What Do North American Babies Hear? A large-scale cross-corpus analysis: Supporting Online Information

Elika Bergelson*†[1], Marisa Casillas*‡[2], Melanie Soderstrom[3], Amanda Seidl[4], Anne S. Warlaumont[5] and Andrei Amatuni[1]

[1]Psychology and Neuroscience, Duke University, USA

[2]Max Planck Institute for Psycholinguistics, The Netherlands

[3]Psychology, University of Manitoba, Canada

[4]Speech, Language, and Hearing Sciences, Purdue University, USA

[5]Communication, University of California, Los Angeles, USA

.

*Joint first authors

†elika.bergelson@duke.edu

‡marisa.casillas@mpi.nl

What Do North American Babies Hear? A large-scale cross-corpus analysis: Supporting Online Information

The sections below provide additional information of likely interest to more specialized researchers in the subfield of early language development and speech analysis. In particular, we provide:

1. Our annotation guidelines for classifying the register and gender of talkers,
2. A description of our model-building process for analysis,
3. Supplementary analyses for infants who heard no male input,
4. Supplementary analyses for our primary reported models with added maximal random effects structure,
5. Supplementary figures depicting the relative presence of LENA-tagged talkers in our recordings with age,
6. A summary table of prior work on children's linguistic input, and
7. Correlations of LENA AWC estimates with the ADS and CDS quantity estimates in the current dataset.

Summary of annotation guidelines

Full annotation guidelines, including audio examples and instructions for how to download and use our custom data distribution and annotation software can be found on the following OSF repository: `https://osf.io/d9ac4/`. The instructions below have been shortened for clarity.

Basic guidelines for coding speech register and speaker gender:

> **Speech register:** Judge whether the segment sounds like speech that is directed to an infant or young child (child-directed speech; CDS) or another adult (adult-directed speech; ADS). If it's neither CDS nor ADS (e.g. mixed talkers, or just noise, or just baby sounds, or silence) call it junk.
>
> **Speaker gender:** Judge whether the speaker was a male or female.
>
> *Note:* You may use your knowledge about the content of the whole block to make your decisions.

Annotation workflow for a conversational block:

1. Load a conversational block (i.e., a sequence of utterance clips).
2. Listen to the entire block. Think about:
   - How many participants are there?
   - Who are the participants (adult vs. child/infant)?
   - Is the conversation as a whole directed toward an infant/child or toward an adult, or a mixture of the two?
3. Listen to and annotate each MAN and FAN clip in the block[1]
   - Is there a single adult speaker? If so, tag the speech as CDS/ADS following criteria described above.

[1]Other types of clips, e.g., OLN cannot be annotated in this framework.

- If not, is there a foregrounded adult speaker (a speaker who is easier to perceive than other talker(s) or noises in the segment and/or who is the speaker for the bulk of the segment)? If so, tag the speech as CDS/ADS for that speaker. If there is a succession of non-overlapping/foregrounded speakers, label the segment ADS/CDS as appropriate.
- Else tag the clip as junk. Only tag a clip with overlapping speech as junk if it is really mixed (e.g., 50/50 or 30/30/30) among multiple simultaneous speakers and there is no discernible foregrounded talker. Code laughter and baby crying FAN/MAN clips as junk
- Keep the following special cases in mind:
  (a) Expressive sounds with a vowel (e.g., oooh, uhh, umm, weee) should be coded as ADS/CDS, but those without a vowel (e.g., shh, kissy-sound, tongue- 'click', hmm, laughter) should be coded as junk.
  (b) All reading and singing should be coded as ADS/CDS so long as it contains vowels (e.g., humming should be coded as junk)
  (c) When the clip has both ADS and CDS, code it for its predominant register (use CDS if it's really 50/50).
  (d) If the speech *sounds* like CDS, but you know it's directed to an adult, pet, or other, still code it as CDS.
  (e) For phone conversations, only code the person on the recorder's end of the phone; not voices on the other end of the line (code them as junk),
  (f) When in doubt, use the context to help identify clips as ADS/CDS.

4. Submit and save your completed block before moving onto the next one.

Model-building process

Due to the exploratory nature of these analyses, we incrementally built each statistical model, only adding fixed-effects that significantly improved model fit. We added justified predictors in three steps: first adding justified simple effects, then two-way effects, then three-way effects.

We illustrate the process with a toy example below. In our example, we have three possible predictors (A, B, and C) with which to model our dependent variable (DV). All models were run using the lme4 package in R, so the following example appears in R pseudocode. The actual scripts used for analysis can be found on the project repository: `https://github.com/marisacasillas/NorthAmericanChildren-ADSvsCDS`.

Step 1. Make a baseline model with random effects but no fixed effects yet.

m0 <- lmer(DV ~ (1|corpus))

Step 2. Test whether any predictor significantly improves model fit on its own.

mA <- lmer(DV ~ A + (1|corpus))

anova(m0, mA)

mB <- lmer(DV ~ B + (1|corpus))

anova(m0, mB)

mC <- lmer(DV ~ C + (1|corpus))

anova(m0, mC)

Let's assume that both A and C improved fit significantly. We then proceed by adding A and C, using this model as our new baseline:

mA.C <- lmer(DV ~ A + C + (1|corpus))

Step 3. Repeat with two-way interactions, using the new baseline model.

mA.C.AB <- lmer(DV ~ A + C + A:B + (1|corpus))

anova(mA.C, mA.C.AB)

mA.C.AC <- lmer(DV ~ A + C + A:C + (1|corpus))

anova(mA.C, mA.C.AC)

mA.C.BC <- lmer(DV ~ A + C + B:C + (1|corpus))

anova(mA.C, mA.C.BC)

Let's assume that none of these two-way interactions improved model fit. We then proceed with the same baseline.

Step 4. Repeat with three-way interactions.

mA.C.ABC <- lmer(DV ~ A + C + A:B:C + (1|corpus))

(mA.C, mA.C.ABC)

Let's assume that the three-way interaction improves fit. Our final model is then:

mbest <- lmer(DV ~ A + C + A:B:C + (1|corpus))

Our data included values for all children for the following predictors: child age, child gender, child's number of older siblings, and maternal education level. When possible we also used adult speaker gender as a predictor; speaker gender is an item-level property (i.e. there's only one gender per utterance).

Speaker gender models for children who heard no male speech

Eight of the 61 children in our corpus heard no speech from males in the audio we annotated. When modeling speaker gender effects we chose not to include male-speech datapoints for these 8 children because we did not want to make inferences about the pattern of male ADS and CDS in cases where we had no data on which to base our inferences. However, an alternative point of view is that the lack of male speech for these 8 children is meaningful. If so, we should count these children as having 'zero' male ADS and CDS. For completeness we therefore ran parallel statistical models of gender effects in each of our three measures (CDS quantity, ADS quantity, and proportion CDS) in which cells were filled with a 'zero' when no male speech was observed. We present the results of these zero-based models side-by-side with those reported in the main paper (No-Male Model = dropped datapoints when no evidence that a male was present; 0-Male Model = male speech cells given a 0 when no male speech was observed). In each case, the best-fit model from our incremental model-building process was identical with the exception of the model for proportion CDS, in which the zero-based representations of no male speech resulted in an additional significant interaction of child age and speaker gender:

CDS quantity

	No-Male Model			0-Male Model		
	$N = 114$			$N = 122$		
	B	*SE*	*t*	*B*	*SE*	*t*
(Intercept)	8.6014	0.5642	15.246	8.6014	0.5538	15.533
AduGender = *male*	-5.4274	0.8274	-6.559	-5.8437	0.7831	-7.4621

ADS quantity

	No-Male Model			0-Male Model		
	$N = 114$			$N = 122$		
	B	*SE*	*t*	*B*	*SE*	*t*
(Intercept)	5.2641	0.4729	11.130	5.2641	0.4618	11.399
ChiAge	-0.6412	0.1013	-6.327	-0.6412	0.0990	-6.479
AduGender = *male*	-2.8482	0.6861	-4.151	-3.1286	0.6264	-4.995
ChiAge:AduGender = *male*	0.5440	0.1466	3.712	0.5219	0.1342	3.888

CDS proportion

	No-Male Model			0-Male Model		
	$N = 114$			$N = 122$		
	B	*SE*	*t*	*B*	*SE*	*t*
(Intercept)	0.6445	0.0312	20.634	0.6445	0.0347	18.547
ChiAge	0.0255	0.0052	4.933	0.0313	0.0074	4.207
AduGender = *male*	-0.1089	0.0430	-2.532	-0.1894	0.0491	-3.853
ChiAge:AduGender = *male*	NA	NA	NA	-0.0225	0.0105	-2.133

Primary models reported, with maximal random slopes added

Although we did not *a priori* expect our predictors to differentially affect CDS or ADS, it is common practice in some circles to fully maximize random effects structure (Barr, Levy, Scheepers, & Tily, 2013), including random slopes. Below we report a version of each of the primary models reported in the paper, but with maximal random effects structures:

For each model we added random slopes of child age, child gender, speaker gender, number of older siblings, maternal education, and their interactions, as allowed by the data. Full interactions between these predictors usually resulted in a non-converging model, so we dropped random slopes until models converged. In dropping random slopes we first removed higher-order interactions, then lower-order ones, then individual predictors. When forced to choose between models with alternative random slopes, we favored predictors and interactions that could more conceivably affect the random unit applicable (e.g., the effect of child age * number of siblings vs. child gender and number of siblings on sample of corpora analyzed). For further details, please see the analysis scripts at `https://github.com/marisacasillas/NorthAmericanChildren-ADSvsCDS`.

Overall, we find that model outcomes with maximal random effects structure added are nearly identical to those reported in the main text, with no qualitative differences at all.

CDS quantity overall

	B	*SE*	*t*
(Intercept)	9.4786	1.3010	7.285
MatEducation	1.1939	0.5639	2.117

CDS quantity with speaker gender

	No-Male Model $N = 114$			0-Male Model $N = 122$		
	B	*SE*	*t*	*B*	*SE*	*t*
(Intercept)	9.4252	0.6699	14.070	9.7111	0.6382	15.217
AduGender = *male*	-5.9856	0.9386	-6.377	-7.0061	0.8251	-8.491

ADS quantity overall

	B	*SE*	*t*
(Intercept)	7.7636	0.6677	11.627
ChiAge	-0.7776	0.1400	-5.552

ADS quantity with speaker gender

	No-Male Model			0-Male Model		
	$N = 114$			$N = 122$		
	B	*SE*	*t*	*B*	*SE*	*t*
(Intercept)	5.2641	0.4729	11.130	5.50369	0.44730	12.304
ChiAge	-0.6412	0.1013	-6.327	-0.61434	0.09857	-6.233
AduGender = *male*	-2.8482	0.6861	-4.151	-3.33516	0.61951	-5.384
ChiAge:AduGender = *male*	0.5440	0.1466	3.712	0.44097	0.13948	3.161

CDS proportion overall

	B	*SE*	*t*
(Intercept)	0.643338	0.023951	26.861
ChiAge	0.026955	0.005013	5.377

CDS proportion with speaker gender

	No-Male Model			0-Male Model		
	$N = 114$			$N = 122$		
	B	*SE*	*t*	*B*	*SE*	*t*
(Intercept)	0.642086	0.028269	22.714	0.675102	0.034273	19.698
ChiAge	0.028079	0.004811	5.836	0.034327	0.007304	4.700
AduGender = *male*	-0.101394	0.041021	-2.472	-0.189352	0.045862	-4.129
ChiAge:AduGender = *male*	NA	NA	NA	-0.022465	0.009828	-2.286

LENA tags across child age

In the data analyzed here, all tags present in the LENA output decrease with age, with the exception of the child's speech tag (CHN). This holds true whether we look at raw number of tags or average total duration of tags with a value at each age.

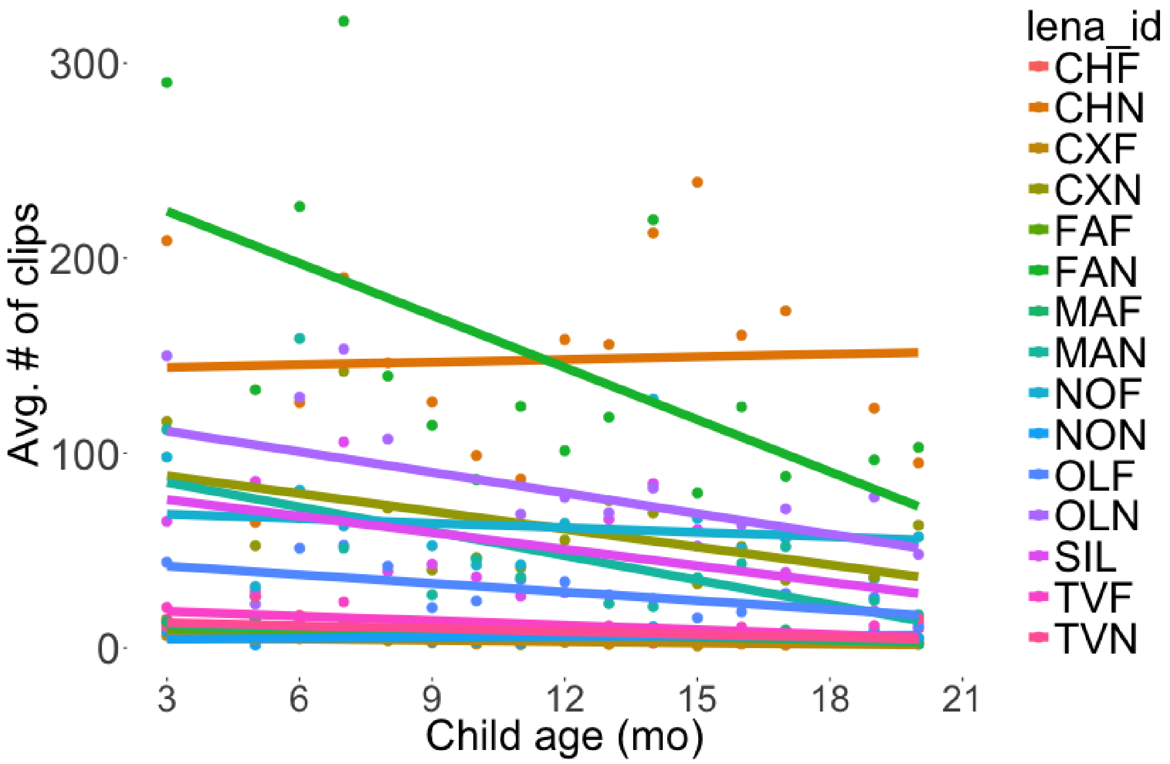

Figure S1. Average number of clips identified with the different tags given by LENA's software in our data, over month-age of the child. The LENA-generated speaker tags are shown in the legend ('lena_id'). For all categories but 'silent' (SIL), the final letter indicates LENA's estimate of whether the speech is near (N) or far (F), based on whether the segment was clearly distinguishable from the SIL category. The codes for the seven speaker categories are: CH: target child; CX: other child, FA: female adult; MA: male adult, NO: noise, OL: overlap, TV: electronic sound. *N.B.*: only FAN and MAN categories were tagged for CDS, ADS, and gender in our analyses.

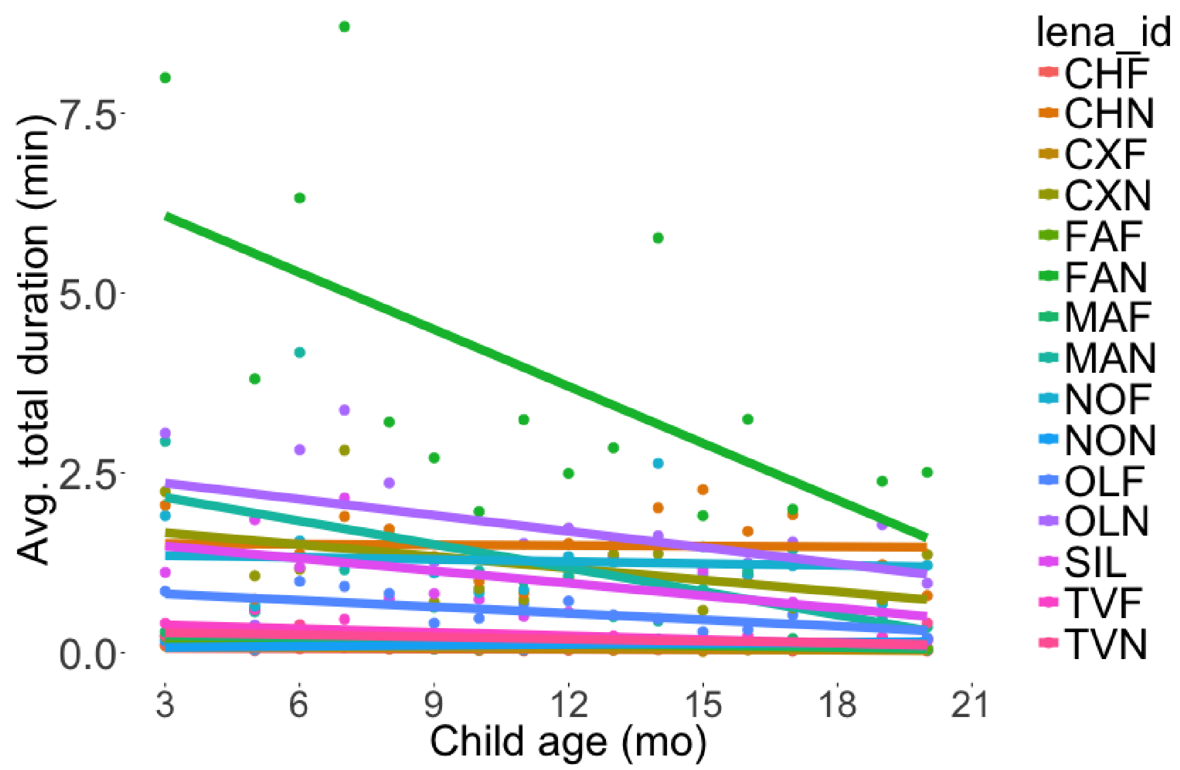

Figure S2. Average total duration in minutes of the different tags given by LENA's software in our data, with averages taken over month-age of the child. The values for 'lena_id' in the legend are clarified in the caption for Figure S1.

Prior work on children's linguistic input

The table below summarizes a representative (non-exhaustive) sample of previous research on quantitative measures of language input to children under 3 years. "LENA AWC" refers to LENA adult word count estimates from a full-day recording, unless otherwise specified.

Table S1

Quantitative measure is given in mean words/hour across the sample, regardless of recording length unless specified otherwise. CDS= child-directed speech; All = all speech heard by the child.

Reference	**Quantitative Measure(s)**	**Sampling Technique**	**N / Gender**	**Age (mo.)**	**SES**	**Other**
Shneidman & Goldin-Meadow (2013)	CDS: 2063		10F/5M	30	range	Multi-speaker
	All: 6254	90-min rec	10F/5M			
	CDS: 2404	"	8F/7M	"	"	Single-speaker
Weisleder & Fernald (2013)	CDS: 67–1,200 All: 200–2,900 range per hour	LENA AWC	19F/10M	19	low SES	Spanish-learning
Johnson et al. (2014)	All: 1725	LENA AWC	17F/16M	0	middle	8 infants preterm
	All: ∼1,000	"	"	7	"	Longitudinal
Tamis-LeMonda et al. (2017)	CDS: 2197	45-min rec	20F/20M	13	middle–upper	
Pancsofar & Vernon-Feagans (2006)	All mother: 2559 All father: 1919	20-min rec	45F/47M	24	middle–upper	
Gilkerson et al. (2017)	All: 1,000–1,500	LENA AWC (12-hr rec)	162F/167M	2–48	range	Longitudinal
Hart & Risley (1995)	CDS: 2153	1-hr rec	8F/ 5M	13–36	professional	
	CDS: 1251	"	12F/11M	"	working class	
	CDS: 616	"	3F/3M	"	welfare	

Table S2

Adult word count(AWC) and Child Vocalization count(CVC) in four previously published papers using LENA recordings, along with the 59/61 recordings in the current dataset that are >8hrs. N.B., SDs were not available for the Greenwood et al. (2011) data, and CVC was not available for the Zimmerman et al. (2009) data.

Study	**AWC Mean**	**AWC SD**	**CVC Mean**	**CVC SD**
Gilkerson et al. (2017)	12709	4274	1817	787
Greenwood et al. (2011)	13142	NA	1714	NA
Soderstrom & Wittebolle (2013)	10125	4890	1744	1058
Zimmerman et al. (2009)	12800	4400	NA	NA
Current dataset	16510	8718	1432	764

Comparing ADS and CDS minutes per hour to LENA AWC and CVC estimates

We additionally checked whether the AWC and CVC estimates from recordings in the current dataset correlated with the ADS and CDS minutes per hour we computed from the 1220 selected blocks. We indeed find that AWC correlates with ADS minutes per hour ($r_s(59) = .365$, p = .004) and with CDS minutes per hour ($r_s(59) = .300$, p = .019) in the current dataset.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Gilkerson, J., Richards, J. A., & Topping, K. (2017). Evaluation of a lena-based online intervention for parents of young children. *Journal of Early Intervention*, 1053815117718490.

Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, *32*(2), 83–92.

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? the influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS one*, *8*(11), e80646.

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, *124*(1), 342–349.