most developmental psychologists, who typically have no training in ethnographic methods, and so it is unclear how they should proceed. I propose that we start by scaling down the child’s social language environment to a smaller phenomenon that we can identify reliably and that has meaningful similarities and differences across diverse homes: activity context. From the starting point of activity, we can unfold and integrate many of the concepts, methods, and analytical approaches that are native to developmental psychology with more contextualized considerations of the child’s developmental milieu.

Activity context

Activity as an entryway into language study—indeed as an entryway to language development by the child—has deep theoretical roots in our field (e.g., Bruner, 1985). We know that activity affects the language children encounter and use with others (e.g., Glas et al., 2018; Tamis-LeMonda et al., 2019). The language in vivo approach builds on these ideas with one main difference: It demands explicit exploration of how the studied activities are meaningful to participants in their situated context and in the broader societal, research, and applied contexts used to frame the study. By holding ourselves accountable for the local relevance of studied activities, researchers can enrich their work, grounding broad theories in children's moment-to-moment language experiences—that is, their language experience in vivo (see Figure 2). Two challenges in taking this approach...
How is this activity meaningful to participants?

Example:
Considerations for “informal home pedagogy” activity types in the United States versus Rossel Island, Papua New Guinea

<table>
<thead>
<tr>
<th>General background</th>
<th>Broad domain/themes</th>
<th>Purpose and motive</th>
<th>Implementation</th>
<th>Economic context</th>
<th>Historical context</th>
<th>Cultural context</th>
<th>Social context</th>
<th>Institutional context</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Numeral, education</td>
<td>Academic/educational occupational potential; serious consequences for everyday life if basic knowledge is not mastered</td>
<td>Many available learning aids; conventional basic knowledge/skills (counting and symbol mastery, simple arithmetic); assessments where errors are explicitly corrected</td>
<td>Industrial economy; STEM development at scale</td>
<td>Shift to industrialized economy and rise of the middle class in 19th and 20th centuries</td>
<td>Via math, tied to concepts of intelligence; source of anxiety in many individuals; gendered outcomes in adolescents and adults</td>
<td>Among intimates</td>
<td>Required knowledge to participate in most institutions</td>
<td>Variation across class</td>
</tr>
<tr>
<td></td>
<td>“Home is the first classroom”</td>
<td>Accurate imitation, then abstraction</td>
<td>Often dyadic (parent-child); roles of teacher (adult) and student (child)</td>
<td>Material exchange, timekeeping, employment</td>
<td>Unwritten history; however, the shell money system is ancient and spread throughout the region</td>
<td>Associates</td>
<td>Perceived status of family/despite via child economic success; individual participation in interactions reliant on numeracy (e.g., games)</td>
<td>Among intimates</td>
<td>Structural economic inequality (often interacting with structural racism) leads to variation in outcomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intellectual and social virtu; no serious consequences if not mastered because of kin network support when this knowledge is relevant</td>
<td>Dyadic (parent-child) or multiparty (multiple adults and/or children); roles of teacher (adult) and student (child)</td>
<td>Daily greetings, marriage, death, conflict</td>
<td>Shell money system with bi-linear accounting</td>
<td></td>
<td></td>
<td></td>
<td>Limited input and relevance for children without paternal land connections (e.g., non-Rossel fathers)</td>
</tr>
</tbody>
</table>

FIGURE 2 Examples of researchers' considerations when contextualizing candidate activities. This chart compares informal home pedagogy activities in the United States and on Rossel Island (Papua New Guinea) across several relevant and interrelated contextual dimensions

are how to identify activities and what to do with information about the activities.

Identifying activities

Here, I focus on identifying activities in long-form naturalistic data (e.g., LENA recordings) because multiple activity contexts, often overlapping, are captured from the child's perspective as they go about their day. Moreover, because this type of recording is still relatively new, the field has not yet developed a set of conventions for finding and choosing activity types (see Mendoza & Fausey, 2021, for more information), leaving room to test a preliminary language in vivo approach.

Researchers using this approach should elaborate the purpose and relevance of the studied activities according to participants. Therefore, even just identifying activities to study may require significant investment (e.g., manually annotating existing data, engaging with work outside the researcher's typical domain). Commonly studied activities like free play and book reading are not exempted. They differ meaningfully in their relevance and familiarity across communities, and researchers should explicitly address how these activities are understood by participants, along with implications for data interpretation. The idea is not to wholesale manually annotate recordings prior to data analysis. Researchers should begin with hypotheses or selective attention to certain activity types, together with clear data- or community-internal evidence of the activity’s relevance and some explicit recognition of the limits of generalization.

How do researchers identify activities in line with an in vivo approach? They might begin with a theme...
that implies relevance to participants but requires local specification—for example, asking about how language is (and is not) used by caregivers dealing with infants’ bodily essentials: eating, sleeping, crying, bathing, and dealing with urine/pee. A combination of observational experience and familiarity with the participant community are crucial to identifying the specific activities relevant to this theme. Or researchers might start by establishing the specific participant goals or roles of interest (e.g., typical daily journeys, spontaneous informal pedagogy, encouragement of socially appropriate behavior). Again, the specific relevant activities will differ according to context. Counting activities in the United States differ from recitations of kin ancestry on Rossel Island in Papua New Guinea, but comparing the two provides a more truthful examination of informal pedagogy in each context than would occur by examining one of those activities across both contexts (e.g., asking U.S. parents to engage their children in kin ancestry pedagogy; Figure 2).

Using activity information

Once these recordings are collected, what do we do with them? If the goal is to collect contextually validated findings that are sensitive to both local frames and broader discourses, there is just one answer: detailed inspection and annotation of the data. Such work is costly (Figure 1), but the returns are invaluable in highlighting critical gaps in knowledge and potential ways to bridge them. Such data can also usually be re-used, especially if they are made available to others (e.g., MacWhinney, 2000).

While this work is descriptive, it contributes substantively to theory. Three examples of theoretical work that would benefit from an activity-centered approach are (1) how lexical development is guided by everyday activities at both the domain level (e.g., cars, foods) and the feature level (e.g., deixis [word referent dependence on context], animacy [perceived sentience], register/style of talk), (2) the conditions under which spontaneous informal pedagogy occurs and what material and interactional frames it involves, and (3) how societal expectations about children as language users shape what children attend to and (re)produce during interactions. For each of these topics, researchers would need to do more study via experimental and computational work, elaborating on the proposed processes. The extended in vivo challenge is to find innovative ways to keep more controlled investigations like experiments in touch with the features of interaction that made the descriptive work meaningful.

WHAT WE WILL GAIN

The purpose of the language in vivo approach is to celebrate continuing advances in big data’s ability to document developmental environments without losing sight of what children and their families actually do with language. I have envisioned this work from the perspective of a Western researcher who engages in cross-cultural comparison and has the privilege of focusing on basic scientific questions about human language-learning capacity. But the in vivo framework, firmly grounded in families’ everyday social desires and obligations, is also useful from other perspectives, including those that aim to document or support language communities undergoing change (de León, 2019) or those that explicitly intervene on parents’ or children’s behavior (see Rowe & Weisleder, 2020, for a review).

The in vivo approach can help researchers develop new perspectives on classic questions and delve deeper into less-studied questions. For example, current research aims to explain the processes by which vocabulary arises the way it does by proposing mechanisms ranging from statistical tendencies to analogy-based reasoning. An in vivo approach would follow this path, but emphasize the role of cultural, situational, and historical context to understand more fully how this early word knowledge informs children’s development as competent language users. After all, even when a word is learned early across diverse groups, its representation is built from very different experiences (e.g., the word aunt in an isolated urban context vs. in a multigenerational family settlement).

Regarding less-studied research questions, a prime example comes from peer- and other-directed linguistic input. Whether children are repeating a secret, a swear word, or a character name from a movie they have never seen, caregivers are frequently reminded that their children’s input goes beyond what adults intend to contribute. Establishing how children attend to and reproduce others’ linguistic practices is closely linked to what children see as relevant—an issue that is addressable in an in vivo approach.

CONCLUSION

While it is tempting to scientifically consider language as a symbolic system (viz., grammar and lexicon) isolated from social systems, doing so limits our ability to directly connect theories with families’ everyday experiences. Theory that is constructed without context (or in ignorance of its implicit context) will not apply naturally or equally across diverse domains—it will be limited in its impact. By embracing the social context of language from the viewpoint of activity, researchers using an in vivo approach can equip themselves to construct rich theories of early language learning that have significant potential for practical impact and that foster greater interdisciplinarity and innovation in the field of developmental language science. In summary, to most people, language is how one trades in social,
cultural, and world knowledge. People—including young children—primarily use language to do things with others. Considering social context is critical to building robust theories of how and why we learn and use language. Not just leveraging but embedding social context at the heart of our work would pay off immensely in these respects.

ACKNOWLEDGMENTS

The ideas expressed in this article are the outcome of interactions with the Tzeltal and Rossel Island community members who have participated and engaged with me during my visits to their communities. I could not have done this work without the mentorship of Steve Levinson and Penny Brown. My collaborators and peers in the Daylong Audio Recordings of Child Language Environments consortium and the Language Variation in Input Environments around the World consortium have been indispensable sources of feedback in developing these and related ideas. Finally, I am grateful to Erika Bergelson and Rana Abu-Zhaya for their comments on earlier versions of this article. Any remaining errors are my own.

ORCID

Marisa Casillas © https://orcid.org/0000-0001-5417-0505

ENDNOTES

1 I use the term CDL (i.e., not child-directed speech) to refer to spontaneous language use directed to a child.

2 Large cross-corpus analyses of small data began long before then, largely thanks to the Child Language Data Exchange System, an open online database of naturalistic child language data (MacWhinney, 2000).

3 Among the annotations are when basic speaker types (woman, man, key child, other child) talk, when the key child and a nearby adult talk one after the other, and high-level estimates of speech content (e.g., number of nearby adult words). Open-source alternatives improve this outlook somewhat (see Rässinen et al., 2021).


References


