UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Examining a developmental pathway of early word learning: From qualitative characteristics of parent speech, to sustained attention, to vocabulary size

Permalink

https://escholarship.org/uc/item/2p82d0sp

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Peters, Ryan E.

Yu, Chen

Publication Date

2020

Peer reviewed

eScholarship.org Powered by the California Digital Library University of California

Examining a developmental pathway of early word learning: From qualitative characteristics of parent speech, to sustained attention, to vocabulary size

Ryan E. Peters (ryerpete@iu.edu)
Chen Yu (chenyu@indiana.edu)
Department of Psychological and Brain Sciences
Indiana University, Bloomington, IN 47405

Abstract

The quality of parent speech has been argued to impact child language growth above and beyond quantity. One potential mechanism tying online experience to long-term vocabulary development is sustained attention to targets of parent speech. We recruited thirty-five parent-toddler dyads to participate in free toy play while wearing head-mounted eye trackers. Parent speech was categorized based on its referential nature, syntax, and communicative intent. Parent referential speech positively related to both vocabulary size and online patterns of sustained attention. Speech categorized based on communicative intent also showed relations with vocabulary size and sustained attention, but specific types of speech impacting each differed. These results support the hypotheses that qualitative characteristics of parent speech relate to both long-term language growth and online sustained attention and provide tentative evidence for the broader hypothesis that sustained attention is the mechanism tying online experience to long-term language growth.

Keywords: eye-tracking; language input; eye-tracking; parent-child interaction; sustained attention; vocabulary size

Introduction

Both quantity and quality of language input matter for early language learning. Hart and Risley (1995) showed that children from low SES families hear significantly fewer words than those from higher SES families, which in turn predicts negative long-term outcomes including lower vocabulary scores (e.g., Fernald, Marchman, & Weisleder, 2013; Hoff, 2013). However, researchers have more recently focused on how the *quality* of parent-child conversations relates to language development (e.g., Hirsh-Pasek, et al., 2015; Hoff, 2006; Rowe, 2012). Work characterizing the quality of Child Directed Speech (CDS) often focuses on the content or communicative functions of parents' utterances. For example, research on parental responsiveness has shown that language growth is positively related to CDS with a higher frequency of utterances that either refer to objects children are attending to or provide affirmation of their communicative acts (Masur, Flynn, & Eichorst, 2005; Tamis-LeMonda, Bornstein, & Baumwell, 2001). Work has shown that a higher frequency of conversation eliciting questions, such as wh-questions, also positively relates to subsequent language development (Hoff-Ginsberg, 1985, 1986; Rowe, Leech, & Cabrera, 2017). In contrast, a higher frequency of directive utterances negatively relates to language growth (Barnes, Gutfreund, Satterly, & Wells, 1983; Hughes, Dote-Kwan, & Dolendo, 1999; Nelson, 1973; Tomasello & Todd, 1983; Newport, Gleitman, & Gleitman, 1977). While this line of work has clarified relations between quality of parent speech and language growth, it remains unclear what effects such characteristics of parent speech have on online behaviors that relate to long-term growth.

A growing body of work highlights children's sustained attention to objects during social interactions as one key mechanism tying online experience to long-term language growth (Yu, Suanda, & Smith, 2019). Crucially, recent work indicates that sustained attention is not solely an individual property of a child, but rather is socially malleable (Suarez-Rivera, Smith, & Yu, 2019; Yu & Smith, 2016). Yu and Smith (2016) showed that, in a free toy play setting, when parents visually attended to the same objects as their children (i.e., engaged in joint attention), the children extended their duration of visual attention. Building on this result, Suarez-Rivera and colleagues (2019) demonstrated that episodes of joint attention are typically embedded within a multimodal suite of behaviors – of which parent speech is the most influential in extending infant attention.

While the above work did not explore whether qualitative characteristics of parent utterances impact their influence on children's sustained attention, indirect evidence for this hypothesis comes from recent work by Chang and Deák (2019). They showed that parent speech directed to 12-month-olds during free object play differentially influenced their infants' patterns of looking and object handling behaviors, both coded using third-person videos. Specifically, they found that speech with object descriptions was associated with switching visual attention to objects, while speech that aimed to direct attention was associated with disengagement from objects. One novel feature of this study was the two-level scheme used to characterize utterances. At the first level utterances were coded based on their syntax, while at the second level utterances were coded based on their topic or communicative function.

In the current study, we build on this previous work to explore the hypotheses that the qualitative characteristics of children's language input from their parents relates to 1) their concurrent vocabulary size and 2) their patterns of online sustained attention in a naturalistic setting. To test these hypotheses, we recruited parent-child dyads to participate in free toy play sessions in the lab while wearing head-mounted eye-trackers, during which we recorded parent speech and

©2020 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

collected gaze data using head-mounted eye-trackers. To characterize parent speech, we built on the coding framework in Chang and Deák (2019) to include an initial level at which we code whether utterances are referential, and if so to which objects they refer. Finally, we collected parent reports of their children's productive vocabularies to use as a measure of concurrent vocabulary size.

Methods

Thirty-five toddlers (mean age = 18.92 mos [12.3-25.3]; female = 16) and their parents participated in a study on naturalistic parent-child interactions during free toy play. Data for 4 additional dyads was collected but excluded from the current analyses due to non-transcribable speech (n=1) or missing audio (n=3).

Data collection and gaze coding

Parent-child dyads played on the floor in a playroom for an average of 7.19 min [range 3.93-11.64]. A set of 24 toys were initially spread randomly on the floor. Parents were told to play as they would at home, but to keep the child sitting on the floor due to the cable attaching the eye-tracker to the computer.

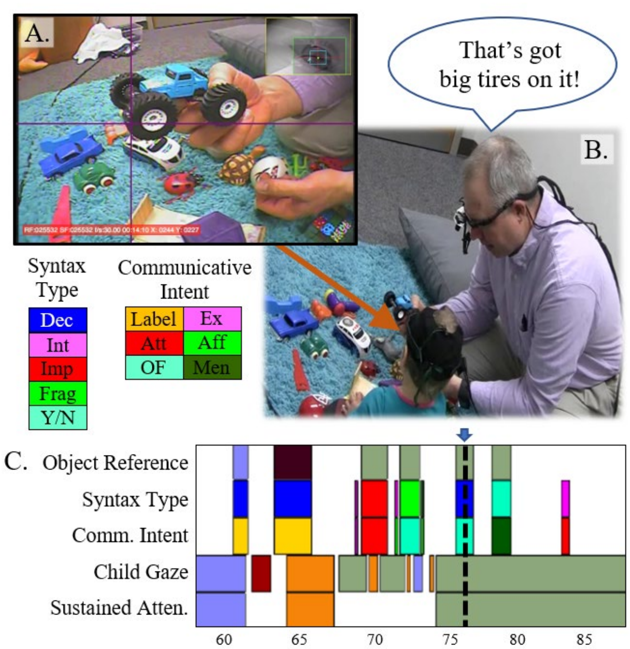

Figure 1. Experimental setup showing (A) first-person child view, (B) third-person view and (C) data stream visualization capturing a moment of sustained attention on the truck. Crosshairs in first-person view estimate point of fixation. C-units missing Object Reference are non-referential. Syntax types shown include declarative (Dec), interjection (Int), imperative (Imp), fragment (Frag), and yes/no question (Y/N). Communicative intent types include label, exclamation (Ex), attention (Att), affirmation (Aff), object feature (OF), and mental (Men).

Both parent and child wore head-mounted eye-trackers (Positive Science LLC) which used scene cameras of the participants' first-person perspectives on the forehead (visual field 108°) and infrared cameras pointed at their right eyes to record fixations (both sampled at rate of 30Hz). The child's eye-tracker was on a hat, while the parent's eye-tracker was attached to a pair of glasses.

The eye-tracker calibration procedure involved placing a large board that had 15 lights and produced sounds approximately 30 cm away from the infant. For each of the 15 lights, the light was lit up until both the parent and child shifted their gaze to the location before moving on to the next light. Researchers monitored the experiment from an adjoining room. If either of the eye cameras was moved during the experiment, the researchers reentered the room, adjusted the camera, and performed a shortened version of the calibration procedure. Additional cameras in the room captured third-person views.

Following the play session, parents completed either the infant or toddler version of the MacArthur-Bates Communicative Development Inventory (MBCDI, Fenson, 2002), a parent questionnaire designed to assess children's productive vocabularies. Eight parents were unable to complete the questionnaire due to time constraints. Using the questionnaires, we then calculated each child's vocabulary percentile relative to a nationally representative sample (Dale & Fenson, 1996) to allow for comparison across ages (mean percentile = 40.56 [5-90]).

After the experiment, all eye-tracking and third-person videos were synchronized, and software was used to generate crosshairs on the parent and toddler first-person views estimating fixation locations. These videos were then used to manually code 25 regions of interest (ROI; 24 toys and social partner's face) using an in-house program.

Speech coding

Audacity was used to segment and transcribe parent speech. First, we divided parent speech into utterances separated by gaps of at least 400 ms, following standard practice in the field (e.g., Yu & Smith, 2016; Suarez-Rivera et al., 2018). After transcribing the resulting utterances, they were further divided into C-units (Communication units), following the SALT (Systematic Analysis of Language Transcripts, Miller & Chapman, 1985) conventions, with two minor differences. C-units were determined using both grammatical and phonological properties of the parent's speech. Specifically, an utterance was divided into separate C-units if it contained multiple independent clauses and/or had clear boundaries in phonological contours. The two differences between the current coding scheme and the SALT conventions was the additional separation of interjections (e.g., "Oh!", "No", "Look!") and tags (e.g., "Right?", "Don't you think?) into their own C-units.

Transcripts (and if necessary third- and then parent first-person videos) were used to code the qualitative characteristics of each C-unit at three levels: referential, syntax, and communicative intent. While each C-unit was coded at all three levels, categories within levels were mutually exclusive.

At the referential level, we determined whether C-units explicitly referred to experimental objects, unrelated objects or people via (proper) nouns (e.g., "that is a turtle", "here is another car") or pronouns (e.g., "what is it", "turn it up this way"). For C-units that referred to experimental objects we also determined which object(s) they referred to. All remaining C-units were categorized as non-referential.

At the syntax level, we determined whether the syntax of C-units matched one of four sentence types: imperatives (e.g., "Keep it there", "Now hold the handle", "Can you make the turtle walk"), yes/no questions (e.g., "Is he walking too?", "Do you like my hat?"), wh-questions (e.g., "What is it?", "Who should we put this hat on?"), or declaratives (e.g., "You are sawing", "I have a hat like you do"). If not, we then checked whether the C-unit consisted of language or vocal play (e.g., "Vroom vroom!", sound of a siren), was an interjection (e.g., "Oh!", "Okay"), was a sentence tag (e.g., "Right?", "I guess"), or, finally, was a sentence fragment (e.g., "Dog", "Or that elephant").

At the communicative intent level, we categorized C-units based on their primary topic or communicative function. First, if a C-unit mainly conveyed information about an object or person we determined whether the information fell into one of four categories: object feature (e.g., "that is white", "that turtle is kind of heavy"), action or activity (e.g., "bunny fell down", "she is sleeping in her bed"), label (e.g., "turtle", "is that a ladybug?"), or object sound (e.g., "Vroom vroom", "Squeak squeak"). If a C-unit did not mainly convey information, we next determined whether it sought to direct or comment on the child's attention (e.g, "Look at this one", "What else do you see?"); consisted of routine or ritualized language (e.g., "Hello", "The incy wincy spider went up the water spout"); was an exclamation (e.g., "Wow!", "Oh!"); was an affirmation (e.g., "Yes", "Good boy") or negation (e.g., "No", "Bad boy"); referred to the number of items or actions (e.g., "Two cars", "One, two, three…"); or, finally, referred to thoughts, memories, feelings or imaginary things – or in other words the mental state of the child (e.g., "Do you love that?", "Do you remember when you played with a football with Daddy?").

Analyses

We begin by presenting descriptive statistics for the frequencies with which parents produce C-units categorized at each of the three levels of speech coding. We then conduct two sets of analyses, one addressing each of our two hypotheses. We present descriptive statistics for all C-unit types and include all C-units in the analyses of speech categorized at the referential level. However, for analyses of speech categorized using syntax and communicative intent, we focus on the types of speech for which parents produced on average at least one referential utterance per minute.

Results

Descriptive statistics

We present the frequencies of parents' production of C-units for each of the types in table 1.

Table 1. Frequency of parent speech by C-unit type.

Variable	Frequency (pm): mean (range), SD	
	Referential	Non-Referential
Total	10.5 (5.2-17.5), 3	14.9 (6.7-32.4), 6.1
Syntax type		
Declarative	4 (1-8.6), 1.8	1.5 (0.2-3.9), 1.0
Fragment	1.9 (0.3-4.7), 0.9	1.6 (0-8.4), 1.5
Y/N question	1.8 (0-4.4), 0.9	0.7 (0-2.4), 0.5
Imperative	1.3 (0-3.8), 1	1.5 (0.2-3.6), 0.9
Wh-question	1.3 (0.4-3.2), 0.7	0.6 (0-1.8), 0.4
Interjection	0.1 (0-0.4), 0.1	7 (3.3-18.5), 3.4
Tag	0.1 (0-0.3), 0.1	0.3 (0-1.3), 0.3
Vocal play	0 (0-0.2), 0	1.7 (0-7.6), 1.5
Comm. intent		
Action/activity	2.8 (0.5-6.2), 1.5	2.0 (0-8.7), 1.6
Label	2.5 (1-6.2), 1.1	0.3 (0-2.0), 0.4
Object feature	2.2 (0.2-5.8), 1.4	0.5 (0-2.2), 0.5
Attention	1.3 (0.2-3.4), 0.9	2.1 (0-6.7), 1.5
Mental	1 (0-2.4), 0.6	0.8 (0-2.3), 0.6
Routine/ritual	0.2 (0-2.2), 0.5	1.0 (0-3.8), 0.9
Affirmation	0.1 (0-0.8), 0.2	3.2 (0.9-10.0), 2.0
Number	0.1 (0-0.6), 0.2	0.2 (0-3.1), 0.6
Negation	0 (0-0.2), 0.1	0.5 (0-1.7), 0.4
Object sound	0 (0-0.1), 0	1.6 (0-5.1), 1.3
Exclamation	0 (0-0), 0	2.4 (0.5-11.4), 2.2

Note. Analyses for syntax and communicative intent levels of parent speech coding were restricted to those C-unit types with referential frequency greater than or equal to 1 per minute (in bold).

Relations between quality of speech and concurrent vocabulary size percentile

Here we address our first hypothesis that the qualitative characteristics of speech in children's language environments relates to their vocabulary size. For each level of parent speech coding, we construct simple linear models to determine whether the frequency with which parents produce different types of C-units predicts their children's concurrent vocabulary size percentile.

First, to determine whether the frequency of referential versus non-referential C-units within children's language environments relates to their vocabulary size we constructed a model predicting concurrent vocabulary size percentile as a function of the frequency with which parents produced referential and non-referential C-units during the toy play session. [1] The model was significant, $F(2, 24)=5.907$, $p<.01$,

[1] Models including age as a predictor had nearly identical results, with age being non-significant.

R^2_{adj}=.274, with referential frequency positively relating to vocabulary size (model coefficients presented in table 2).

Table 2. Results for multivariate models predicting vocabulary percentile as a function of production frequencies for 1) referential vs non-referential C-units, 2) referential C-units for five syntax types, and 3) referential C-units for five types of communicative intent

Variable	Linear Model	
	Coefficients	95% CI
Referential Type Model		
Intercept	-5.51	[-39.68, 28.67]
Frequency (pm)		
Referential	5.18 **	[1.99, 8.38]
Non-referential	-0.47	[-2.17, 1.23]
(Referential) Syntax Type Model		
Intercept	-2.02	[-38.93, 34.90]
Frequency (pm)		
Declarative	4.73	[-1.13, 10.59]
Fragment	-0.92	[-14.47, 12.63]
Imperative	8.56	[-2.71, 19.84]
Wh-question	9.21	[-6.50, 24.92]
Y/N question	1.22	[-11.92, 14.36]
(Referential) Communicative Intent Model		
Intercept	-15.97	[-55.99, 24.06]
Frequency (pm)		
Action/activity	12.27 **	[4.82, 19.71]
Attention	-2.73	[-19.28, 13.82]
Label	9.70	[-5.89, 25.29]
Mental	-1.86	[-22.36, 18.63]
Object feature	2.67	[-4.55, 9.89]

Note. N=27. CI = confidence interval. ** $p < .01$.

Next, we constructed a model predicting vocabulary size as a function of the frequency at which parents produced C-units for each of five syntax types with average referential production frequencies of at least 1 pm. The model was not significant, $F(5, 21)$=2.319, p=.08, R^2_{adj}=.20 (model coefficients presented in table 2). Indicating the frequencies with which parents produce C-units of these five syntax types do not robustly relate vocabulary size.

Finally, we constructed a model predicting vocabulary size percentile as a function of the frequency at which parents produced C-units for each of the five communicative intent types with average referential production frequencies of at least 1 pm. The model was significant, $F(5, 21)$=3.653, p<.05, R^2_{adj}=.34, with the frequency of action/activity C-units positively relating to vocabulary size (model coefficients presented in table 2).

Effects of quality of speech on sustained attention

Here we address our second hypothesis that the qualitative characteristics of parent speech relates to their children's patterns of online sustained attention – defined here as gaze to an object lasting at least 3 s. We compare participant-level mean proportions of sustained attention to relevant objects in the 3 second time window starting at C-unit onset, across C-unit types within each level. We use ANOVA analyses of linear mixed effects models (with random intercepts for subject) and the Tukey method for correction of p-values for multiple comparisons. We constructed models using the lme4 package (Bates et al., 2015) and analyzed them using the emmeans package (Length, 2019) in R (R Core Team, 2019).

First, we explore how the referential nature of parent speech relates to their children's simultaneous patterns of sustained attention. While referential utterances have clear targets (i.e., the objects the C-units explicitly refer to), non-referential utterances do not have clear targets. We decided one fair comparison is to infer the target based on the most recent referential C-unit.[2] Thus, we constructed a model comparing proportions of sustained attention directed to targets for referential C-units versus the proportion of sustained attention directed to the inferred targets for non-referential C-units. The model was significant, $F(1, 34)$=82.69, p<.001, indicating the proportion of sustained attention directed to targets of referential C-units (mean=.385, CL=[.341, .429]), was greater than for the inferred targets of non-referential C-units (mean=.261, CL=[.217, .305]), $t(34)$=9.093, p<.001 (Figure 2).

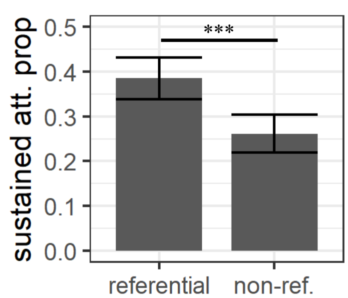

Figure 2. Comparison of proportion of sustained attention directed to targets for referential and inferred targets for non-referential C-units. Error bars show 95% confidence intervals. *** p < .001.

Next, to determine whether the syntax type of referential C-units relates to patterns of online sustained attention we modeled mean proportions of sustained attention directed to the targets of referential utterances as a function of the same five syntax types. The model was not significant, $F(4,133.88)$=.714, p=.58, indicating the syntax type of referential C-units does not robustly relate to the proportion

[2] One arguably fairer approach is to compare the sums of sustained attention directed to all experimental objects for referential and non-referential C-units. The results of such an analysis are nearly identical to those presented for the comparison using inferred targets for non-referential C-units.

of sustained attention concurrently directed to the target object of the C-unit (Figure 3).

Finally, to explore whether the communicative intent type of referential speech relates to patterns of online sustained attention directed to speech targets, we modeled mean proportions of sustained attention directed to the targets of referential utterances as a function of the same five communicative intent types. The model was significant, $F(4,133.78)=5.042$, $p<.001$, indicating the communicative intent of a C-unit does relate to the proportion of sustained attention concurrently directed to the target object (Figure 4). Post-hoc pairwise analyses indicated the mean proportion of sustained attention directed to target objects of C-units primarily conveying information about object features (mean = 0.499, CL=[.430, .567]) was significantly greater than for those conveying information about actions or activities (mean = .353, CL=[.285, .422]), $t(134)=-3.536$, $p<.01$; object labels (mean = .357, CL=[.288, .426]), $t(134)=-3.445$, $p<.01$; and those C-units referring to the mental state of the child (mean = .337, CL=[.267, .408]), $t(135)=-3.856$, $p<.01$.

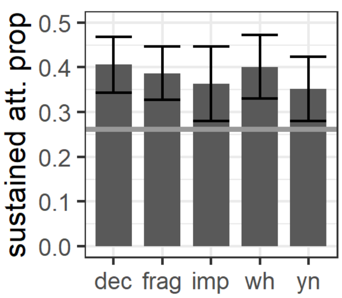

Figure 3. Comparisons of proportion of sustained attention directed to the targets of referential utterances for five syntax types: declaratives (dec), fragments (frag), imperative (imp), wh-questions (wh), and yes/no questions (yn). Error bars show 95% confidence intervals. The horizontal gray line shows the mean proportion of sustained attention directed to inferred targets of non-referential C-units.

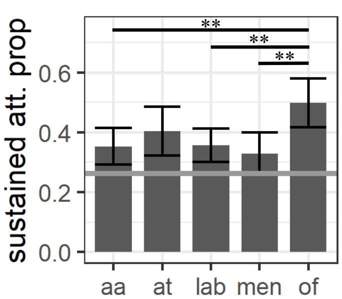

Figure 4. Comparisons of proportion of sustained attention directed to the targets of referential utterances for five communicative intent types: action/activity (aa), attention (at), labels (lab), mental phenomena (men), and object features (of). Error bars show 95% confidence intervals. The horizontal gray line shows the mean proportion of sustained attention directed to inferred targets of non-referential C-units. ** $p < .01$.

Discussion

The current study was designed to explore the hypotheses that the qualitative characteristics of parent speech relate to (1) the language growth of their children and (2) their patterns of online sustained attention to objects during free toy play. To test these hypotheses, we brought parent-child dyads into the lab to participate in play sessions while wearing head-mounted eye-trackers, during which we recorded parent speech and collected gaze data. We characterized parent speech using a three-level coding scheme (with levels categorizing speech segments based on whether they were referential, their syntax, and their communicative intent), and we measured child vocabulary size using a parent questionnaire of their children's productive vocabularies. Overall, the results provide support for both hypotheses.

There were two significant results relevant to the hypothesis that the qualitative characteristics of parent speech relate to longer term language growth. First, we found that children's vocabulary size percentile is robustly predicted by the frequency (pm) with which parents produced referential (but not non-referential) speech during the toy play sessions. Second, we found that vocabulary percentile is also significantly related to the frequency with which parents produced speech that conveyed information about actions and activities, but not speech that fell into other communicative intent categories.

Likewise, there were two significant results relevant to the second hypothesis that the qualitative characteristics of parent speech relates to their children's online patterns of sustained attention to relevant objects. First, we found that children showed a boost in sustained attention (though not specifically to target objects) upon hearing referential speech when compared with non-referential speech. Second, we found that children robustly show greater sustained attention to target objects upon hearing speech conveying information about object features when compared with speech conveying information about actions or activities, labels, or the mental state of the child.

It is also worth considering our null results for the analyses of speech categorized based on syntax, in particular for imperatives and wh-questions. While the work discussed in the introduction highlighted the potential for a negative impact of imperatives (Barnes, et al., 1983; Hughes, et al., 1999; Nelson, 1973; Tomasello & Todd, 1983; Newport et al., 1977), our null result is not entirely surprising given mounting evidence that the impact of imperatives/directives depends on whether they follow or attempt to redirect the attention of the child (Masur et al., 2005; Masur, Flynn, & Lloyd, 2013). Likewise, recent work by Rowe and colleagues (2017) highlighted the potential for a positive impact of wh-questions, however the discrepancy in our results is once again not entirely unexpected given their work focused on older children's interactions with their fathers.

The alignment in results showing a positive impact of referential speech both on online sustained attention and more long-term language growth provides indirect evidence that sustained attention is the mechanism via which qualitative characteristics of parent speech influence more long-term language growth. However, this story is complicated by number of issues. First and most crucially, the current study explored relations between parent speech and *concurrent* vocabulary size, making it impossible to disentangle the likely bi-directional relations between these variables. Second, for speech categorized based on communicative intent there is a discrepancy in the results, where frequency of speech conveying information about actions and activities positively relates to vocabulary size while in contrast speech conveying information about object features positively relates to patterns of online sustained attention. One potential explanation for this discrepancy lies in the fact that in the current study we have ignored the larger discourse patterns within which any individual speech segment occurs – which has been shown to impact child gaze patterns (Chang & Deák, 2019; Suanda, Smith, & Yu, 2016). Thus, in future work we plan to address these limitations.

Conclusion

Previous work has highlighted how the quality of parent speech, above and beyond quantity, impacts children's language growth. We built on these findings by exploring how parents' speech – coded based on their referential nature, syntax, and communicative intent – related to child vocabulary size and patterns of sustained attention during free toy play. The quality of parent speech matters, and alignment between online and long-term results provides tentative evidence that child sustained attention plays a mechanistic role in tying the two together.

Acknowledgements

This research was supported in part by NICHD grants R01HD074601 and R01HD093792.

References

Barnes, S., Gutfreund, M., Satterly, D., & Wells, G. (1983). Characteristics of adult speech which predict children's language development. *Journal of child language*, 10(1), 65-84.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, 12(1), 2.

Chang, L. M., & Deák, G. O. (2019). Maternal discourse continuity and infants' actions organize 12-month-olds' language exposure during object play. *Developmental science*, 22(3), e12770.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28(1), 125-127.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Fenson, L. (2002). *MacArthur Communicative Development Inventories: User's guide and technical manual.* Paul H. Brookes.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental science*, 16(2), 234-248.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., ... & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological science*, 26(7), 1071-1083.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental review*, 26(1), 55-88.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental psychology*, 49(1), 4.

Hoff-Ginsberg, E. (1985). Some contributions of mothers' speech to their children's syntactic growth. *Journal of Child Language*, 12(2), 367-385.

Hoff-Ginsberg, E. (1986). Function and structure in maternal speech: Their relation to the child's development of syntax. *Developmental Psychology*, 22(2), 155.

Hughes, M., Dote-Kwan, J., & Dolendo, J. (1999). Characteristics of maternal directiveness and responsiveness with young children with visual impairments. Child: *Care, Health and Development*, 25(4), 285-298.

Lenth, R. (2019). *emmeans: estimated marginal means, aka least-squares means*. R package v. 1.4.3.01.

Masur, E. F., Flynn, V., & Eichorst, D. L. (2005). Maternal responsive and directive behaviours and utterances as predictors of children's lexical development. *Journal of Child Language*, 32(1), 63-91.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the society for research in child development*, 1-135.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and noneffects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109–150). Cambridge: Cambridge University Press.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. v. 3.6.2.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child development*, 83(5), 1762-1774.

Rowe, M. L., Leech, K. A., & Cabrera, N. (2017). Going beyond input quantity: Wh-questions matter for toddlers' language and cognitive development. *Cognitive science*, 41, 162-179.

Slone, L. K., Abney, D. H., Borjon, J. I., Chen, C. H., Franchak, J. M., Pearcy, D., ... & Yu, C. (2018). Gaze in Action: Head-mounted Eye Tracking of Children's

Dynamic Visual Attention During Naturalistic Behavior. *JoVE (Journal of Visualized Experiments)*, (141), e58496.
Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental psychology*, 55(1), 96.
Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child development*, 72(3), 748-767.
Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First language*, 4(12), 197-211.
Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, 26(9), 1235-1240.
Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental science*, 22(1), e12735.