# A vocabulary checklist for early lexical development in Tseltal

Marisa Casillas
University of Chicago, USA

Ruthe Foushee
The New School for Social Science, USA

Humbertina Gómez Pérez
Juan Méndez Girón
Gilles Polian
Centro de Investigaciones y Estudios Superiores en Antropología Social, Sureste, Mexico

Kennedy Casey
Princeton University, USA

Penelope Brown
Max Planck Institute for Psycholinguistics, The Netherlands

**Abstract:** This study presents a checklist for measuring the expressive vocabularies of young children acquiring the Mayan language Tseltal. Adapted from the vocabulary checklist of US English MacArthur-Bates Communicative Development Inventories and tested with 84 Tseltal-acquiring children (9–23 months), this Tseltal checklist shows desired within- and across-subcategory variability in age of acquisition, expected age- and gender-related change, and typical patterns of relative over- and under-representation in the most cross-linguistically stable domains. We discuss potential uses of the checklist and the next steps in its future development.

**Keywords:** vocabulary checklist; parent report; Tseltal; Mayan

**Corresponding author(s):** Correspondence concerning this article should be addressed to Marisa Casillas, 1101 E 58th St, Chicago, IL 60637. E-mail: mcasillas@uchicago.edu

**ORCID ID(s):** Marisa Casillas: https://orcid.org/0000-0001-5417-0505; Ruthe Foushee: https://orcid.org/0000-0002-8361-1725; Gilles Polian: https://orcid.org/0000-0002-0642-2875; Kennedy Casey: https://orcid.org/0000-0001-5326-3080

**Citation:** Casillas, M., Foushee, R., Gómez Pérez, H., Méndez Girón, J., Polian, G., Casey, K., & Brown, P. (2025). A vocabulary checklist for early lexical development in Tseltal. *Language Development Research*, *5*(3), 245–275. http://doi.org/10.34842/ldr2025-862

## Introduction

Vocabulary checklists, such as the MacArthur-Bates Communicative Development Inventories (CDI, Marchman et al., 2023), are efficient and reliable tools for gleaning general measures of children's lexical and grammatical development (Fenson et al., 1994; Frank et al., 2021; Marchman & Dale, 2023). Rather than measuring a child's actual lexical inventory or current morphosyntactic knowledge, these tools generate a quick snapshot of children's overall language development that—combined with other information sources (e.g., interviews, longitudinal data, other assessments)—can help identify delays in language development.

Further, in understudied and underresourced language communities, checklists can help to rapidly map out some typical pathways for lexical development (e.g., Alcock et al., 2015; Southwood et al., 2021; Vogt et al., 2015; Weber et al., 2018), particularly when they complement observational and experimental data in the creation of language acquisition sketches (Hellwig et al., 2021). Such sketches can be used to establish typical development milestones and learning materials for clinical and educational professionals who work in the language community and who support language maintenance. Such documentation may be especially useful for communities experiencing language endangerment—often due to colonization—when there is a strong interest in heritage language maintenance. Heritage language loss has been closely tied to the loss of cultural knowledge (Bromham et al., 2020; Kik et al., 2021; Maffi, 2005), and child language development is an appropriate target for projects seeking to more broadly fortify cultural identity in colonized societies and minoritized communities.

The semi-standardized format of vocabulary checklists also makes cross-linguistic developmental comparison possible at an unprecedentedly wide scope: as we write this, researchers have begun the process of adapting the CDI into at least 117 language varieties, and CDI *data* is freely available from over 92k children sampled across 42 language varieties on the Stanford Wordbank archive (https://wordbank.stanford.edu/). Such large datasets can be used to identify trends in lexical development that crosscut structurally and culturally diverse language communities as well as clear points of variation in development that afford new insights into the human cognitive system (Braginsky et al., 2019; Frank et al., 2021).

### Adaptation, Not Translation

Originally developed for US English-acquiring children, the CDI has been authorized for adaptation in 117 other language varieties, many of which now have available adaptations (https://mb-cdi.stanford.edu/adaptations.html). The CDI adaptation materials speak of "adaptation" rather than "translation" of an instrument because the specific items on one list may not be culturally or linguistically appropriate for another.

For example, the relevant common animals, foods/drinks, household objects, and more, can be expected to differ between geographically and economically diverse language communities. Even when an item is relevant in multiple language communities, its meaning may substantially differ between languages, leading to false equivalences at the item level. That said, highly overlapping conceptual categories can sometimes be identified across diverse lexicons ("unilemmas"), in which case they can provide unique insights into cross-linguistic conceptual development; an affordance particularly useful when studying children learning more than one language (Byers-Heinlein et al., 2024; Tan et al., 2024).

Another important consideration for CDI adaptation is when and how to use inflected forms on the checklist. For example, in the language we focus on presently—Tseltal—bare transitive verb stems are ungrammatical. There is no single inflected verb form for transitive or intransitive verbs that is so common or so representative of the major inflectional paradigms as a whole that it can confidently be used to assess vocabulary development.

Lastly, while item adaptation is discussed extensively in the CDI's online materials, adaptations in the manner of data collection are less often addressed. The CDI was originally intended as a fillable paper form that could be given to parents at a lab visit or via mail. This format has been seamlessly adapted to online data collection via secure webforms (deMayo et al., 2021). However, for populations in which the primary caregivers are not comfortably literate in the language under study, in which a more conversational interaction is normatively appropriate, and/or in which primary childcare is divided among multiple family members, the basic format of one participant filling out a written form is going to fall short of accurately capturing information about child development.

## Study Population

The present study focuses on Tseltal-speaking families with small children. Tseltal is a Mayan language spoken by more than 500k people in southeastern Mexico. Most Tseltal-speaking communities sit in central and northeastern Chiapas. There are three primary dialect areas (north, central, and south); we focus here on the central geolect, specifically the variant spoken in the Tenejapan municipality. Linguistically, Tseltal has several typological features that are understudied in the language development literature at large, including ergative-absolutive alignment, verb-first basic word order, an agglutinating and mildly polysynthetic morphology, a large inventory of "specific" verbs, an absolute frame of reference spatial system, ejective consonants, and more (Brown, 1997, 1998a, 2008; De León, 1999a, 1999b, 2001; Polian, 2013, 2017).

Tenejapa is a rural municipality in which subsistence farming (e.g., corn, beans, squash, potatoes, coffee, and more) is a primary way of life, often supported by one or more other sources of income (e.g., taxi driving, labor outside of the community, etc.). Most children in the sampled communities grow up primarily hearing and speaking Tseltal, though they may commonly encounter Spanish in some village areas (e.g., at public events on school grounds) and on television and the radio. Many Spanish words have been borrowed into the contemporary Tseltal lexicon, and share similar, if not identical, phonological forms with their original Spanish forms (e.g., *wakax* /wɑkɑʃ/ (Tseltal) vs. *vaca* (Spanish) "cow"; *tijeras* (Tseltal[1] and Spanish) "scissors"). Until they are walking reliably, around their first birthdays, Tseltal children spend most of their waking day in a sling worn by their mother. Sometimes another (typically female) relative carries the baby instead. Even after they are walking, young Tseltal children tend to stick close to their primary caregiver, and are often carried while they sleep or while the mother goes about her business in and out of the home.

Thus, in our targeted age group, children spend the majority of their waking time in very close proximity to their primary female caregiver(s). Many Tseltal children grow up in multigenerational households that include the child's parents and paternal family (father's parents and father's brothers' wives and children). Children and adults share some load in caring for young infants when the mother is not available, and it is common for married women to have 3–5 children; therefore our respondents have a great deal of experience caring for and observing children by the time they themselves reach motherhood. While Tseltal has previously been characterized as non-child-centered (Brown, 1998b, 2011; Brown & Gaskins, 2014), infants' primary caregivers—in most cases, their mothers—have immense amounts of exposure to their children's vocalizations from carrying and caring for them most of the day, most days. These caregivers are therefore extremely well positioned to report on what their children say.

Notably, there is a great deal of variation in the number of years of school Tseltal women currently complete, and even those who complete secondary education have very rare opportunities to read and write in Tseltal after the first two years of primary school (the language of schooling is Spanish). Therefore, any Tseltal checklist must be conducted via verbal interview. This technique has been used with vocabulary checklists before when interviewers cannot expect fluent literacy in the language(s) being examined (Alcock et al., 2015; Hamadani et al., 2010; O'Shannessy et al., 2024; Prado et al., 2018; Southwood et al., 2021; Vogt et al., 2015; Weber et al., 2018).

[1] 20 years ago, this borrowing of Spanish "tijeras" was pronounced more like *texerex* /teʃɐreʃ/ in Tseltal. However, the conventional Spanish form has since taken its place. There are many such cases in the everyday Tseltal lexicon, which affects the responses given in a vocabulary checklist task.

## Method

Our methodological approach strikes a balance between the MacArthur-Bates CDI guidelines for checklist adaptation and our team's own immediate research priorities. We set out to create a vocabulary checklist that would be structured and collected such that our data could be used in large-scale cross-linguistic analyses (e.g., Wordbank) while also maintaining sensitivity to local Tseltal accounts of typical early word knowledge—Tseltal early noun and verb development is the target of our current work involving this questionnaire.

Our methods thus follow similar guidelines to what is suggested for a "Level 2" adaptation of the vocabulary checklist of the CDI, with three important exceptions: (1) Our present research questions motivated a close look at an age group (12–20 months) on the border of the CDI: Words & Gestures (W&G; 8–18 months) and CDI: Words & Sentences (W&S; 16–30 months), so our adaptation initially included lexical and grammatical items from the US English W&S checklist *and* gesture items from the US English W&G checklist; (2) we only measured expressive item knowledge—and for most participants we additionally collected a reported phonological form of children's word productions; (3) we paused the inclusion and further development of grammatical items after getting uninterpretable data from a handful of early participants.

As described below, we incorporate *local* accounts of typical early word knowledge by using our transcribed corpus of Tseltal child speech and, via an iterative process of checklist refinement, asking early participants to tell us any words their children say that were not included in the list they had just completed. Following the CDI adaptation guidelines, the organization of items into semantic categories (e.g., "Animal names", "Food or drink") mostly aligns with the groupings suggested by the CDI board (which originate in groupings from the US English CDI checklist). However, some exceptions are noted below. Moreover, because "semantic category" is simply a type of metadata for each item, we note that the items can easily be subset and/or re-categorized into more locally meaningful categories by anyone wishing to use these data in the future—one of many reasons why it is essential that we share all data for all items as openly as possible.

### Participants

We collected checklist data from 84 Tseltal-acquiring children between ages 9 and 23 months (mean = 16.03; 38 girls and 46 boys). This sample size for checklist responses is within the typical range for prior CDI adaptations at the pre-validation and pre-norming stage; however, we note that sample sizes for studies targeting larger age ranges and studies associated with public health research can be much higher (e.g.,

our unsystematic review reveals a range of $N$ = 36–2,418[2]). Personal communication with CDI board members verified that there is no standard sample size for creating a preliminary adaptation, and that a more primary issue at this stage was identifying a range of easier and harder words within each checklist subcategory. We thus aimed for 75 participants for our preliminary checklist dataset, ultimately exceeding that goal with a final sample of 84 participants. Based on our past experience in these communities, we had assessed a sample size of 75 to be feasible during a two-week bout of intensive testing during which we simultaneously ran other studies.

To varying extents, all children also had direct and overheard exposure to Spanish, and sometimes other Mayan languages (e.g., Tsotsil) depending on the composition of their household. However, all children were reported by their primary caregiver(s) to hear Tseltal most of the time, to be typically developing, and to have no known problems in hearing, seeing, or speaking. Language background and typical development were thus determined by simply asking the primary caregiver: (a) whether they spoke to the child only in Tseltal or also in Spanish or other languages (and, if the latter, how often) and (b) whether the child had any problems with hearing, seeing, or talking, or any other problems in developing. When caregivers reported exposure to a second language, they typically gave a verbal description of quantity or context (e.g., "sometimes" or "with his uncle"). Using this information, the experimenters (typically HGP and MC) verified that each child was Tseltal dominant (i.e., vast majority input in Tseltal) before proceeding. All participants resided in the mountainous Tenejapa municipality of central Chiapas. Checklists were collected in 12 rural villages within that municipality, between June and October 2023. The study plans were pre-approved by the University of Chicago Institutional Review Board (IRB23-0244).

## Checklist Items

### *Initial Item Adjustments*

The vocabulary checklist reported here was developed over six iterations across the 84 participants. As a starting point (Version 0), we reviewed the checklist used in Brown, Gentner, and Braun (2005), which was based on an adaptation of the US English MacArthur-Bates CDI (Fenson et al., 1994; Marchman & Dale, 2023; Marchman et al., 2023). The original Brown et al. (2005) Tseltal checklist contained 613 items across 19 categories taken from the English CDI vocabulary checklist at that time, listed in

[2] $N$ = 36 for Australian Aboriginal English/Kriol/Other in Jones et al., 2020; $N$ = 50/58 for Kiswahili/Kirigiama in Alcock et al., 2015; $N$ = 100 for Czech in Jarůšková et al., 2024; $N$ = 110/115/105/98 for Afrikaans/isiXhosa/South African English/Xitsonga in Southwood et al., 2021; $N$ = 241 for Wolof in Weber et al., 2018; $N$ = 566 for Changana/Ronga/Portuguese/Other in Vogt et al., 2015; $N$ = 29/869 for Chichewa/Chiyao and Krobo/Ewe/Twi/English in Prado et al., 2018; and $N$ = 2418 for Bangladeshi in Hamadani et al., 2010.

order as: Sound effects and animal sounds (11); Animal names (37); Vehicles (8); Toys (9); Food or drink (46); Clothing (21); Body parts (22); Furniture and rooms (32); Small household items (49); Outside things and places to go (42); People (34); Games/routines (21); Action words (185); Words about time (13); Descriptive words (39); Pronouns (17); Question words (8); Prepositions and location (15); Quantifiers (4). These categories are slightly different from those featured in the US English CDI checklist, with Brown's adaptations for Tseltal including: (a) combining Outside and Places to Go, given that many relevant places to go are primarily outdoors or semi-outdoors; (b) considering all verbs under Action words, given that "Helping verb" is not a sensible linguistic category in Tseltal; and (c) removing the small and mixed category of Connecting Words, which in Tseltal could only include one subordinator, one subordinator that also acts as a definite article, one coordinator that also acts as a preposition, and two borrowed conjunctions from Spanish. In our adaptation, we respect these category decisions by Brown, which still largely align with the CDI's standards.

In the early 2000s, Penny Brown interviewed the families of 5 young Tenejapan Tseltal-acquiring children to document their vocabulary production and comprehension. The study was discontinued after the first 5 participants because the interviews lasted 2 or more hours and because the experimenter (PB) maintained doubts about the children's true word production and comprehension, compared to what was reported on the form (Brown, personal communication). We ran through this entire checklist, via interview, with the mother of an apparently linguistically advanced 18-month-old boy. Most words on that initial checklist were presented as grammatical, bare lexical stems, but verbs were presented in the incompletive, first person singular form (e.g., *ya x-ben-on* /jɑ ʃbenon/ (I walk)) and body parts were presented with a first person singular possessive morpheme (e.g., *j-k'ab* /xk'ɑb/ (my-hand)). Following checklist completion, we reviewed candidate words that were not on the list, and also words that were on the list but were unlikely to be useful in tracking young children's vocabulary in that mother's experience. This participant's checklist was considered pilot data; it is not included in the primary analyses of the 84 children.

In Version 1 of the checklist, using the pilot data from the advanced 18-month-old child, we cut the list down to 212 words and 9 gestures. We only asked about whether the child *produced* the word or not (i.e., we did not ask about comprehension). We also exploratorily asked about what kinds of sentences and errors the child was producing. These questions about sentences and errors elicited very different types of responses among the first three caregivers tested ($N = 3$; e.g., regarding errors: "none", "not talking enough to say", and "calling family members by the wrong name"). However, the other (word and gesture) caregiver responses appeared to function more reliably. In the process of conducting the first three interviews, we found it useful to provide multiple forms for each verb, namely: (a) a fully inflected incompletive first person form, (b) the same form without the aspect markers, and (c) the bare verb stem (e.g., for *tsak* /tsɑk/ (to take/grab) we would give the options "*ya j-tsak, j-tsak, o tsak*?" /jɑ xtsɑk/,

/xtsɑk/, or /tsɑk/). Caregivers were asked to respond 'yes' in the case that their child said any of the forms listed. The bare verb stem for transitive verbs (e.g., *tsak*) is ungrammatical, but such forms are typically found in the spontaneous speech of Tseltal-acquiring children (Brown, 1997), as they are in children learning other Mayan (De León, 1999b) and some polysynthetic languages (for a review, see Kelly et al., 2014). For each verb, caregivers were thus asked to report which of the options (if any) their child produced—in many cases, caregivers reported the bare stems.

In Versions 2 (267 words; 9 gestures; $N = 5$) and 3 (263 words; 9 gestures; $N = 7$) we removed the sentence type and error questions, standardly added the three verb forms described above for all verb items, and added more words, especially harder words that were likely to be known only by older toddlers. We removed some words that were considered to be old fashioned, or which were homophonous with other items on the list (e.g., *ja'* /xɑʔ/ is both the word for water and a demonstrative; we retained the former). Version 4 (263 words; 9 gestures; $N = 3$) maintained the same words as Version 3, but we changed how we asked about each item. Instead of simply asking whether the child produced each item or not, we now asked: "Does your child say this word? If so, *how* do they say it?" We would then write an impressionistic orthographic transcription of what the caregiver produced. For example, for *xawin* /ʃɑwin/ (cat), a typical response was: "'yes' the child says it" and they say it like "*win* /win/".[3] With this additional question, administrations of the checklist still typically only lasted 10–20 minutes in duration, and caregivers seemed to overall enjoy doing impressions of their young child's productions. In Versions 5 (273 words; 9 gestures; $N = 20$) and 6 (299 words; 9 gestures; $N = 46$), we continued adding harder words and missing words typical of early production, in addition to making minor changes to item order and categorization (e.g., *ton* /ton/ (rock) was moved from the "Toys" category to the "Outside things and places to go" category). We kept the same gesture items from Versions 1–6, but the gesture labels were reworded for greater clarity in Version 5 (the interviewer typically demonstrated the gesture, rendering the intended meaning clear in all cases).

### *Final Item Selection*

As a final step, and following Alcock et al. (2015), we used the collected data to pare the list back down to ~250 items that include a range of earlier- and later-produced words within each sub-category of the checklist (e.g., Animal names; Vehicles; Toys; Food or drink; etc.). This process needed to be completed manually, and so was pre-

[3] To demonstrate the diversity of reported productions here, the unique reported productions of *xawin* (/ʃɑwin/) among the 84 participants were: *xawin, chawin, xamin, xiwin, xa, xaw, xux, waw, win, wixwix,* and *meumeu*. Onomatopoeic form substitutions like "meumeu" were marked, but noted as a different form, as were word forms for the same referent in another language (e.g., *gato,* from Spanish).

registered on OSF (https://osf.io/z8hdk) to mitigate bias in item selection. Our process for item selection followed Alcock et al. (2015)'s description fairly closely: First, find items that reach 50% production at any month—these will be the core words for the list. Then, add the 20 least-frequently produced words among those that were known by at least 20% of children at any age. Then, add the 20 least-frequently produced words among those that were known by at least 5% of children at any age. These are the hard items—the ones we only expect older children to produce.

We then generated the mean age of acquisition (AoA) for each item on the full list of 299 words and systematically reviewed the currently included words for a relatively even distribution in the 12–20-month-old age range and within each sub-category (e.g., Animal names; Vehicles; Toys; Food or drink; etc.). When the AoA distribution appeared uneven, we added, removed, or swapped out items to improve the representation of easier and harder words. In this process we also ensured that the verbs were somewhat balanced in transitivity, and that the transitive verbs included both specific and general verbs from Brown (1998a)'s list.[4] We also checked that the nouns were fairly balanced in concreteness, animacy, and ability to be handled. We kept an eye out for near synonyms and removed them unless each individual item was separately motivated (e.g., *bistuk* and *bi yu'un* both mean something like English "why", so we removed the latter; quantifiers *bayel* and *uts* both mean something like English "a lot", but the latter can also be inflected as an adjective). Where possible, we tried to keep mini sets of words within categories that are of theoretical interest (e.g., kinship terms, spatial terms, etc.). This process of scanning, swapping, and re-checking each sub-category and across age took many iterations.

This final list of 251 words and 9 gestures was checked with our co-authors, which include a native speaker of Tenejapan Tseltal, a near-native speaker of Tenejapan Tseltal (who is a native speaker of the closely related language Tsotsil), and a linguist specializing in Tseltal. The final list was accepted for current analysis, but cases of potential "missing" items (e.g., lower-frequency household items and animals) were noted to be tested in future versions of the checklist. The 251 words include 231 words from Brown et al.'s (2005) CDI-inspired checklist. The checklist items thereby overlap substantially with the US English Words and Sentences CDI, including 113 of the 639 unique unilemma concepts within the comparable categories of the US English Words and Sentences CDI.[5] They are divided among sub-categories as follows: Sound effects

[4] This distinction is relevant for transitive verbs which are either very restricted in the patients they take (heavy/specific verbs: e.g., *we'* /weʔ/ "eat-tortilla" and *top'* /topʾ/ "shatter") or which are instead very open (light/general: e.g., *ak'* /ɑkʾ/ "give/put" and *ai'y* /ɑʔi/ "see/hear/perceive").

[5] The 251 items also include 30 concepts not included in Wordbank's current unilemma inventory: flea, louse, buzzard, VW Beetle, commuter pickup truck, cold cornmeal (beverage), warm cornmeal (beverage), sombrero (hat type), wrap (clothing type), Tseltal skirt (clothing type), hammock, metal roofing material, stirring stick, peso, milpa (mixed-plant field), namesake, little one (term of endearment),

and animal sounds (10); Animal names (16); Vehicles (8); Toys (5); Food and drink (21); Clothing (10); Body parts (14); Furniture and rooms (13); Small household items (19); Outside things and places to go (15); People (12); Games and routines (14); Action words (41); Words about time (6); Descriptive words (15); Pronouns (10); Question words (6); Prepositions and location (10); Quantifiers (6).

**Figure 1.** ***Participants were verbally interviewed by native and near-native speakers who most often noted responses on paper copies of the checklist.***

### *Procedure*

Participants were recruited via word of mouth. As illustrated in Figure 1, participants were either interviewed by appointment, in their home, or came to visit the interviewer(s) in another location during open testing hours (e.g., by a local school, in a neighbor's house, etc.). Most participants signaled to the interviewer in advance when they would be able to meet, via verbal agreement, direct/text message, or phone call. When the interview began, participants were first engaged in a consent process that described the context for the research study, their right to stop at any time, and their compensation, among other topics (see Appendix A for the full points covered). Consent was sought in a series of informally phrased points, with wording varying

Mrs (honorific type), Mr (honorific type), older sister or father's side aunt/cousin (kin type), older brother or older first cousin (kin type or honorific), older brother of a female (kin type), walk-on (greeting), here-take (presentational word), okay/fine/agreed, let's go, sound to call chickens, perceive/experience, uphill, and downhill.

slightly from participant to participant to ensure a more conversational flow. In our experience, this more interactional form of consent is more effective in eliciting questions and demonstrations of understanding from participants.

Once participants consented to participate, we began with the instructions. For each word, participants were asked to indicate if their child says (or used to say) that word and—for participant 16 onward (Version 4 onward)—how their child says it. During recruitment we asked to interview the child's primary caregiver, which is typically the mother in this community. However, in practice, our interviews often *additionally* included aunts, grandmothers, and older siblings who had spent significant time with the child and who were present during our interview period. Mothers were the primary interviewees, but attending family members sometimes offered their opinion on whether the child said a word or not, sometimes in response to a bid from the mother and sometimes spontaneously—ultimately we always accepted the mother's final judgment.

Participant responses were recorded in real time. When there were two interviewers present, one focused on talking with the caregivers and one on writing down responses. When there was only one interviewer present, they were responsible for both talking and writing. Consent and interviews were conducted by native or near-native Tseltal speakers who reside in one of the testing villages (HGP or another team member). Foreign researchers (RF, MC) served only as second interviewers, noting down responses as they were given. Interviewers were typically able to complete this entire checklist interview process in 10–20 minutes. We additionally note that the brevity of this interview made it easy to combine checklist data collection with experiment-based data collection in the same short visit (typically 20–40 minutes).

Most data were collected on paper copies of the list, but early versions were directly typed into a spreadsheet on a laptop, and a handful of sessions were collected via pdf markup on a tablet computer. Any checklists collected by our local, independent interviewer (HGP) were photographed and sent via encrypted message to our primary analysts (MC, KC, RF) for digital entry into the project database. Database entries were quality checked (MC, KC) prior to analysis.

## Results

Our aim in the present paper is to test whether the checklist functions as expected, as an instrument designed to map variation in typical lexical development among Tseltal-acquiring children. We divide our analyses into three parts: (a) age of acquisition checks, (b) age-related change, and (c) relative representation across checklist categories. In all of the analyses below we use children's *conceptual* vocabulary (i.e., include an item if they are reported to say it in Spanish rather than the provided Tseltal item). The vast majority of reported productions—99.10%—aligned with the provided

Tseltal wordforms, or a referentially acceptable alternative in the speech community (e.g., substituting "darkness" for "night"; see Appendix C). Child wordforms reported in Tseltal varied between individuals, ranging from 92.59%–100% of items with Tseltal responses (mean = 99.02%). If we use Tseltal-only responses rather than all responses, it makes no qualitative difference in any of the findings reported. We exclude reported productions that don't map well onto the intended target item, including: associated words (e.g., "bite" for "snake"), overgeneralizations ("car" for "taxi"), and onomatopoeia used in place of object labels (e.g., "moo" for "cow"). The produced forms for the excluded items are nearly always captured by another item in the checklist (e.g., "car" for "car") and make up 0.66% (less than 1%) of the checklist responses we gathered.

Note that because the checklist was developed in versioned waves, the 251 final items vary in the number of times they were assessed; 219 items have data from all 84 children, 5 items have data from 76 children, 25 from 46 children, and 2 from 38 children. We do not impute missing data for any of the 251 items in the analyses below. Instead, we base proportions by item and by participant on the total data available for each case.

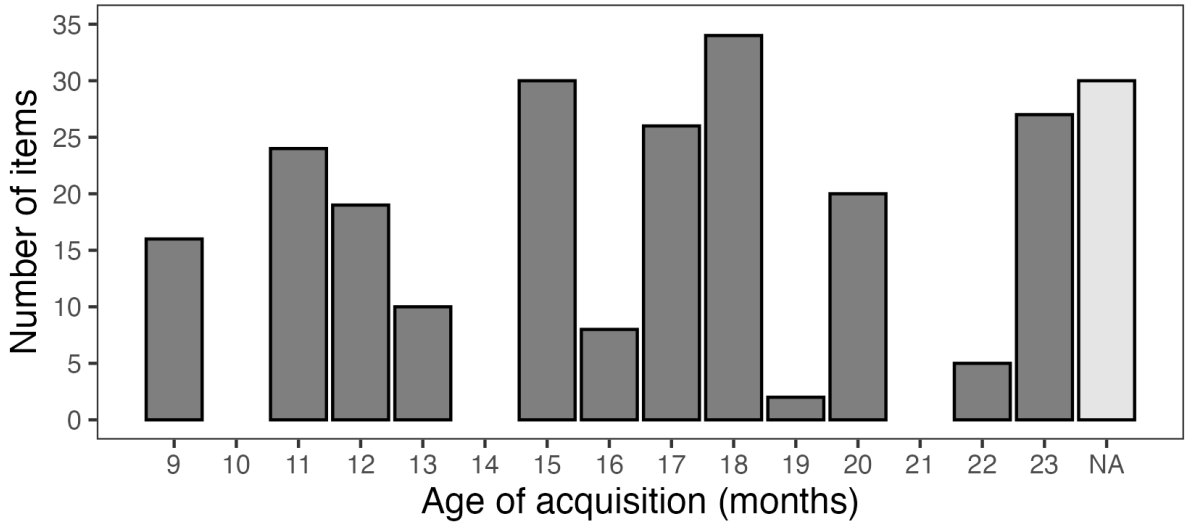


**Figure 2.** ***Distribution of age of acquisition—the first age with ≥ 50% of participants producing the item—over all items. Items listed in the "NA" bar (far right) do not reach ≥ 50% production at any observed age.***

### *Age of Acquisition Checks*

By design, the checklist should include a range of earlier- and later-acquired words within and across all sub-categories. In the methods, above, we described how we

attempted to achieve this distribution of word difficulties. Here we report on our success in doing so. For each of the 251 words on the final checklist, we define the age of acquisition (AoA) as the first age at which at least 50% of the participants were reported to produce the word.

The distribution of AoA over the entire checklist is shown in Figure 2. Indeed, we see a reasonably balanced distribution of AoAs between 9 and 23 months, the ages tested. Note that 11.9% ($N = 30$) items did not achieve 50% production at any age. This is partly attributable to having relatively little data for children older than 20 months ($N = 3$ children); based on collaborator discussion, we predict that many of these words would have an AoA before 24 months with more data collection (see Appendix B for further consideration on this issue).

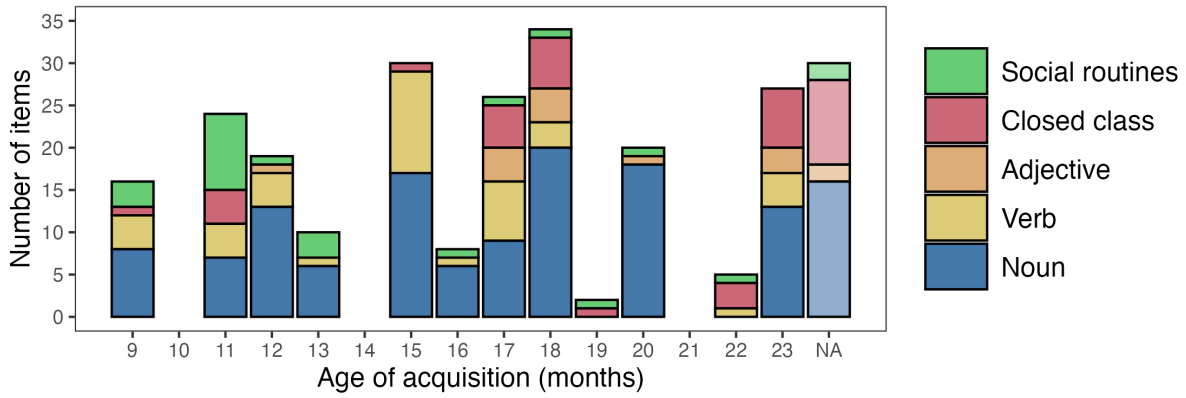


**Figure 3.** ***Distribution of age of acquisition over all items, separated by major syntactic types (Noun, Verb, Adjective, Closed class, Social routines). Items listed in the "NA" bar (far right) do not reach ≥ 50% production at any observed age.***

In Figure 3 we look at the data from a similar perspective, only this time breaking the data down by major syntactic types on the checklist, including three open-class categories (Noun, Verb, Adjective), a closed-class category (Closed class; e.g., pronouns, quantifiers), and a category for fixed expressions associated with everyday games and routines (Social routines). Again, we see that there are early, middle, and late AoAs within each type, though some categories are more limited in their spread than others. For example, the first adjective AoA does not occur until 12 months. Other categories are slightly unbalanced in their AoA distribution. For example, there is a cluster of social routine items acquired at 11 months. These slightly asymmetrical distributions are expected, considering that some word types (e.g., social routines) are typically acquired earlier or later than others (e.g., adjectives) given differences in salience, frequency, conceptual complexity, etc. (Arunachalam & Waxman, 2010; Bates et al., 1994; Braginsky et al., 2019; Frank et al., 2021; Gentner, 2006). In general, however, the items in the present checklist meet the aim of including a range of relatively

easier and harder items within each sub-category.

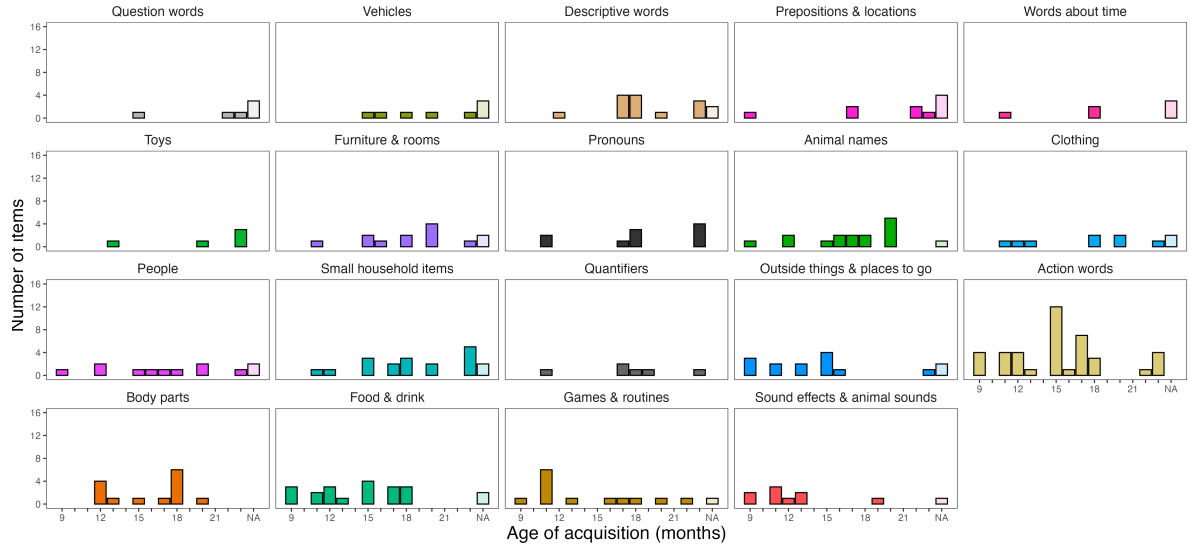


**Figure 4.** ***Distribution of age of acquisition over all items within each content sub-category of the list, sorted from the latest acquired (upper left; Question words) to the earliest acquired (lower right; Sound effects & animal sounds).***

One final check on AoA distributions comes from the semantic sub-categories standardly associated with CDI checklists, such as: Food and drink, Vehicles, Clothing, and more. Figure 4 shows the AoA distributions for items within each content sub-category. Similar to Figure 3, general skew across the observed age range is more apparent for some categories than others (e.g., Question words tend to be acquired later; Food or drink words tend to be acquired earlier). But even in the smallest sub-categories (e.g., 6 or fewer items in each in the categories: Question words, Words about time, Toys, and Quantifiers) there is a clear spread in AoA.

Overall, we find that the checklist effectively achieves its aim of including easier and harder words within each major sub-category and across the checklist as a whole.

### *Age-Related Change*

Another checklist outcome worth assessing is age-related change. If the checklist is working as expected, we are very likely to see an increase in productive vocabulary size with age, particularly: evidence for an acceleration in word production starting around 18 months (Fenson et al., 1994; Marchman & Dale, 2023), and (more tentatively) larger vocabularies for female than male children (Kachergis, Francis, & Frank, 2023; Mayor & Plunkett, 2011). To test these predictions, we fit Generalized Additive Model for Location, Scale and Shape (GAMLSS) models (Rigby &

Stasinopoulos, 2005) to generate approximate percentiles for overall vocabulary size across age, overall (Figure 5) and for female versus male children (Figure 6; see also Jackson-Maldonado et al., 2024). These models are restricted to assume monotonic growth, such that vocabulary size strictly increases across age.[6]

In the overall data, the typical trajectory—at the 50th percentile—clearly suggests an acceleration in word production shortly after the first birthday. As children get older, and larger vocabulary sizes become more likely, we also see greater reported variability in observed vocabulary sizes, with an estimated spread between children in the 10th and 90th percentile of 200 checklist words by 24 months.

When we divide the data by child sex and examine the 50th percentile trajectories, we see that, numerically, female children are consistently reported to have larger vocabularies. However, this difference is small and non-significant, providing no clear evidence for early sex-based vocabulary differences in Tseltal.

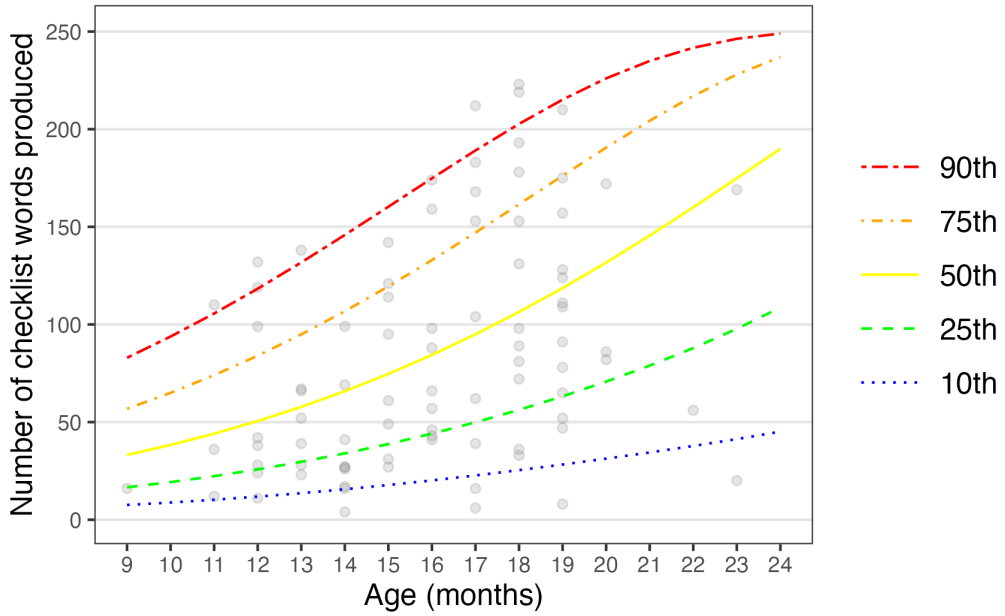


**Figure 5.** ***Number of checklist items produced across age by individual children (gray dots), showing percentiles (10th, 25th, 50th, 75th, and 90th) for lexical production.***

[6] gamlss(produces ~ pbm(age, lambda = 10000), sigma.formula = ~ pbm(age, lambda = 10000), family = BE, data = vocab.data)

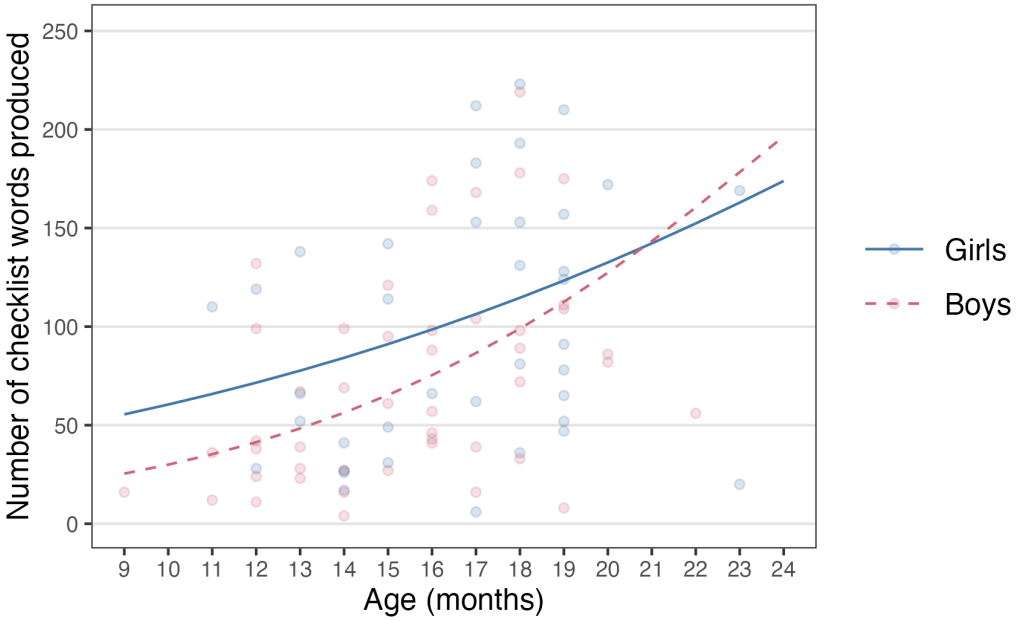


**Figure 6.** ***Number of checklist items produced across age by individual male (red) and female (blue) children, showing the estimated trajectory for the 50th percentile of lexical production in each group.***

In sum, age- and gender-related differences in reported vocabulary size, as measured by the checklist, largely accord with patterns observed previously in the literature using the checklists in other CDI instruments (Fenson et al., 1994; Jackson-Maldonado et al., 2024; Kachergis et al., 2023; Marchman & Dale, 2023; Mayor & Plunkett, 2011).

### *Relative Representation Across Checklist Categories*

While we do aim to have a range of easier- and harder-to-acquire items within each sub-category of the checklist, we can expect systematic differences in word learning between categories due to their salience, conceptual complexity, and more (Arunachalam & Waxman, 2010; Bates et al., 1994; Braginsky et al., 2019; Frank et al., 2021; Gentner, 2006). Our final analysis thus sketches preliminary evidence for variability across the checklist's content categories, such as Animals, Small household items, Food or drink, and more.

Following Braginsky et al. (2019), we use a measure of relative representation to understand whether children's production of words within each sub-category is greater or lesser than we would expect on the basis of random, unbiased development. The analysis makes use of the idea that, if lexical development is unbiased, children should know words in any category proportionally to their overall word knowledge. So, a child who produces 50% of the words on the checklist should, on average,

produce 50% of the words within any given sub-category. If that child produces more than 50% of the words in a sub-category, we can consider that category to be relatively over-represented. If the child produces less than 50% of the words in a sub-category, it would be relatively under-represented.

We can make some broad predictions for this analysis based on work from the Wordbank team (Braginsky et al., 2019; Frank et al., 2021). Namely, while sub-category rankings vary across languages, some domains show consistent over-representation in development (Sounds, Games & routines, and Body parts) while others show consistent under-representation (Places and Time words, Frank et al., 2021). The Tseltal data are consistent with this prediction (Figure 7). Sounds, Games & routines, and Body parts make three of the four most over-represented categories. Spatial and Time words are within the six most under-represented categories.

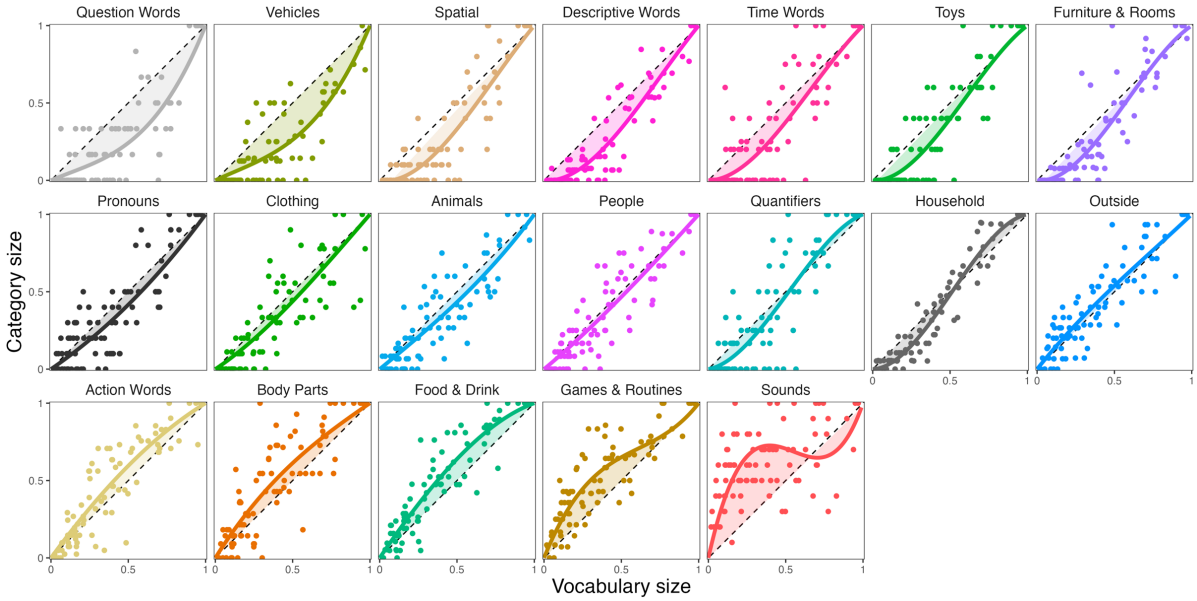


**Figure 7.** ***Distribution of age of acquisition over all items within each content sub-category of the list, sorted from the latest acquired (upper left; Question words) to the earliest acquired (lower right; Sound effects & animal sounds).***

We can assess the extent of bias in learning by measuring the area between the diagonal (unbiased learning) and the fitted line (observed data). Again, following Braginsky et al. (2019), we randomly sub-sampled and measured this area 1,000 times to create bootstrapped 95% confidence intervals around the estimated bias size. The resulting estimates and confidence intervals are shown in Figure 8. Sub-categories with effect size distributions overlapping with zero show no evidence for bias in learning; those below zero show evidence for under-representation, and those above zero for

over-representation.

The bootstrapping analysis suggests a significant overrepresentation of Sounds, Games & Routines, and Body parts, among other categories, as well as a significant underrepresentation of Time words and Spatial words. The pattern accords well with the cross-linguistic predictions based on empirical observations from Wordbank (Frank et al., 2021).

In brief, variability across checklist sub-categories accords with the most consistent patterns found in prior work on the vocabulary checklists of CDI instruments.

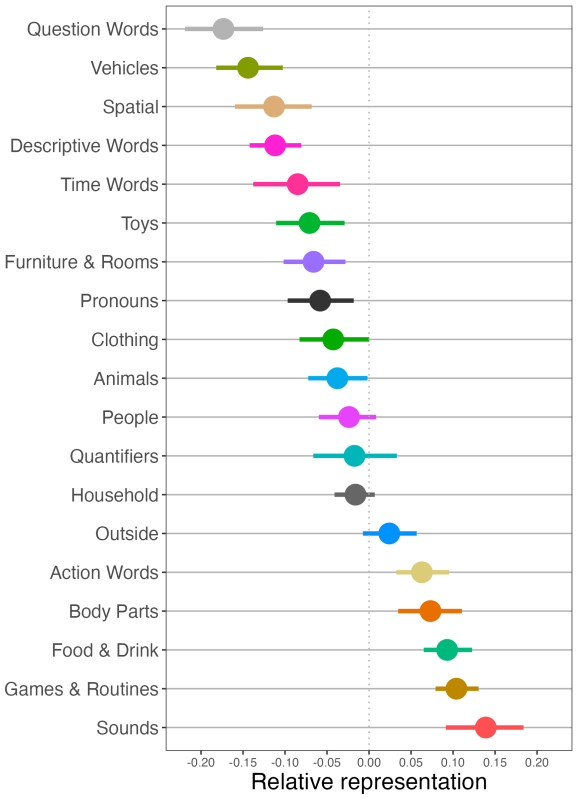


**Figure 8. Bootstrapped relative representation effect sizes (x-axis) for word production across each content sub-category of the checklist (y-axis).**

## Discussion

This paper documents the creation of the first validated Tseltal vocabulary checklist. Based on a checklist first developed by Penelope Brown and Dedre Gentner, we collected data from 84 participants in an iterative development process that resulted in a 251-word checklist. Through this iterative process we were able to develop an

inventory of items that are partially shared with other checklists but also reflect local caregiver reports on Tseltal children's early word production. Following CDI adaptation guidelines, these items are organized into standard CDI semantic categories, but could equally be rearranged and re-sampled to create more linguistically and contextually organized semantic categories for Tseltal children, depending on the relevant research question (e.g., based on shared semantic roots across syntactic classes, like *lo'* /loʔ/ "eat-soft" and *lo'bal* /loʔ.bɑl/ "banana" [literally: eat-soft-NOMINALIZER]). The list and its associated dataset are therefore ready to be used and further developed for both comparative and language-specific work on Tseltal.

The checklist displays reasonable variability within and across categories in age of acquisition, replicates basic patterns of age- and sex-related change, and demonstrates expected over- and under-representation patterns in the checklist's sub-categories given past findings (Bates et al., 1994; Braginsky et al., 2019; Fenson et al., 1994; Frank et al., 2021; Gentner, 2006; Jackson-Maldonado et al., 2024; Kachergis et al., 2023; Marchman & Dale, 2023; Mayor & Plunkett, 2011). In this brief discussion, we review the benefits of creating such an instrument for Tseltal before considering the most important next steps for continued instrument development.

### *Potential Uses of This Instrument*

Vocabulary checklists are multi-purpose tools; we hope that this vocabulary checklist for Tseltal will be of use to a variety of professionals working in the fields of Tseltal (and perhaps more broadly, Mayan) child development. We have here tentatively concluded that the checklist broadly behaves like the vocabulary checklists of other CDI instruments, implying that it is likely useful for generating a lexically grounded snapshot of children's overall language development (Fenson et al., 1994; Frank et al., 2021; Marchman & Dale, 2023). The checklist may therefore, in the future, prove helpful in designing educational materials and identifying delays in linguistic development. We would, however, very strongly warn against using this checklist on its own as the foundation for decisions about intervention or education. Substantial further work that includes longitudinal data, test-retest reliability estimates, external validation, and more, will be required before the instrument can be treated as a reliable clinical tool. Furthermore, adequate application and further development of the instrument, where it concerns clinical treatment of children and clinician training, will necessarily involve the integration of observational data and interviews, among other data sources.

All that being said, these preliminary data collected using the checklist already begin to outline the distinct patterns in lexical development—along with expected sources of variability—that characterize Tseltal language development. These patterns can be leveraged to inspire language learning materials aimed at fortifying indigenous language maintenance efforts. The same patterns can help speak to the human capacity

for language learning at large; similarities and differences in Tseltal lexical development may help us infer the adaptive capacities that underlie language learning across the diverse developmental milieux in which language is acquired.

The realization of this potential will only be achieved if the anonymized data, documentation, and analysis tools are freely and openly available to community members, clinical and educational professionals, and researchers. Sharing these aspects of the project redundantly and making sure they are well and securely disseminated is possible thanks to resources such as the Open Science Framework, the Wordbank archive, GitHub, and the CIESAS website (see Appendix A for links to each resource). Fully open materials will help ensure the healthy further development of this instrument over time.

### *Next Steps*

The analyses in the present paper suggest that the checklist is, basically, functional in its current form. However, there are a number of clear future directions to take to ensure its usefulness and to further establish its validity (see Jarůšková et al., 2023). Regarding usefulness, we still wonder whether we have missed important words. Our ideal items are highly informative as developmental indices and additionally carry information about some linguistic or cultural feature that informs our stories of how Tseltal children develop (in particular) or our theories of human language cognition (in general). One example along these lines would be small lexical sets of spatial terms or kinship terms, which have setting-specific importance and also directly bear on theories of cognition (Bates et al., 1994; Clark & Sengul, 1978; Gentner, 2006; Mitchell & Jordan, 2021). There is also the important matter of ensuring that these items make sense across the major dialects of Tseltal.

In addition to new words and word substitutions, it would be useful to make two further structural changes to the checklist. First, we tried initially to ask about word combinations and errors, but our preliminary adaptation of these questions elicited highly variable response types. It is worth trying again, in a future iteration, to ask about morphosyntactic development. In a language with such a rich inflectional morphology (Polian, 2013), and with many aspects of morphosyntactic development well-captured in observational work (Brown, 1997, 2008; De León, 1999b, 1999a, 2001), there is a clear utility for a quick, rough measure of grammatical development. The second structural change would be to separate the checklist into two instruments: one aimed at younger children (akin to the Words and Gestures CDI instrument) and one aimed at older children (akin to the Words and Sentences CDI instrument). Our present checklist is aimed somewhere between these two traditional checklist populations—from the time just before first words to first word combinations. Our present age sample reflects the current needs of our research team, which is focused on a bigger project concerning lexical development in 12–20-month-olds (note that here we have

allowed data collection up until 24 months). However in the long term it would be useful to have separate instruments as have been used in most other CDI adaptations.

Finally, there is a great deal more we could do to validate the instrument, internally and externally. Future steps should include longitudinal data collection, test-retest reliability measures, independent vocabulary measures, and more. Along with these efforts will come another necessary addition: much more data from many more children. These validation efforts are key to our interpretation—and thus application—of the checklist data. Should the checklist be used for clinical interventions, it will become especially urgent to establish these validity measures, in collaboration with clinicians, educational professionals, participant families, and other stakeholders. To better scale in these circumstances we may also need to consider a compromise between written parental report (the traditional method) and spoken parental interview (our current method). Following our Australian colleagues, we could consider a digital survey that features sound files for each word and an intuitive data-entry interface (O'Shannessy et al., 2024).

### *Conclusion*

We present a preliminary vocabulary checklist for tracking the lexical development of children acquiring Tseltal as their primary language. The checklist displays many of the expected patterns for the vocabulary checklists of instruments based on the MacArthur-Bates CDI. We discuss important avenues for further development in the future.

## References

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., & Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: Communicative Development Inventory (CDI) adaptation for two Kenyan languages. *Journal of Child Language*, *42*(4), 763–785. https://doi.org/10.1017/S0305000914000403

Arunachalam, S., & Waxman, S. R. (2010). Language and conceptual development. *WIREs Cognitive Science*, *1*(4), 548–558. https://doi.org/10.1002/wcs.37

Bates, E., Marchman, V. A., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, *21*(1), 85–123. https://doi.org/10.1017/S0305000900008680

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Bromham, L., Hua, X., Algy, C., & Meakins, F. (2020). Language endangerment: A multidimensional analysis of risk factors. *Journal of Language Evolution, 5*(1), 75–91. https://doi.org/10.1093/jole/lzaa002

Brown, P. (1997). Isolating the CVC root in Tzeltal Mayan: A study of children's first verbs. In E. V. Clark (Ed.), *Proceedings of the 28th annual Child Language Research Forum* (pp. 41–52). CSLI/University of Chicago Press.

Brown, P. (1998a). Children's first verbs in Tzeltal: Evidence for an early verb category. *Linguistics, 36*(4), 713–754. https://doi.org/10.1515/ling.1998.36.4.713

Brown, P. (1998b). Conversational structure and language acquisition: The role of repetition in Tzeltal. *Journal of Linguistic Anthropology, 8*(2), 197–221. https://doi.org/10.1525/jlin.1998.8.2.197

Brown, P. (2008). Verb specificity and argument realization in Tzeltal child language. In M. Bowerman & P. Brown (Eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability* (pp. 167–189). Erlbaum.

Brown, P. (2011). The cultural organization of attention. In *The handbook of language socialization* (pp. 29–55). Wiley Online Library. https://doi.org/10.1002/9781444342901.ch2

Brown, P., & Gaskins, S. (2014). Language acquisition and language socialization. In *Cambridge handbook of linguistic anthropology* (pp. 187–226). Cambridge University Press. https://doi.org/10.1017/CBO9781139342872.010

Brown, P., Gentner, D., & Braun, K. &. (2005). *Initial acquisition of nouns and verbs in Tzeltal speaking children [Paper presentation]*. Tenth International Congress for the Study of Child Language, Berlin, Germany.

Byers-Heinlein, K., Gonzalez-Barrero, A. M., Schott, E., & Killam, H. (2024). Sometimes larger, sometimes smaller: Measuring vocabulary in monolingual and bilingual infants and toddlers. *First Language, 44*(1), 74–95. https://doi.org/10.1177/01427237231204167

CDI Advisory Board. (2015). Adaptations in other languages. Retrieved from https://mb-cdi.stanford.edu/adaptations.html

Clark, E. V., & Sengul, C. J. (1978). Strategies in the acquisition of deixis. *Journal of Child Language, 5*(3), 457–475. https://doi.org/10.1017/S0305000900002099

De León, L. (1999a). Verb roots and caregiver speech in early Tzotzil Mayan acquisition. In B. Box, D. Jurafsky, & L. Michaelis (Eds.), *Cognition, discourse, and function* (pp. 99–119). CSLI/University of Chicago Press.

De León, L. (1999b). Verbs in Tzotzil (Mayan) early syntactic development. *International Journal of Bilingualism, 3*(2-3), 219–240. https://doi.org/10.1177/13670069990030020601

De León, L. (2001). Finding the richest path: The role of language and cognition in the acquisition of verticality in Tzotzil. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 544–565). Cambridge University Press. https://doi.org/10.1017/CBO9780511620669.020

deMayo, B. E., Kellier, D., Braginsky, M., Bergmann, C., Hendriks, C., Rowland, C. F., Frank, M. C., & Marchman, V. (2021). Web-CDI: A system for online administration of the MacArthur-ates communicative development inventories. *Language Development Research, 1*(1), 55–98. https://doi.org/10.34842/kr8e-w591

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), Serial no. 242, i–185. https://doi.org/10.2307/1166093

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank project*. MIT Press. https://doi.org/10.7551/mitpress/11577.001.0001

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544–564). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.003.0022

Hamadani, J. D., Baker-Henningham, H., Tofail, F., Mehrin, F., Huda, S. N., & Grantham-McGregor, S. M. (2010). Validity and reliability of mothers' reports of language development in 1-year-old children in a large-scale survey in Bangladesh. *Food and Nutrition Bulletin, 31*(2), S198–S206. https://doi.org/10.1177/15648265100312S212

Hellwig, B., Defina, R., Kidd, E., Allen, S. E. M., Davidson, L., & Kelly, B. F. (2021). Child language documentation: The sketch acquisition project. In G. Haig, S. Schnell, & F. Seifart (Eds.), *Doing corpus-based typology with spoken language data: State of the art* (pp. 29–58). University of Hawai'i Press.

Información Cultural, R. N. de. (2020). Tseltal. Gobierno de México. Retrieved from https://sic.gob.mx/ficha.php?table=inali_li&table_id=50

Jackson-Maldonado, D., Friend, M., Marchman, V. A., Weisleder, A., Auza, A., Conboy, B., Rubio-Codina, M., & Dale, P. S. (2024). The MacArthur Inventario del Desarrollo de Habilidades Comunicativas III: A measure of language development in Spanish-speaking two- to four-year-olds. *Language Development Research, 4*(1), 361-398. https://doi.org/10.34842/12fb-qz93

Jarůšková, L., Smolík, F., Chládková, K., Oceláková, Z., & Paillereau, N. (2023). How to build a communicative development inventory: Insights from 43 adaptations. *Journal of Speech, Language, and Hearing Research, 66*(6), 2095–2117. https://doi.org/10.1044/2023_JSLHR-22-00591

Kachergis, G., Francis, N., & Frank, M. C. (2023). Estimating demographic bias on tests of children's early vocabulary. *Topics in Cognitive Science, 15*(2), 303–314. https://doi.org/10.1111/tops.12635

Kelly, B., Wigglesworth, G., Nordlinger, R., & Blythe, J. (2014). The acquisition of polysynthetic languages. *Language and Linguistics Compass, 8*(2), 51–64. https://doi.org/10.1111/lnc3.12062

Kik, A., Adamec, M., Aikhenvald, A. Y., Bajzekova, J., Baro, N., Bowern, C., ... Novotny, V. (2021). Language and ethnobiological skills decline precipitously in Papua New Guinea, the world's most linguistically diverse nation. *Proceedings of the National Academy of Sciences, 118*(22), e2100096118. https://doi.org/10.1073/pnas.2100096118

Maffi, L. (2005). Linguistic, cultural, and biological diversity. *Annual Review of Anthropology, 34*(1), 599–617. https://doi.org/10.1146/annurev.anthro.34.081804.120437

Marchman, V. A., & Dale, P. S. (2023). The MacArthur-Bates communicative development inventories: Updates from the CDI advisory board. *Frontiers in Psychology, 14*, 1170303. https://doi.org/10.3389/fpsyg.2023.1170303

Marchman, V. A., Dale, P. S., & Fenson, L. (2023). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual, 3rd edition*. Brookes. https://doi.org/10.3389/fpsyg.2023.1170303

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science, 14*(4), 769–785. https://doi.org/10.1111/j.1467-7687.2010.01024.x

Mitchell, A., & Jordan, F. M. (2021). The ontogeny of kinship categorization. *Journal of Cognition and Culture, 21*(1-2), 152–177. https://doi.org/10.1163/15685373-12340101

O'Shannessy, C., Davis, V., Foster, D., Bartlett, J., & Nelson, A. (2024). Development of a CDI for multilingual children speaking Indigenous languages in Central Australia. International Association for the Study of Child Language (Prague, Czechia). Retrieved from https://little-kids-learning-languages.net/

Polian, G. (2013). *Gramática del Tseltal de Oxchuc [Grammar of Oxchuk Tseltal]*. Centro de Investigaciones y Estudios Superiores en Antropología Social.

Polian, G. (2017). Tseltal and Tsotsil. In J. Aissen, N. England, & R. Zavala Maldonado (Eds.), *The Mayan languages* (pp. 610–647). Routledge. https://doi.org/10.4324/9781315192345-22

Prado, E. L., Phuka, J., Ocansey, E., Maleta, K., Ashorn, P., Ashorn, U., … Dewey, K. G. (2018). A method to develop vocabulary checklists in new languages and their validity to assess early language development. *Journal of Health, Population and Nutrition, 37*(13), 1–10. https://doi.org/10.1186/s41043-018-0145-1

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics, 54*(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Southwood, F., White, M. J., Brookes, H., Pascoe, M., Ndhambi, M., Yalala, S., … Alcock, N. (2021). Sociocultural factors affecting vocabulary development in young South African children. *Frontiers in Psychology, 12*, 642315. https://doi.org/10.3389/fpsyg.2021.642315

Tan, A. W. M., Marchman, V. A., & Frank, M. C. (2024). The role of translation equivalents in bilingual word learning. *Developmental Science, 27*(4), e13476. https://doi.org/10.1111/desc.13476

Vogt, P., Mastin, J. D., & Aussems, S. (2015). Early vocabulary development in rural and urban Mozambique. *Child Development Research, 1*, 89195. https://doi.org/10.1155/2015/189195

Weber, A. M., Marchman, V. A., Diop, Y., & Fernald, A. (2018). Validity of caregiver-report measures of language skill for Wolof-learning infants and toddlers living in rural African villages. *Journal of Child Language, 45*(4), 939–958. https://doi.org/10.1017/S0305000917000605

## Data, Code and Materials Availability Statement

All data, code, and materials are available from https://osf.io/g2spx/.

## Ethics Statement

Ethics approval was obtained from the IRB of the University of Chicago (IRB23-0244). All participants gave informed verbal consent before taking part in the study, as described in the Procedure subsection of the Methods section.

## Authorship and Contributorship Statement

**Penelope Brown** created the first ever Tseltal-adapted checklist, prior to this article. Collectively, **Humbertina Gómez Pérez**, **Juan Méndez Girón**, **Ruthe Foushee**, **Gilles Polian**, and **Marisa Casillas** iteratively revised the checklist. **Humbertina Gómez Pérez**, **Ruthe Foushee**, and **Marisa Casillas** collected the data, aided by those mentioned in the Acknowledgements. **Kennedy Casey** and **Marisa Casillas** digitally pre-processed the data. **Kennedy Casey**, **Ruthe Foushee**, and **Marisa Casillas** conducted and checked the analyses. **Marisa Casillas** wrote the first draft of the analyses. All authors commented on revisions of the manuscript thereafter. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## Acknowledgements

This work was supported by an NSF CAREER award to MC (2238609). We are indebted to Sebastiana Velasco Gutierrez, Micaela Velasco Gutierrez, Antonieta López Santiz, and the participating families who contributed to this dataset. Special thanks go to Mika Braginsky, Mike Frank, Adriana Weisleder, and Virginia Marchman for advice on analyses. Special thanks, too, to Philip Dale for advice on describing our checklist in relation to the CDI. We are deeply grateful to Dedre Gentner and Penny Brown for sharing their original adapted Tseltal checklist. This work was partly presented at IASCL and BUCLD 2024.

## Appendix A: Links to Further Materials

Find the current and archived checklist, instructions, anonymized data, scripts, and more at one of the links below.

1) Open Science Framework (OSF) repositories:
   a) https://osf.io/z8hdk contains…
      i) **Method preregistration:** the pre-registered methods for developing the Tseltal checklist reported in this paper (up to version 6).
   b) https://osf.io/g2spx/ contains versioned archives of…
      i) **Data:** the anonymized data and participant for the children's 84 checklists reported on in the current paper,
      ii) **Scripts:** all associated analysis scripts (including what is required to reproduce this document),
      iii) **Checklists:** blank printable and editable copies of the most current checklist version(s) (in pdf, docx, and xlsx),
      iv) **Instructions:** blank printable and editable copies of the most current checklist instructions (in pdf and docx),
      v) **Consent example:** a pdf copy of the consent page we used for verbal consent in the currently collected data,
      vi) **Study metadata:** a general description of the study and reference to this paper for more details,
      vii) **Contact information:** up-to-date contact information for those who have follow-up questions
2) The CIESAS website (https://sureste.ciesas.edu.mx/polian-gilles/) contains archives of…
   a) **Checklists:** blank printable and editable copies of the most current checklist version(s) (in pdf, docx, and xlsx),
   b) **Instructions:** blank printable and editable copies of the most current checklist instructions (in pdf and docx),
   c) **Consent example:** a pdf copy of the consent page we used for verbal consent in the currently collected data,
   d) **Study metadata:** a general description of the study and reference to this paper for more details,
   e) **Contact information:** up-to-date contact information for those who have follow-up questions
3) The WordBank repository (https://wordbank.stanford.edu/) contains…
   a) **Data:** the anonymized data and participant for the children's 84 checklists reported on in the current paper,
   b) **Study metadata:** a general description of the study and reference to this paper for more details,

**Contact information:** up-to-date contact information for those who have follow-up question

## Appendix B: Alternative Age of Acquisition Estimates

We were unable to establish an AoA based on proportional production (≥ 0.5 production) for 11.9% ($N = 30$) of the items on the checklist. As one reviewer pointed out, we can alternatively use binomial regression to estimate age of acquisition for all the items on our checklist, including the 11.9% that yielded no AoA in the current sample. We ran a logistic mixed-effects regression of whether or not a child produced an item (1/0) that included a fixed effect of child age in months (numeric) and a random effect of checklist item (factor). We then used the ggeffects package in R (Lüdecke, 2018) to estimate an AoA for each item. Below we plot the AoA distributions, which range from 5 months (unrealistic) to 31 months, with AoAs for most words sitting between 12 and 30 months. Peak acquisition rates for this list were estimated to be between 18 and 24 months.

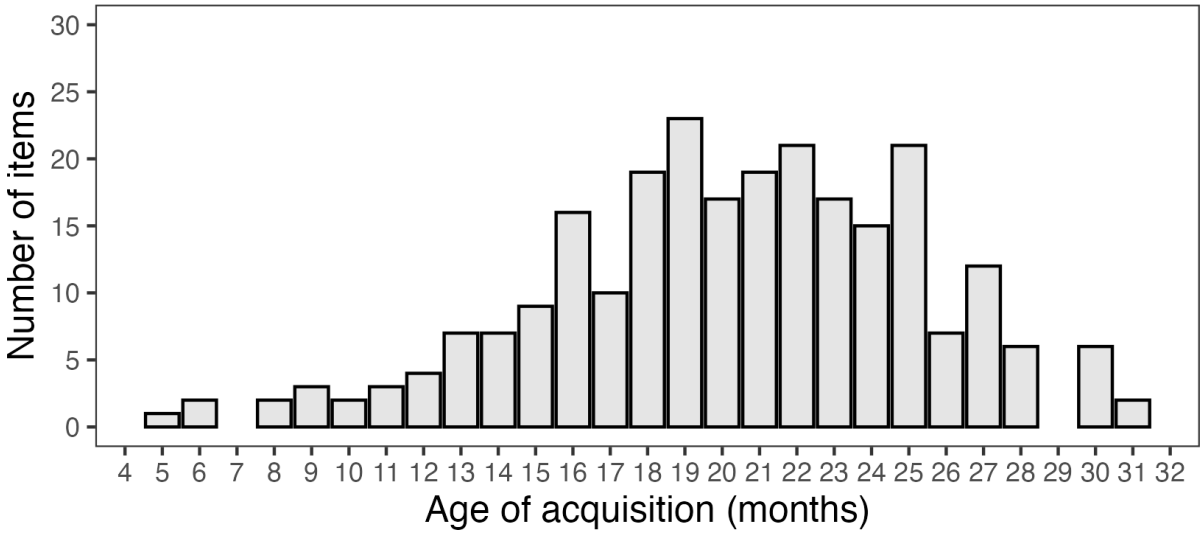


***Figure SM1. Distribution of model-estimated age of acquisition—the first age with ≥ 50% of a sample estimated to produce the item—over all items.***

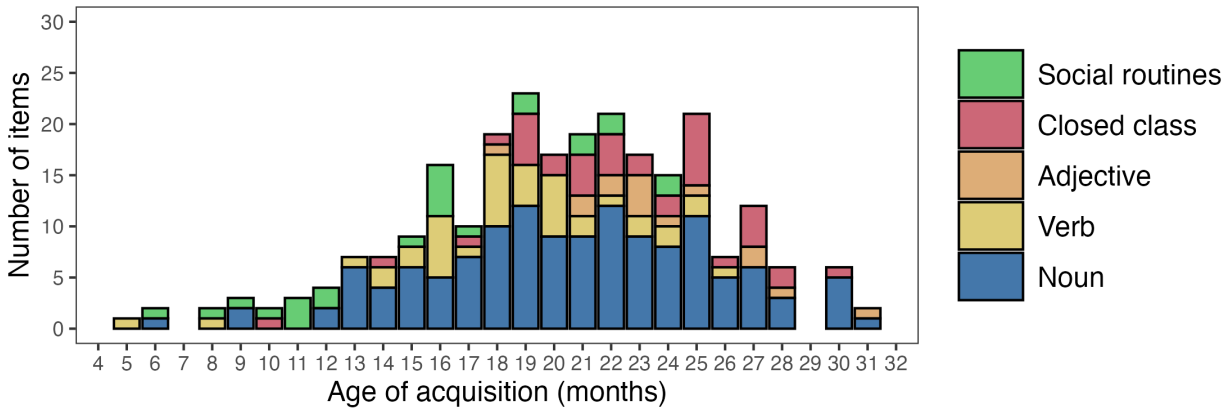


***Figure SM2. Distribution of model-estimated age of acquisition over all items, separated by major syntactic types (Noun, Verb, Adjective, Closed class, Social routines).***

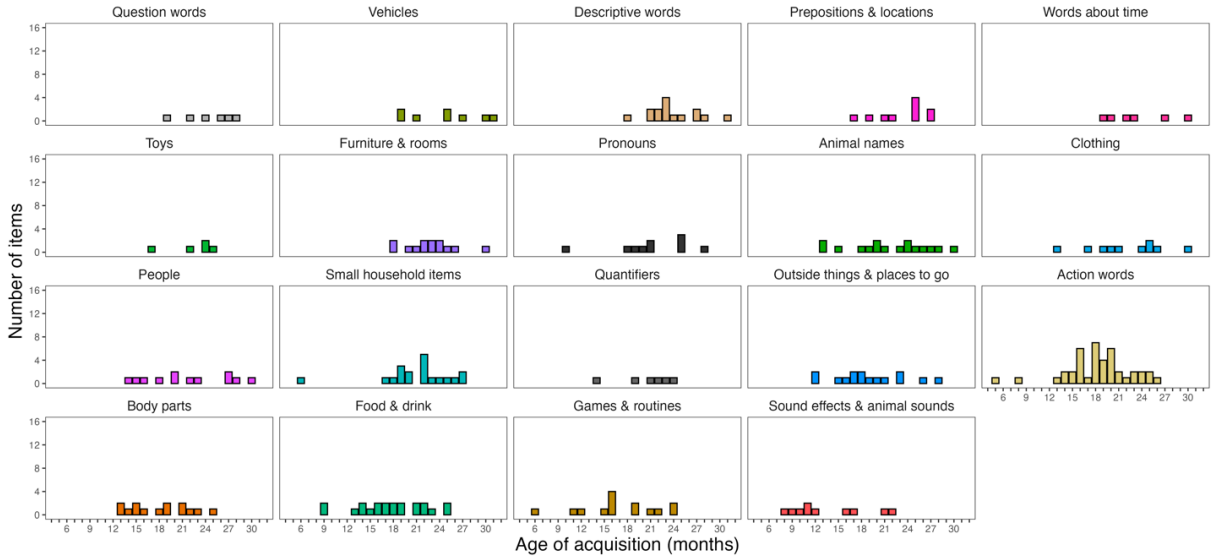


***Figure SM3. Distribution of model-estimated age of acquisition over all items within each content sub-category of the list, sorted from the latest acquired (upper left; Question words) to the earliest acquired (lower right; Sound effects & animal sounds).***

## Appendix C: Data Pre-Processing and Marking of Alternatives

Participant responses were recorded by an experimenter in real time, either digitally or on paper (see Methods for details). These individual responses were then copied over (by MC or KC) to a single spreadsheet that we use for analysis. We passed through

the spreadsheet many times to identify potential errors that emerged during data recording or data transfer. Experimenters are moving fast through the list during the interview, and so—while the vast majority of responses are sensible and interpretable—there were occasional inconsistencies in how responses were recorded. First, and most commonly, an onomatopoeic form was accepted in place of its associated noun label (e.g., "moo" for "cow"). We trained experimenters to verbally accept these responses verbally but to not write them down, considering that these onomatopoeic forms were already captured in the Sounds category. Second, mothers reported children producing a wide variety of alternative referring forms children use in place of the list item we asked about (e.g., "darkness" for "night"). To address these inconsistencies adequately, we created a marking system:

- **Canonical:** word forms in Tseltal and/or relating directly to the standard forms offered in the list (which sometimes come from Spanish). These *canonical forms* are spelled orthographically in the coding, with no additional mark up.
- **In-language alternative:** productions recognizable as pragmatically and semantically appropriate equivalent forms in Tseltal—but differing from the standard, expected item word form—were accepted if they could feasibly be the dominant way of referring to this concept in that child's family. These *in-language alternative forms* are spelled orthographically in the coding, and are enclosed in a single pair of parentheses.
- **Other-language alternative:** productions in Spanish or another language that were recognizable as pragmatically and semantically appropriate equivalent forms in Tseltal—but differing from the standard, expected item word form—were accepted if they could feasibly be the dominant way of referring to this concept in that child's family. These *other-language alternative forms* are spelled orthographically in the coding, and are enclosed in a double pair of parentheses. Note that many of the standard list items are shared directly with Spanish; those items are considered "canonical" Tseltal productions, since they represent expected borrowings.
- **Excluded:** productions that did not satisfy the research aim of identifying children's recognizable target wordforms for the items on the list. These most often included onomatopoeia as substitutes for target items (e.g., "moo" for "cow") but also included non-adult-like over-extensions (e.g., "car" for "taxi") and the production of an associated word in place of the target word (e.g., "bite" for "snake"). These *excluded forms* are spelled orthographically in the coding, and are enclosed in a single pair of square brackets. We note that, while the decision to exclude these responses may under-count some children's productive vocabularies (if, e.g., other researchers find onomatopoeic substitutes acceptable), they make up 0.66% (less than 1%) of the checklist responses we gathered.

Three other formatting decisions were made in conducting digital data entry that facilitate the use of this parenthesis-based coding and the identification of unique forms per list item:

- multi-word responses are separated by underscores (e.g., “ya_xban”)
- multi-alternative responses are ordered as follows: canonical > (in-language alternative) > ((other-language alternative)) > [excluded]
- recorded forms that *only* varied based on non-phonological, non-meaningful variation in continuant length were collapsed into a single form (e.g., for the sound a car engine makes: “rr”, “rrrr”, “rrr” were all converted to “rrrr”)

***References Cited***

Lüdecke, D. (2018). Ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

## License