An Automated Classifier for Child-Directed Speech from LENA Recordings

Janet Y. Bang, George Kachergis, Adriana Weisleder, and Virginia A. Marchman

1. Introduction

Children's speech environments vary in numerous ways. Our ability to make claims about naturally-occurring speech in children's daily environments has been greatly assisted by recorders that can capture and store large amounts of audio (e.g., an entire day). One notable example is the LENA digital language processor (Gilkerson et al., 2017), which is worn inside a child's front shirt pocket and stores up to 16 hours of the audio environment around the child per recording. Analyzing these recordings is facilitated by LENA software that automatically segments and classifies the audio into relevant categories (e.g., silence, speech, television) and estimates the number of utterances or words spoken "near and clear" to the child. However, these automated measures gloss over many important features of audio environments, for example, if silence represents periods when the child is sleeping, or if adult speech is directed to the target child. The directed nature of speech has received particular attention, as a growing body of work has proposed that child-directed speech (CDS), more so than other-directed speech (ODS), supports lexical development (Ramírez-Esparza, García-Sierra & Kuhl, 2014; Shneidman & Goldin-Meadow,

* Janet Y. Bang, San José State University, janet.bang@sjsu.edu, George Kachergis, Stanford University, kachergis@stanford.edu, Adriana Weisleder, Northwestern University, adriana.weisleder@northwestern.edu, Virginia A. Marchman, Stanford University, marchman@stanford.edu.

We are especially grateful to the families for their contribution to this research. We would also like to thank Anne Fernald for supporting the studies that enabled this work, the Language Learning Lab staff for their tireless work in hand-coding these data, and the members of the Language and Cognition lab at Stanford and the LangView consortium for their thoughtful comments and suggestions. This work was supported by grants from the National Institutes of Health (R01 HD42235, DC008838, HD092343), the Schusterman Foundation, the W.K. Kellogg Foundation, the David and Lucile Packard Foundation, and the Bezos Family Foundation to Anne Fernald, the National Institutes of Health (2R01 HD069150) to Heidi Feldman, the National Institutes of Health (R21 DC018357) and a Elizabeth Munsterberg Koppitz Child Psychology Graduate Student Fellowship from the American Psychological Foundation to Adriana Weisleder, and a Postdoctoral Support Award from the Stanford Maternal and Child Health Research Institute to Janet Bang.

 Proceedings of the 46th annual Boston University Conference on Language Development, ed. Ying Gong and Felix Kpogo, 48-61. Somerville, MA: Cascadilla Press.

2012; Weisleder & Fernald, 2013). To explore these more detailed aspects of children's learning environments, researchers have typically relied on labor-intensive manual coding by humans who listen to and consider multiple features of the speech environment (Ramírez-Esparza et al., 2014; Weisleder & Fernald, 2013). To facilitate this work, we present two automated classifier systems that take the output from the LENA software and identify periods of sleep in the recording and segments which are primarily CDS and ODS. The development of these tools can both ease the burden of labor-intensive coding and potentially yield insights into the characteristics of speech in children's everyday language learning environments.

1.1. Child-directed versus other-directed speech for language learning

The importance of child-directed speech is central to theories that aim to explain how children learn language from social interactions (Tomasello, 1995). However, communities vary widely in how much speech is directed to children (Casillas et al., 2019; Ochs & Schieffelin, 1984; Shneidman & Goldin-Meadow, 2012). Despite this variability, cross-cultural work finds that key language milestones (e.g., onset of first words and multi-word utterances) emerge around the same age in a variety of communities (Casillas et al., 2019; Crago et al., 1997). Such evidence raises questions regarding how speech in children's environments supports language acquisition.

In addition, the ways in which child-directed speech specifically, rather than all available speech, contributes to children's language learning is an active area of research. Some lab-based experimental studies demonstrate that children can learn new words from speech that is not explicitly directed to them. For example, Akhtar and colleagues (2001) found that 1- to 2-year-old children were able to learn novel nouns and verbs when observing two adults play a game. Other studies have varied the degree of joint attention, such as having speakers turn their backs to infants, replicating these findings (Gampe et al., 2012). In contrast, research examining speech in natural environments reports that child-directed speech, more so than other-directed speech, is associated with children's vocabulary development. For example, using LENA recordings with 29 Spanish-speaking families in the US, Weisleder and Fernald (2013) recorded speech environments and then hand-coded periods of child-directed vs. overheard speech. Child-directed speech was directed to the target child, either in one-on-one interactions or with others; overheard speech was directed to adults or children other than the target child. Using estimates derived from LENA to measure total adult word counts, variability in child-directed adult words at 19 months was related to children's vocabulary size at 25 months, while adult words during periods of overheard speech was not. Similar findings were seen in Shneidman and Goldin-Meadow (2012), where child-directed, but not overheard, speech was linked to children's vocabularies in Yucatec-Mayan-speaking families in subsistence farming communities in Mexico.

1.2. Identifying periods of child-directed and other-directed speech in daylong recordings

Daylong recordings provide an extraordinary opportunity to examine how speech varies across different contexts in young children's natural environments. For example, some researchers seek to examine the audio environments of young children during periods when the child is awake compared to periods when the child is sleeping. Other researchers are interested in identifying which features of talk differ in child-directed contexts compared to periods when that speech is not directed to the target child. To examine how features of talk differ across contexts, it is first necessary to identify these contexts within daylong recordings. In studies to date, human listeners are trained to identify periods of sleeping, child-directed, and other-directed speech by attending to numerous cues that are available on the audio recording. However, less is known about which of the multiple available features most reliably characterize these different contexts. Moreover, while fruitful, these efforts are highly labor and time intensive. Though there are emerging tools to support the efficiency of this type of manual coding (Cychosz et al., 2021), efforts to create automated tools are also in critical need. Additionally, in some cases, ethical considerations prevent researchers from listening to the recordings (Cychosz et al., 2020). Thus, tools that enable classification of periods of child-directed and other-directed speech from features that are automatically extracted from the recordings could expand the range of cases in which such features can be examined.

One advantage of LENA is that it provides an array of automated measures that characterize the child's audio environment (Xu et al., 2009), including: the frequency of adult words (AWC), conversational turns (CTC), and child vocalizations (CVC), and time-based estimates of noise, silence, distant speech (i.e., duration of speech far from the target child or overlapping), meaningful speech (i.e., duration of speech that is near and clear to the child), and electronics (e.g., TV). Typically, these measures are used independently, ranking individual families as having higher vs. lower mean AWC/hour or CTC/hour, or ranking children as having higher vs. lower CVC/hour. However, it is also possible that these measures can be used in conjunction to distinguish more subtle differences among periods of time during the daylong recordings. For example, for a given 5-minute segment, estimates of AWC may be more likely to be target-child-directed speech (tCDS) when accompanied by relatively high values of CTC or CVC. Or, a child is more likely to be sleeping when low values of AWC are also accompanied by low values of CVC or CTC. Finally, segments with relatively more minutes of distant speech may be more likely to be ODS than segments with more minutes of meaningful speech. By combining these features in various ways, we can gain insights into which features characterize different periods in a child's learning environment. Here, we explored ways to combine automatically-generated estimates from LENA to classify different kinds of speech segments. By comparing the results of these methods to previously-made judgments from human listeners, we estimated their

reliability; thereby, potentially identifying tools that could reduce the burden of hand-coding in future studies.

1.3. Current study

We examined the feasibility of training automated classifiers to reliably identify periods of (1) a child sleeping, and (2) CDS to the target child (tCDS) versus speech that is available to the child but directed to others that are not the target child, i.e., ODS. We first conducted preliminary analyses using only the frequency counts, i.e., AWC, CTC and CVC, derived from recordings of 29 Spanish-speaking families (Weisleder & Fernald, 2013). We assessed the degree to which variation in these measures was associated with variation in whether a particular 5-minute segment was classified as tCDS or ODS by human coders. Next, we applied more sophisticated machine-learning classifiers that combine multiple frequency- and time-based measures from LENA to identify periods of sleep, tCDS, and ODS. We first used cluster analyses to examine how multiple LENA features hang together to predict the judgements of human listeners. We then trained a sleep classifier and a tCDS/ODS classifier, comparing the results of both classifiers to human coders in a large sample of 153 English- and Spanish-speaking families. Finally, we examined if AWC values based on model-predicted segments of tCDS versus ODS replicated previously published links with hand-coded data and children's later language outcomes (Weisleder & Fernald, 2013).

2. Method

We analyzed daylong LENA recordings across five samples of children from a total of 153 families with 17- to 28-month-old children. For all samples, human listeners had coded 5-min or 10-min audio segments for periods of sleep, tCDS, or ODS following similar protocols. We first conducted a logistic regression to assess relations among AWC, CTC, and CVC. We then conducted a cluster analysis and trained automated classifiers using the hand-coded segments and LENA-derived measures, including features of speech (AWC, CTC, CVC) and features based on time (meaningful speech, distant speech, TV, noise, and silence).

2.1. Participants

Participants were families and their 17- to 28-month-old children from 79 English- and 74 Spanish-speaking households in the US. In total, families contributed over 1,000 recorded hours of LENA recordings (12,936 segments). Descriptives can be seen in Table 1.

Table 1: Descriptives of families across five different samples

Sample	n	Lang.	Age (mo)	Mat. Ed (y)	Record. length (hours)	Seg. dur (min)	Num seg. incl.
1	27	En	18 - 19	12 - 18	10.62 (2.29)	5	3491
2	29*	En	17 - 19	10 - 18	9.32 (2.52)	5	3275
3	45	En	23 - 26	10 - 18	11.05 (3.22)	10	1891
4	29	Sp	18 - 20	4 - 16	10.67 (3.13)	5	2758
5	45	Sp	25 - 28	6 - 18	13.44 (3.68)	10	1521

*n = 22 from Sample 2 are also included in Sample 3 at a second time point, thus the total sample results in 153 unique families; En = English, Sp = Spanish

2.2. Data collection and coding

Across all studies, research staff obtained informed consent from caregivers and provided instructions of how to use the LENA. Caregivers were asked to record on a 'typical' day. To respect families' privacy, caregivers were told that they were able to pause the recording at their convenience. Instructions varied slightly across samples, but in all cases, families recorded on a single day or across multiple days and were encouraged to record during all parts of the day.

Native speakers of each language hand-coded 5- or 10-min segments. To determine periods of sleep, coders identified when audio segments consisted of multiple consecutive AWC values of 0, and next listened to confirm if children were sleeping. If the child was confirmed to be sleeping (e.g., often evident by deep breathing), coders continued listening to segments prior to and after these segments to determine the beginning and end of periods of sleep.

To identify segments of tCDS or ODS, coders classified each segment based on the most prevalent type of language interaction in that segment. tCDS was defined by speech that was directed to the target child, whether this was only to the target child or seemed inclusive of the target child (e.g., if a speaker addressed a group that included the target child, for example they said "you all" or "look at this", the speech would be considered tCDS). Coders based their judgment on numerous features including exaggerated prosody, pace, affect, number of people present, distance of the speaker relative to the child, environmental sounds, semantic properties of the speech, and the activity of the interaction. Periods of ODS were identified by speech that did not appear directed to the target child or inclusive of the target child. For samples 3 and 5 coders were asked to identify the prevalent type of language interaction (i.e., approximately 70% of speech was either tCDS or ODS); for samples 1, 2, and 4 coders were asked to identify whether the segment was majority (>50%) tCDS or ODS but were also able to identify a third category that indicated if a sample

was ‘split’ to be approximately 50% tCDS and 50% ODS. For the classifiers, we treated all ‘split’ segments as ODS, so that all segments coded as tCDS reflected primarily child-directed speech. Coders could use the AWC value to determine how much speech in the segment needed to be classified as tCDS or ODS.

3. Results

3.1. Preliminary analysis

We conducted hierarchical mixed effects logistic regression models to examine the degree to which LENA-provided frequency measures of AWC, CTC, and CVC predicted the hand-coded classifications of tCDS or ODS. We used Sample 4, which consisted of all-day LENA recordings from a previously-published study where 5-minute segments had been coded for tCDS or ODS (Weisleder & Fernald, 2013). Models included a random intercept of participant and all frequency measures were converted to rates per minute and mean-centered within each family to allow interpretation of values as relative to each family’s mean rates. We found that each frequency measure, AWC/min, CTC/min, and CVC/min, independently contributed to the probability of a segment being classified as tCDS versus ODS. As seen in Figure 1, lower AWC rates (B = -.59, 95% CI [-.73, -.46]) were associated with a higher probability of ODS, indicating that a one unit increase in AWC above a family’s mean would result in a lower probability of the segment being coded as tCDS. In contrast, higher rates of CTC (B = .36, 95% CI = [.21, .52]) and CVC (B = .39, 95% CI = [.26, .52]) resulted in a higher probability that the segment was coded as tCDS. These findings indicated that each of the LENA frequency measures predicted the probability of tCDS, but did so in different directions, suggesting that relations were more complex than these techniques could capture. Thus, we next recruited machine learning techniques to explore the extent to which multiple LENA features could be used to classify periods of sleep, tCDS, or ODS.

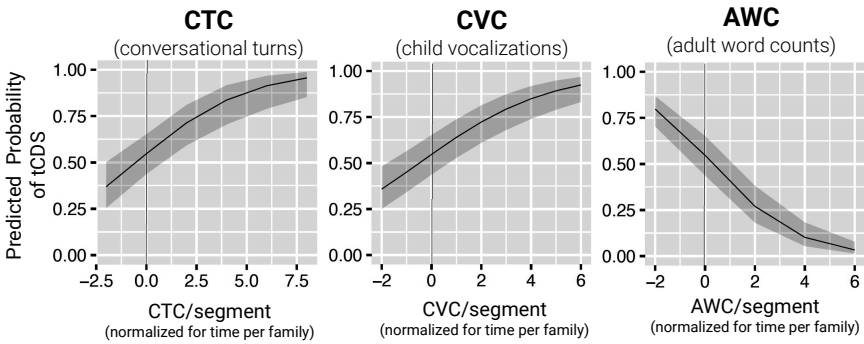

Figure 1. Predicted probabilities and confidence intervals (shaded region) of tCDS from AWC, CTC, and CVC, when holding other measures at their mean (vertical line at 0)

3.2. Cluster analysis

We next examined whether segments could be meaningfully clustered, which might suggest that a classifier based on thresholding multiple values (e.g., a decision tree) might work better than techniques that look at predictors individually. Using an unsupervised clustering algorithm (*k*-means), we clustered all 12,936 segments according to their raw LENA values, considering *k*={2,..,15} clusters. Table 2 shows the preferred *k*=7 clusters along with the proportion of each type of segment in the cluster and the mean values of LENA features for segments in that cluster (e.g., AWC, CTC, silence, noise). Clusters 4 and 5 capture mostly sleep (64% and 53%) with low AWC, CTC, and CVC, but both clusters also include a fair number of CDS segments (22% and 30%). Note that Cluster 5 is also associated with high levels of noise, whereas Cluster 4 is associated with high levels of silence. Clusters 6 and 1 are both predominantly CDS (73% and 60%) and cover 36.4% of the dataset, however, one has somewhat higher mean AWC, CTC, and CVC values than the other. Note also that these two clusters also contain many ODS segments. Clusters 7 and 2 are comprised mostly of ODS segments. While both clusters are associated with low values of CTC and CVC, Cluster 7 is associated with high values of AWC, while Cluster 2 is not. Finally, Cluster 3, which looks much like the sleep clusters (4 and 5) in terms of low AWC, CTC, and CVC, is also associated with a higher level of TV.

Overall, clustering the segments according to the LENA measures showed that: 1) multiple LENA features captured meaningful variation between the clusters, as some corresponded mostly to sleep, tCDS, or ODS, and yet 2) the clusters have significant overlap in tCDS and ODS, and to a lesser extent, sleep.

Table 2: Means of LENA variables by cluster, annotated with proportion of sleep, CDS, and ODS segments

cluster	N	sleep	CDS	ODS	AWC	CTC	CVC	noise	silence	distant	TV	meaningful
4	2041	**0.64**	0.22	0.14	3.01	0.07	0.49	0.01	0.85	0.08	0.02	0.03
5	142	**0.53**	0.30	0.17	3.44	0.11	0.86	0.63	0.12	0.16	0.04	0.04
6	1256	0.00	**0.73**	0.27	54.55	3.78	9.51	0.02	0.27	0.25	0.01	0.45
1	3450	0.01	**0.60**	0.39	21.97	1.14	4.76	0.03	0.37	0.33	0.03	0.25
7	1485	0.01	0.33	**0.66**	76.10	1.40	2.61	0.01	0.21	0.33	0.03	0.42
2	3475	0.04	0.45	**0.51**	13.63	0.37	1.80	0.03	0.17	0.66	0.02	0.12
3	1087	0.27	0.28	**0.45**	7.33	0.16	0.73	0.02	0.15	0.07	0.69	0.06

3.3. Identifying sleep segments

We next attempted to build a classifier to automatically distinguish sleep from awake segments using only automatically-generated LENA features. A simple decision tree classifier was trained to distinguish segments when the target child was asleep from those when they were awake, mirroring the first step that researchers often perform when manually cleaning a LENA dataset. We

trained the model using 5-fold cross-validation on 90% of 12,936 segments (1,879 sleep segments, 11,057 awake). The decision tree achieved an AUC of 0.881 on the held-out test set of 1,294 segments. Shown in Figure 2, the decision tree splits first on the amount of "meaningful" speech per minute, and then on silence. If meaningful speech is >0.4s per minute (i.e., where .4 is 0.0063 x 60 sec) and silence is less than 54s per minute (i.e. 0.9), then segments are highly likely to be awake (96.5%; 9,707 of 10,056 segments). If meaningful speech per minute is very low (i.e. <0.4s) and silence per minute is high (>.64, i.e. 38.4s), then these segments are highly likely to be sleep (97%, i.e. 1022 of 1054 segments). If meaningful speech is very low but silence per minute is also low, then when child vocalizations (CVC) are low (<0.1), segments are also somewhat more likely to be sleep (66%; 263 of 397 segments).

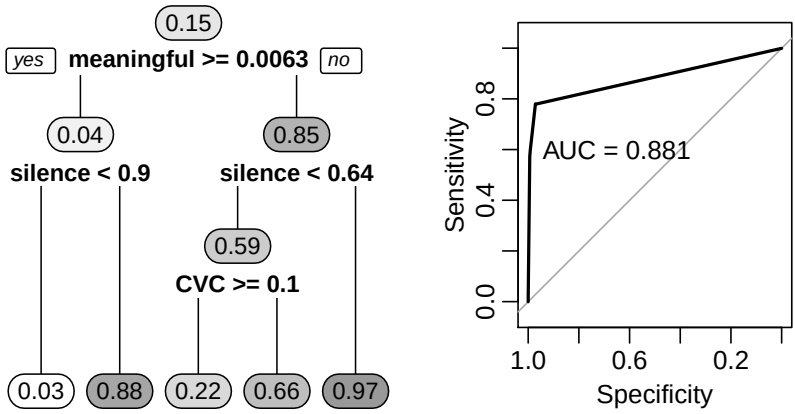

Figure 2. (left) Decision tree for classifying LENA segments with sleep (dark gray) vs. no sleep (white), and (right) the ROC curve for this classifier

3.4. Classifying tCDS vs. ODS segments

We turn now to the more challenging task of building a classifier to automatically distinguish tCDS from ODS segments. While a simple decision tree performed very well for the simple case of sleep, it did not work as well for the overlapping clusters of child-directed and other-directed speech. Thus, we instead used a more sophisticated machine learning model: XGBoost (eXtreme Gradient-Boosted trees; Chen & Guestrin, 2016), a state-of-the-art algorithm that trains a cascade of decision trees successively on subsets of the data, upweighting the segments that were misclassified by earlier decision trees.

We trained a classifier on LENA features to distinguish tCDS segments from all other segments (ODS and split segments). First, we removed the 1,879 segments during which children were asleep (assuming they would be cleaned by hand or removed by the sleep classifier). We then reclassified the 1,012 "split" segments which raters judged to be 50% ODS and 50% tCDS as ODS

(0), making a total of 5,239 ODS segments and 5,818 tCDS segments (58% tCDS). The purpose of the classifier is thus to distinguish "pure" tCDS from mixed or pure ODS, after removing periods of sleep. A random 90% of the data (9,951 segments) was used to train the classifier, and the remaining 10% (1,106 segments) was used for evaluation. When trained on 90% of the segments using the normalized LENA features, the XGBoost classifier achieved an AUC of 0.719, with an overall accuracy of 0.674 on the held-out data.[1] Figure 3 shows the ROC curve (left) and the relative importance of each LENA feature (right) in the final classifier. A limitation of XGBoost is that it does not enable simple visualizations, e.g., a decision tree, of how classifications are made. However, we can use the feature importance measure to assess which features were most informative in the ensemble of boosted trees. The classifier's reliance on the amount of silence, CTC, AWC, and meaningful talk per segment corresponds well to researchers' intuitions about how tCDS can be summarized.

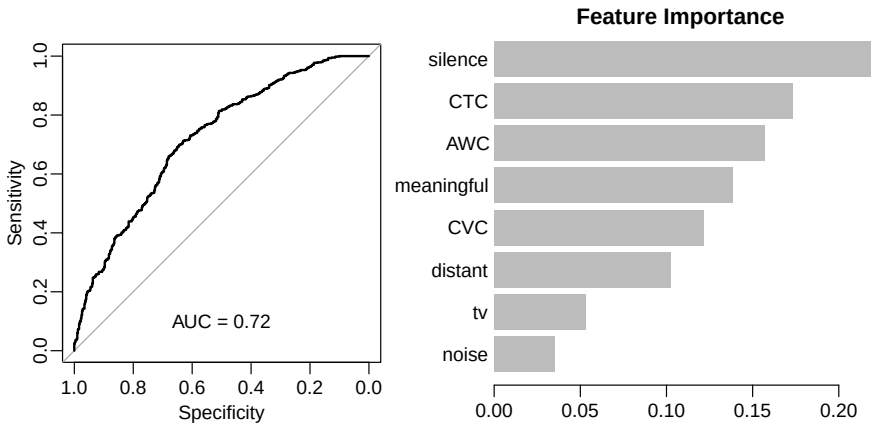

Figure 3. (left) ROC curve of the tCDS/ODS classifier and (right) relative importance of the LENA features in this classifier

3.5. Replication of links between tCDS and child language outcomes

One critical question is whether the tCDS/ODS classifier works sufficiently well to replicate results from manually-annotated studies. To test this, we used the Weisleder & Fernald (2013) dataset of 29 Spanish-speaking children whose caregivers completed the MacArthur-Bates Mexican Spanish CDI

[1] To ensure that the classifier was not just learning characteristics of these particular children, we also trained a cross-validated version on 90% of the children, leaving out data from 10% of the children (n=15) in each fold. This classifier achieved approximately the same performance (AUC = 0.73; average test accuracy = 0.66), suggesting that the classifier will generalize well to data from additional children.

(Jackson-Maldonado et al., 2007) to assess vocabulary size when the children were 24 months. In this manually-annotated dataset, children who heard more tCDS at 19 months had significantly larger vocabularies at 24 months (r = .52, 95% CI=[.19, .75], p = .004). However, there was no significant association between the amount of ODS at 19 months and vocabulary size at 24 months (r = .25, p = .199).

We attempted to replicate this result using the classifier's predictions of which segments were classified as tCDS vs. ODS. As in the original manual annotations, children who heard more tCDS at 19 months had significantly larger vocabularies at 24 months (r=.48, 95% CI=[.14, .72], p=.008), but there was no significant relation between the amount of ODS and vocabulary size (r = .32, 95% CI=[-.06, .61], p=.094). Notably, the strength of these correlations using the hand-coded and model-predicted values are very similar, providing further evidence that the classifier is an effective tool for this purpose.

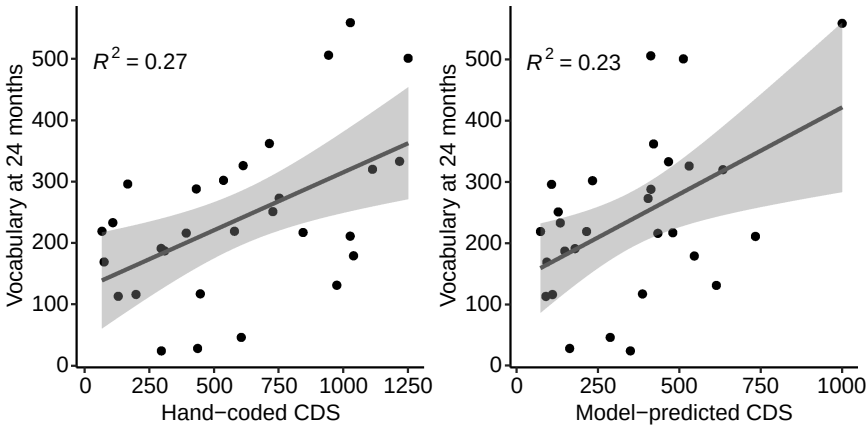

Figure 4. Scatterplots between hand-coded or model-predicted tCDS at 19 months and children's later vocabulary sizes at 24 months

4. General Discussion

Our study suggests that standard LENA outputs can facilitate identification of potentially meaningful sources of variation in children's speech environments. We discuss our five main insights in turn.

First, we found differences in how commonly-used frequency measures of AWC, CTC, and CVC predicted the probability of a segment having target-child-directed speech. Our preliminary analyses indicated that segments with higher AWC relative to a family's mean were more likely to be other-directed rather than target-child-directed. Frequency measures of CTC and CVC resulted in the opposite, where segments with higher values relative to a family's mean were more likely to be target-child-directed. These findings suggest a possibly counter-intuitive finding that periods of child-directed speech

are defined by relatively *lower* rates of adult words and relatively *higher* rates of conversational turns and child vocalizations.

Second, a much more complex picture arose when including both LENA frequency and duration measures in cluster analyses. While some distinct features characterized different audio environments, there was also a high degree of overlap across clusters. For example, as expected, clusters with more sleep segments were characterized by the lowest AWC, CTC, and CVC. However, one sleep cluster was characterized by more silence, while the other was characterized by more noise. This aligns with human coders' reports that periods of sleep often involved what appeared to be fans or sound machines, sounds which were likely categorized as "noise" by LENA. In predicting tCDS, clusters were characterized by the highest averages of CTC and CVC, but were more mixed with regards to AWC. In predicting ODS, one cluster consisted of the highest average AWC, while the others had lower CTC and CVC values, or longer durations of distant speech and TV. Thus, we observed multiple ways in which features were combined for all three categories of sleep, tCDS, and ODS, with each category represented by more than one cluster (and not perfectly). Future work might fruitfully examine in more detail potential differences between segments in different cluster types. For example, are segments in some clusters more difficult for human raters to classify than other segments? Or, are segments in some clusters associated with different types of language and/or activities than other segments?

Third, we found a high degree of success in training a classifier to identify periods of sleep. Using a simple decision tree classifier, AUC approached .90 for both the full dataset and the held-out test segments. Consistent with the multifaceted nature of clusters defined by more sleep, the classification was not simply due to periods of silence. The decision tree first considered the duration of 'meaningful' speech, which is defined as speech by a live speaker that is near the child; it then considered the duration of silence, before finally considering the number of children's own vocalizations. This suggests that periods of sleep can be reliably identified from characteristics of the audio environment and shows advantages of considering multiple features of those environments.

Fourth, we found moderate success in training a classifier to distinguish periods of tCDS versus ODS. To accommodate the overlapping clusters of tCDS and ODS, we used a more sophisticated machine-learning model, XGBoost, where the model could sequentially add a cascade of decision trees and weight misclassifications by earlier decision trees. We found moderate sensitivity and specificity on the full dataset and a slightly weaker AUC on the held-out test segments. The feature importance list illustrated the average gain in our prediction of tCDS versus ODS, highlighting many features that also emerged in our cluster analysis. Ongoing work suggests that reliability between model-derived versus human-coded segments are similar to interrater reliability between human coders, thus moderate success of the classifier may be a reasonable goal given the complexities of the speech environment. The superior performance of the classifier relative to analyses with individual predictors

suggests that human classifications of child-directed and other-directed speech rely on nuanced distinctions that take into account *combinations* of features in the audio environment (e.g., low silence with high CTC and moderate AWC).

Finally, we demonstrated that we could use model-derived predictions of tCDS and ODS to replicate associations between caregiver speech at 19 months and children's vocabularies at 24 months that were observed in previously published work (Weisleder & Fernald, 2013). Despite moderate success in accuracy with the tCDS/ODS classifier, the model-derived predictions revealed, as previously observed with hand-coding, that variability in adults' directed speech to target children was positively and significantly related to children's later vocabularies, whereas this link was not statistically significant when adult speech was directed to others.

5. Suggested uses of the classifier

Taken together, our findings suggest that applying classifiers to LENA data may facilitate data cleaning, coding, and analysis. First, the sleep classifier can automate one laborious step of 'cleaning' daylong LENA recordings with a reasonably high degree of reliability. Second, the tCDS/ODS classifier could also be used to reduce the significant hours of manual labor when coding periods of target-child- or other-directed speech. We have found that the classifier's per-segment probability of tCDS matches well with the uncertainty of human raters (e.g., the 50/50 "split" segments were classified as ~50% probability of being tCDS). This suggests that the classifier could be used to judge the high-confidence of 2/3rds of the data, and then pass the segments it has less certainty about to a "human in the loop" for closer scrutiny.

6. Limitations

While we included over 1,000 hours of data that had been hand-coded in 153 English- and Spanish-speaking families from varied socioeconomic backgrounds, our sample represents a small subset of the variability that exists within English- and Spanish-speaking families in the US. Our sample also represents a tiny subset of the linguistic and cultural variability in child-rearing environments around the world. Further validation studies are critical to understand whether our classifiers can generalize to new languages and communities (Cristia et al., 2021). Additionally, while our classifier is open-source, LENA software is not; thus, the ability to use this classifier on researchers' own data requires a substantial cost to purchase the LENA recorders and software. Future work should compare whether our classifiers can be used with open-source speech algorithms (e.g., ALICE; Räsänen et al., 2021) to achieve similar performance in sleep and tCDS/ODS classifiers.

7. Conclusion

The findings here present exciting opportunities for advancement in understanding how children learn from the available speech in their environment. We were able to train and validate two automated classifiers using LENA-based measures to identify periods of sleep and to distinguish between periods of tCDS versus ODS. This work has the potential to significantly reduce the time-consuming nature of identifying periods of directed speech to target children from the rich and naturalistic information collected with daylong recordings. We hope that by improving our systematicity in understanding shared and different features of target-child- and other-directed speech, we can better understand how children across different communities acquire and develop their language skills.

References

Akhtar, Nameera, Jipson, Jennifer, & Callanan, Maureen A. (2001). Learning words through overhearing. *Child Development, 72*(2), 416–430. https://doi.org/10.1016/S0163-6383(98)91471-0

Casillas, Marisa, Brown, Penelope, & Levinson, Stephen C. (2019). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Chen, Tianqi, & Guestrin, Carlos (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: ACM. https://doi.org/10.1145/2939672.2939785

Crago, Martha B., Allen, Shanley E. M., & Hough-Eyamie, Wendy P. (1997). Exploring innateness through cultural and linguistic variation. In M. Gopnik (Ed.), *The biological basis of language* (pp. 70–90). Oxford: Oxford University Press.

Cristia, Alejandrina, Lavechin, Marvin, Scaff, Camila, Soderstrom, Melanie, Rowland, Caroline, Räsänen, Okko, Bunce, John, & Bergelson, Elika. (2021). A thorough evaluation of the Language Environment Analysis (LENA) system. *Behavior Research Methods, 53*(2), 467–486. https://doi.org/10.3758/s13428-020-01393-5

Cychosz, Margaret, Villanueva, Anele, & Weisleder, Adriana. (2021). Efficient estimation of children's language exposure in two bilingual communities. *Journal of Speech, Language, and Hearing Research, 64*(10), 3843–3866. https://doi.org/10.31234/osf.io/dy6v2

Cychosz, Margaret, Romeo, Rachel R, Soderstrom, Melanie, Scaff, Camile, Ganek, Hillary, Cristia, Alejandrina, Casillas, Marisa, de Barbaro, Kaya, Bang, Janet, & Weisleder, Adriana. (2020). Longform recordings of everyday life: Ethics for best practices. *Behavior Research Methods*, *52*(5), 1951–1969. https://doi.org/10.3758/s13428-020-01365-9

Floor, Penelope, & Akhtar, Nameera. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy, 9*(3), 327–339. https://doi.org/10.1207/s15327078in0903_4

Gampe, Anja, Liebal, Kristin, & Tomasello, Michael. (2012). Eighteen-month-olds learn novel words through overhearing. *First Language, 32*(3), 385–397. https://doi.org/10.1177/0142723711433584

Gilkerson, Jill, Richards, Jeffrey A., Warren, Steven. F., Montgomery, Judith. K., Greenwood, Charles. R., Kimbrough Oller, D., Hansen, John H. L., & Paul, Terrance. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology, 26*(2), 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

Jackson-Maldonado, Donna, Thal, Donna. J., Marchman, Virginia. A., Newton, Tyler, Fenson, Larry, & Conboy, Barbara. (2003). *MacArthur Inventarios del Desarrollo de Habilidades Communicativas: User's Guide and Technical Manual*. Brookes.

Ochs, Elinor, & Schieffelin, Bambi. B. (1994). Language acquisition and socialization: Three developmental stories and their implications. In R. A. Schweder & R. A. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 276–322), Cambridge, UK: Cambridge University Press.

Ramírez-Esparza, Nairán, García-Sierra, Adrián, & Kuhl, Patricia K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science, 17*(6), 880–891. https://doi.org/10.1111/desc.12172

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

Shneidman, Laura. A., Arroyo, Michelle. E., Levine, Susan. C., & Goldin-Meadow, Susan. (2013). What counts as effective input for word learning? *Journal of Child Language, 40*(3), 672–686. https://doi.org/10.1017/S0305000912000141

Shneidman, Laura A., & Goldin-Meadow, Susan (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science, 15*(5), 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Tomasello, Michael. (1995). Joint attention as social cognition. In C. Moore, P. Dunham, J. Philip (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130), Hillsdale, NJ: Erlbaum.

Weisleder, Adriana, & Fernald, Anne (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143–2152. https://doi.org/10.3399/096016407782317928

Xu, Dongxin, Yapanel, Umit, & Gray, Sharmi. (2009). Reliability of the LENA TM Language Environment Analysis System in Young Children's Natural Home Environment. http://www.lenafoundation.org/wp-content/uploads/2014/10/LTR-05-2_Reliability.pdf

Proceedings of the 46th annual Boston University Conference on Language Development

edited by Ying Gong
and Felix Kpogo

Cascadilla Press Somerville, MA 2022

ISSN 1080-692X
ISBN 978-1-57473-077-7 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com