1   Early language experience in a Papuan village

2   Marisa Casillas[1], Penelope Brown[1], & Stephen C. Levinson[1]

3   [1] Max Planck Institute for Psycholinguistics

4   Author Note

5   Correspondence concerning this article should be addressed to Marisa Casillas, P.O.

6   Box 310, 6500 AH Nijmegen, The Netherlands. E-mail: Marisa.Casillas@mpi.nl

7   Abstract

8   To be completed later.

9   *Keywords:* Child-directed speech, linguistic input, non-WEIRD, vocal maturity, turn

10   taking, interaction, Papuan

11   Word count: XXXXX (XXXX not including references)

Early language experience in a Papuan village

Introduction

In their first five years of life, children hear an extraordinary amount of language in a wide variety of interactional contexts. Tracking the distribution and characteristics of this linguistic input over the day, across age, and between children is a difficult task. Until recently, developmental language science has relied on short video recordings of caregiver-child interaction, at home or in the lab, to get a grasp on what kinds of language children typically hear. This has been a fruitful approach in teasing out individual and group-based differences in interactional style (REFS). However, short recordings are limited in their insight because they represent only a small slice of the child's language abilities and experiences (REFS).

Improved recording hardware and advances in speech technology have recently allowed us to use daylong recordings to get a peek into children's broader language landscapes. Daylong recordings are made with a device, usually positioned on the target child's chest, while that child freely navigates their social environment for most of a waking day (REFS). This style of audio recording has allowed researchers to track children's verbal language use across a range of activity and interlocutor contexts, yielding more representative and generalizable measures of their language environments (REFS). While daylong recording collections are typically too large for comprehensive transcription and annotation, a combination of automated tools, (REFS) sampling techniques (REFS), and standardized annotation approaches (REFS) can lead to rich, but efficiently-gained glimpses into the at-home language environment. However, properly collecting, processing, and archiving daylong data is not easily achieved and may not be well suited for a range of research questions (REFS). At time of writing, there are few options for capturing visual information across the day (but see REFS), limiting this method primarily to acoustic phenomena

37   (REFS).

38   Daylong recording methods are still relatively new, and their reliability and predictive

39   value for language development have not yet been fully established. For example, one

40   collection of recordings made in the US Northwest suggests that there is so much variability

41   across activities and days in basic talk characteristics (e.g., how much speech comes from

42   what types of speakers) that researchers need several days of recordings before they can

43   expect their input estimates to stabilize (Anderson & Fausey, in prep). Even if one can

44   achieve a reliable estimate of a language environment measure (e.g., overhearable adult

45   words per hour), how and why that estimate relates to deeper factors shaping the learning

46   situation, including caregiving ideologies and language outcomes, is often indirect at best.

47   Relatedly, meaningful differences between individual children may be minimized when

48   averaging across the entirety of the day's high and low moments; it may well be that a few

49   key interactions throughout the day provide sharper resolution on individual and

50   group-based differences compared to whole-day averages.

51   Two recent studies have directly investigated the effect of recording duration on

52   caregiver speech, finding that short recordings display denser, somewhat different input than

53   what is found in longer recordings (REFS). Bergelson and colleagues (REFS) analyzed the

54   contexts of noun use encountered by 44 6- and 7-month-old children in the US in both

55   hour-long at-home videos and daylong recordings. The hour-long video differed from the

56   day-long average in the density of noun input, the utterances in which nouns were embedded

57   (e.g., questions vs. declaratives), and the actual nouns used. When the authors compared the

58   hour-long video to the "peak" hour of talk in the day, they found a similar density of noun

59   input across the recording types, but the other differences remained. This difference in input

60   density also appears in short recordings of varying length with US children (REFS).

61   Importantly, however, the speech characteristics heard by individual children in a sample

62   correlates across recording types (i.e., their relative rank for a given measure is stable;

REFS). Based on these findings, one could infer that at-home short recordings are influenced by some (but not all) of the same underlying factors that drive language patterns during daylong recordings (e.g., caregiver ideologies about child development, child responsiveness, household composition).

Studies of children growing up in two indigenous Mayan communities of Southern Mexico (Tseltal and Yucatec Mayan) suggest that short and long recordings may yield substantial differences in how the speech environment is characterized (REFS). Previous studies on these communities have tended to use ethnographic and microanalytic analyses of short interactions to examine the character of children's speech environments. They have found that caregivers shape infants' and young children's worlds such that the children learn to attend to what is going on around them rather than expecting to be the center of attention (REFS). Consistent with this goal, direct talk to infants, particularly from adults, is rare until children themselves begin to elicit responses from others (REFS). Because young children are often cared for by older siblings and cousins, a substantial portion of talk to young children was also expected to come from other children (REFS). Similar observations have been reported for multiple other distinct (but ethnolinguistically related) communities in the region (REFS). Following up on this ethnographic work, Shneidman (REFS) used short videos of interaction to conduct a quantitative, longitudinal study of the speech young Yucatec children heard. She found that interactional patterns aligned well with observations in previous work in that community: infants were rarely spoken to at first, but their language input increased enormously with age, mostly due to an influx of speech from other children (REFS). However, when Casillas and colleagues (REFS) used daylong recordings with a Tseltal Mayan community, where a similar caregiver interactional style has been described previously on the basis of short videos, the pattern of findings diverged from expectations. In brief, they found that infants and young children were indeed spoken to rarely, but that there was no increase in speech input with age and the majority of speech came from adult women, even when children were old enough to independently follow their

90 older siblings and cousins around the house. These divergent results betweeh daylong and

91 short video recordings don't imply that the latter is wrong, only that it is not representative

92 with respect to the child's language experience over an entire day.

93 These findings do raise an important issue faced by developmental psychology as it

94 continues to expand the study of child language to more diverse speech communities: when

95 researchers are not members of the community they are studying, it is difficult to know a

96 priori what is typical, representative, or meaningful in children's language experience. By

97 observing as much speech as possible in a context as ecologically valid as possible and by

98 sampling, annotating, and analyzing the data on the basis of the most established

99 development measures we have, researchers using daylong recordings might hope to approach

100 this issue without first needing to conduct deep ethnographic studies in the community on

101 caregiving practices and ideologies around language use and language development (REFS).

102 When studying members of our own cultural group, we can bridge between simple, observable

103 behaviors and rich interpretations of, thereby expanding our explanatory model beyond the

104 measures directly analyzed (e.g., why child-directed talk might relate to faster vocabulary

105 development). We cannot hope to gain such enriched understandings cross-culturally

106 without ethnographic work; and in the absence of such work we must accept that there may

107 be a dissociation between how we have traditionally understood an operationalized language

108 behavior (e.g., child-directed speech) and what drives the use and form of that behavior in a

109 given community or interactional context (e.g., pedagogical concerns, entertainment of the

110 caregiver, getting the child to assist). Until there are trained researchers working on this

111 topic who were born and raised as members of these communities (what we should be trying

112 to cultivate for the longer term) this is a quandary we will continue to face.

113 Pairing ethnographic work with broader-scope studies of children's language

114 environments may be the most fruitful way to ensure that their speech environments and

115 speech development are captured well enough to propose and test meaningful theories

116 cross-culturally. These two methods have complementary roles to play in exploring the

117 landscape of at-home language, and neither should be taken to reflect the "true" language

118 input for a given child; after all, in the example of Tseltal above, many interactions with

119 infants during the daylong recordings came during moments where visitors using a video

120 camera, or even other community members, would not typically be invited (e.g., after the

121 parent was roused by the child, who was waking from an afternoon sleep). If we want to

122 encourage more work on small-scale and/or understudied language learning contexts, it will

123 be important to continue establishing how different methods of measuring the input impact

124 the conclusions that are likely to be made.

125 In this study we present analyses of daylong recordings from a small-scale indigenous

126 community in which prior ethnographic work has painted a clear picture of early

127 caregiver-child interaction, different from what we have seen in studies of Mayan caregivers:

128 child-centric, face-to-face interaction from the first days of infancy on Rossel Island, Papua

129 New Guinea (PNG). In what follows we will review the ethnographic work done with this

130 community previously, our methods for following up on these findings with daylong

131 recordings, and a discussion of the differences that arose. This study was completed as part

132 of a larger comparative project focusing on children's speech environments and linguistic

133 development at two sites: the Tseltal Mayan community mentioned above (Casillas et al.,

134 forthcoming) and this Rossel Island community. Therefore all methods for annotation and

135 analysis in this study parallel those reported elsewhere for Tseltal Mayan children's speech

136 environments (Casillas, Brown, & Levinson, forthcoming).

137 **The community**

138 Yélî Dnye is a language spoken by approximately 5000 people, nearly all of whom reside

139 on Rossel Island, a remote island 250 nautical miles off the mainland coast in Milne Bay

140 Province, Papua New Guinea. While all the neighboring languages fall into the Austronesian

141 family, Rossel is a presumed-Papuan isolate that features a phonological inventory and set of

142 grammatical features that are completely unattested in other languages of the region. Partly

143 due to its remoteness, most children on Rossel Island grow up speaking it monolingually at

144 home, only beginning to learn English (the official lingua franca of Milne Bay Province) as

145 they progress through school, which typically begins when a child is 7 or 8 years old.

146 We were interested to investigate the language environment of children acquiring Yélî

147 Dnye because prior ethnographic work had suggested that child-directed speech is highly

148 frequent in this community, from mothers and other adult caregivers, but also from other

149 children. Therefore we were interested in understanding how children's input environment

150 influenced their acquisition of this language with all its rare structures.

151 However, to our surprise we found that Yélî children were not spoken to very often at

152 all. In fact, they were spoken to less often than the Tseltal children we have studied in other

153 work, who are growing up in a community where children are indeed reportedly spoken to

154 infrequently.

Our predictions

156 Based on prior ethnographic work, we expected to see that young children and infants

157 are spoken to relatively frequently—on par with Western children, and certainly more often

158 than we saw with Tseltal children, where estabished work from a number of related

159 ethnolinguistic groups suggests a non-child-centric approach to early language socialization.

160 Because of the similar subsistence farming lifestyle we expected that, like Tseltales, Rossel

161 Island children would show strong time-of-day effects. Similarly, because children in both

162 sites are growing up in large households, we expected to see lots of ODS and, consistent with

163 both Tseltal and NA data, a decrease in ODS with age. We predicted an increasing

164 proportion of speech from other children in Tseltal but didn't find it; ethnographic

165 predictions on Rossel would say the same... so we expect it again. Do we expect an increase
166 in CDS with age??

Method

168   This study was completed as part of a larger comparative project focusing on children's
169 speech environments and linguistic development at two sites: a Tseltal Mayan community in
170 southern Mexico and this Rossel Island community. Therefore all methods for annotation
171 and analysis in this study parallel those reported elsewhere for Tseltal Mayan children's
172 speech environments (Casillas, Brown, & Levinson, forthcoming).

173   The data we present come from 7–9-hour recordings of a waking day at home for the
174 child. Children wore the recording device, which was an elastic vest containing a small stereo
175 audio recorder (Olympus WS-832 or WS-853) and a miniature camera that captured photos
176 of the child's frontal view at a fixed interval (every 15 seconds; Narrative Clip 1). The
177 camera was additionally outfitted with a fisheye lens that, while distorting the images,
178 allowed us to capture 180 degrees of children's frontal view. Because the camera and
179 recorder are separate devices, they were synchronized using an external wristwatch that was
180 used to record the current time at start of recording on each device individually, with
181 accuracy down to the second (photographed by the camera and spoken into the recorder).
182 The photos are timestamped by the camera such that the precise intervals between photos
183 are captured. These timestamps can be used with the cross-device time synchronization cue
184 to create photo-linked audio files of each recording, which we then format as video files (see
185 REFS-URL for scripts and more information).

186   We chose to collect multi-hour recordings of children's language experience at home in
187 order to capture a range of different activity contexts and interlocutors at different times of
188 day (REFS). Previous work investigating the trade-offs of short- versus long-format

recordings of parental speech have demonstrated that the apparent quantity of speech children hear and some of the characteristics of that speech differs depending on recording duration (day-by-day tamislemonda REFS). That said, short recordings often have the benefit of video data, which enables analysts to take visual information into account in interpreting the communicative behaviors captured. Those using daylong recording methods instead have traditionally had to sacrifice this visual context because of (current) technological limitations; there are no miniature, lightweight (e.g., 400g; 5cm x 5cm or less) video recorders on the market that can record for 7–16 hours. We aimed to generate generalizable baseline estimates of how much speech children hear in this community but wanted to maintain visual information for later transcription and interpretation, leading to us develop this novel daylong recording method (see also Abels REFS).

This set-up was used to create daylong recordings of 57 children under age 4;0 on Rossel Island in 2016 (Casillas et al HB), from which we selected 10 representative children between ages 0;0 and 3;0 for transcription and analysis in the current study. The 10 children were selected to be spread between the target age range (0;0–3;0) while also representing a range of typical maternal education levels found in the community and being evenly split between male and female children (see also ACLEW REFS). For each child we then selected a series of non-overalapping sub-clips from the day for transription in the following order: nine randomly-selected 2.5-minute clips, five manually-selected "peak" turn-taking activity 1-minute clips, five manually-selected "peak" vocal activity 1-minute clips, and one manually-selected 5-minute expansion of the best one-minute clip, for a total of 37.5 minutes of transcribed audio for each child (6.25 audio hours in total). The criteria for manual clip selection are identical to those described for the parallel study on Tseltal by Casillas and colleagues (forthcoming).

We were limited to selecting sub-clips from 10 children for analysis because of the time-intensive nature of transcribing these naturalistic data; 1 minute of audio typically took

us approximately 60–70 minutes to be segmented into utterances, transcribed, annotated, and loosely translated into English (~400 hours total). Given that Yélî Dnye is nearly exclusively spoken on Rossel Island, where there is no electricity and unreliable access to mobile data, transcription could only be completed over the course of three 4–6 week visits by our research group to the island in 2016, 2018, and 2019.

We used the ACLEW Annotation Scheme (REFS) in ELAN (ELAN REFS) to transcribe and annotate all hearable speech—both near and distant—in the clips. We first segmented out the utterances and ascribed them to individual speakers (e.g., older brother, mother, aunt, etc.). We then annotated the vocal maturity of each utterance produced by the target child (non-canonical babble/canonical babble/single word/multi-word/unsure) and annotated the addressee of all speech from other speakers (addressed to the target child/one or more other children/one or more adults/a mix of adults and children/any animal/other/unsure). Transcription and annotation was done together by the first author and one of three community members (all native speakers of Yélî Dnye). The community-based research assistants personally knew all the families in the recordings, and were able to use their own experience, the discourse context, and information from the accompanying photos in reporting what was said and to whom speech was addressed for each utterance. Detailed manuals and self-guided training materials, including a "gold standard test" for this annotation scheme can be found at URL (REFS).

pdf

2

pdf

2

pdf

2

Table 1

Demographic overview of the 10 children whose recordings are sampled in the current study, including from left to right: child's age (years;months.days); child's sex (M/F); mother's age (years); level of maternal education (none/primary/secondary/preparatory/university); and the number of people living in the child's household.

Age	Sex	Mother's age	Level of maternal education	People in household
01m;09d	F	31	secondary	8
16m;29d	M	25	secondary	6
20m;03d	F	33	primary	9
35m;29d	M	41	primary	8
12m;29d	F	30	primary	9
04m;13d	M	24	preparatory	5
25m;22d	F	21	secondary	4
03m;19d	M	37	primary	9
09m;03d	F	29	secondary	5
07m;18d	M	24	secondary	5

```
## pdf
##   2
```

```
## [1] 3.13
```

```
## [1] 2.95
```
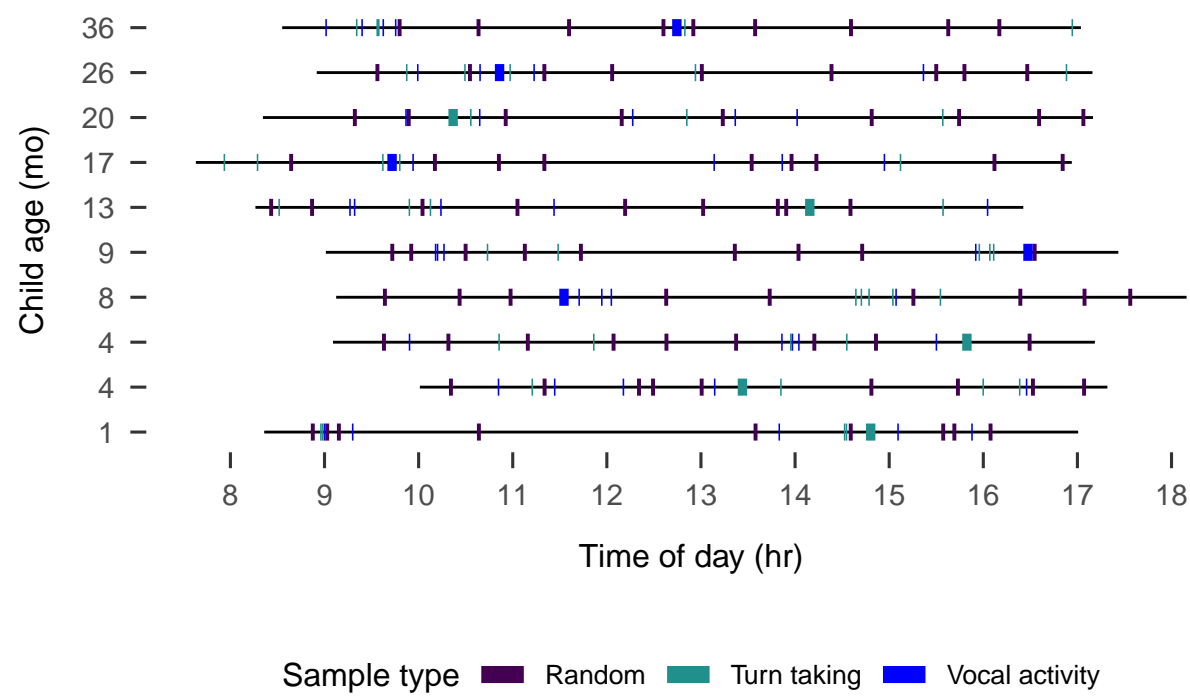
```
## [1] 1.58
```

```
## [1] 6.26
```

Figure 1. Recording duration (black line) and sampled clips (colored boxes) for each of the 10 recordings analyzed, sorted by child age in months.

```
## [1] 14.49
```

```
## [1] 15.07
```

```
## [1] 9.54
```

```
## [1] 18.73
```

```
## [1] 73.32
```

```
## [1] 78.84
```
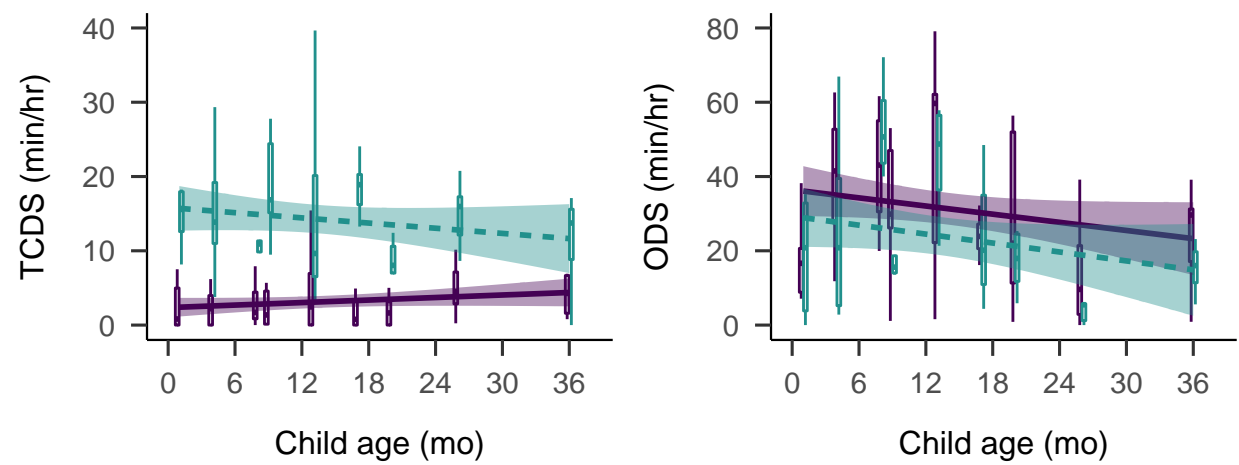
```
## [1] 41.41
```

```
## [1] 100
```

```
## [1] 35.9
```

Figure 2. Estimates of TCDS min/hr (left) and ODS min/hr (right) across the sampled age range. Each box plot summarizes the data for one child from the randomly sampled clips (purple; solid) or the turn taking clips (green; dashed). Bands on the linear trends show 95% confidence intervals.

[1] 32.37

[1] 20.2

[1] 53.78

[1] 28.81

[1] 21.22

[1] 6.68

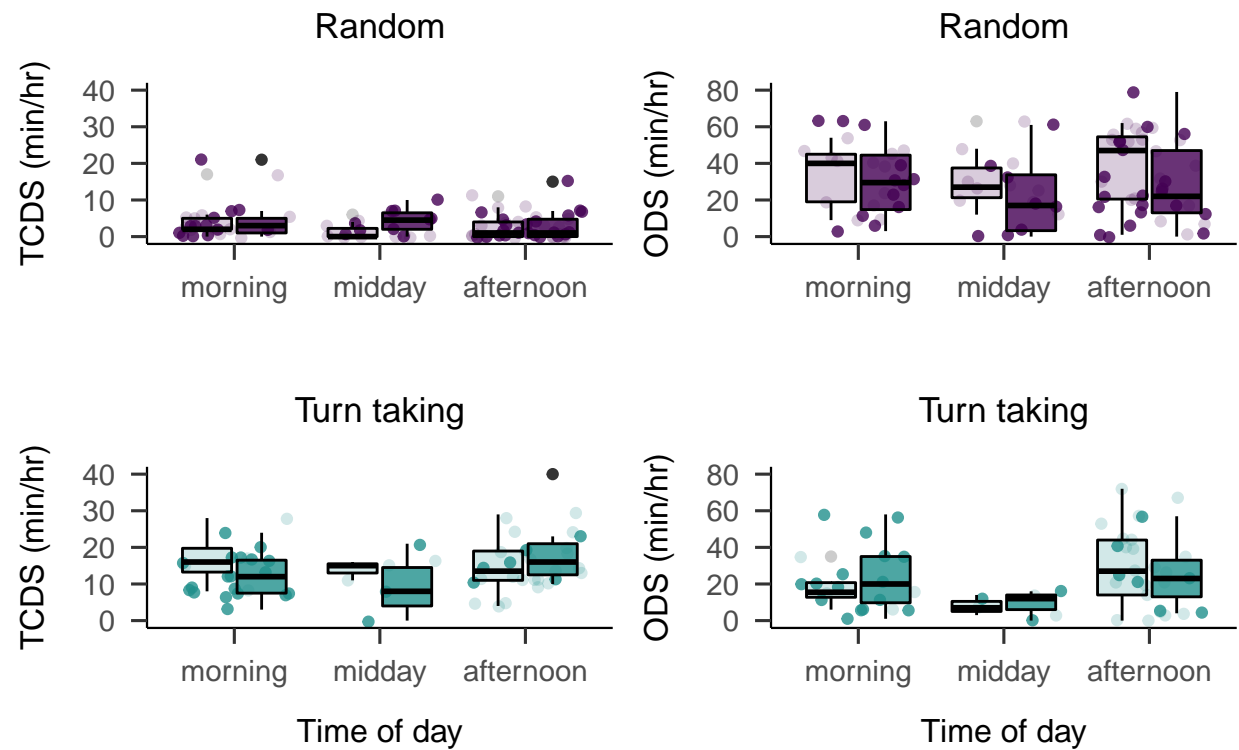## [1] 60.18

Vocal maturity

pdf
2

Figure 3. Estimates of TCDS min/hr (left panels) and ODS min/hr (right panels) across the recorded day in the random clips (top panels) and turn-taking (bottom panels) clips. Each box plot summarizes the data for children age 1;0 and younger (light) or age 1;0 and older (dark) at the given time of day.

```
## pdf
##   2

## pdf
##   2
```

Acknowledgements

This paper was written using the papaja library in RStudio (Aust & Barth, 2018).
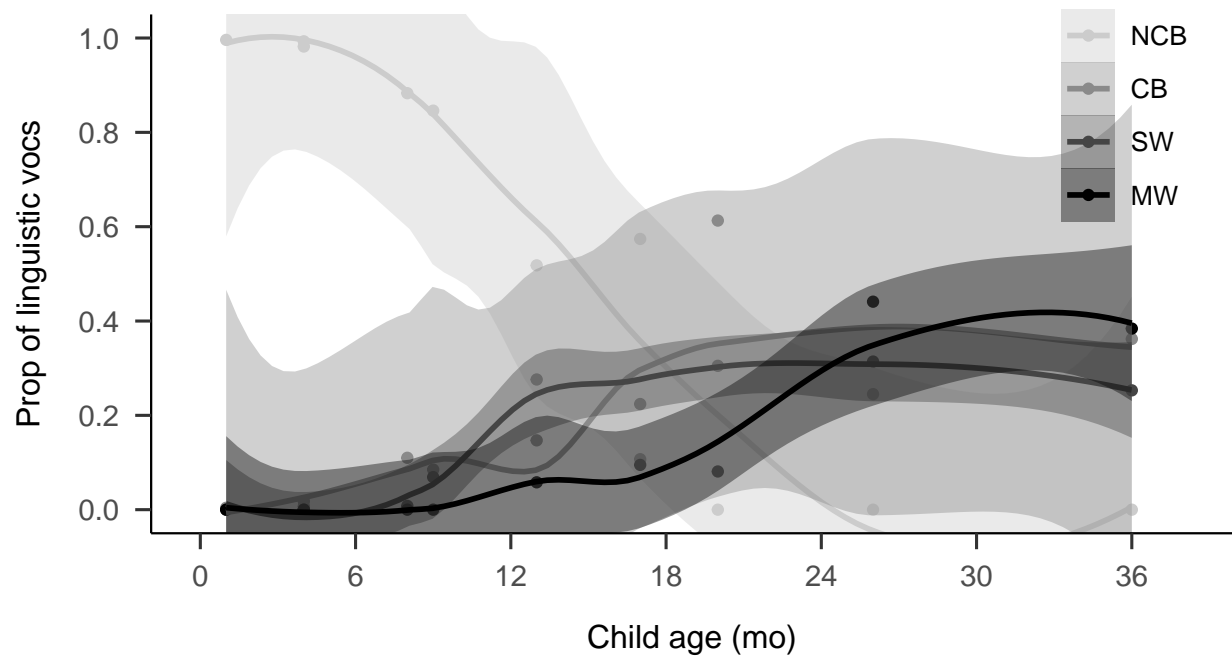
Figure 4. Proportion of vocalization types used by children across age (NCB = Non-canonical babble, CB = Canonical babble, SW = single word utterance, MW = multi-word utterance).

References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja