The Roles of Speech Complexity and Pointing Gesture in Guiding Children's Attention During Shared Book Reading

Yayun Zhang[1]**, Jennifer Sander**[1,2]**, Thalassia Kontino**[3]**, Caroline F. Rowland**[1,4]**, Chen Yu**[5]

Yayun.Zhang@mpi.nl, Jennifer.Sander@mpi.nl, t.kontino@uu.nl
Caroline.Rowland@mpi.nl, chen.yu@austin.utexas.edu

[1]Max Planck Institute for Psycholinguistics, The Netherlands, [2]Max Planck School of Cognition, Germany
[3]Utrecht University, The Netherlands, [4]Radboud University, The Netherlands,[5]University of Texas at Austin, USA

Abstract

Shared book reading is widely acknowledged for its positive impact on language development, as it exposes children to complex linguistic structures not typically encountered in daily conversation. However, the mechanisms through which shared reading supports language acquisition remain less well understood. This study investigates the effects of speech complexity and gesture use on children's real-time word learning from books. Using a dual head-mounted eye-tracking paradigm, we assessed gaze dynamics in 18- to 24-month-old children during naturalistic book reading with their parents. Our findings indicate that while parental speech is rich in linguistic diversity, children at this age exhibit a preference for simpler sentence structures. Simpler sentences and imperatives, particularly when paired with child gestures, appear to capture children's attention most effectively. This study emphasizes the interplay between speech complexity, gesture, visual attention, and word learning, demonstrating that multimodal input plays a critical role in facilitating language acquisition.

Keywords: shared book reading, word learning, visual attention, sentence complexity, gestures, word-object mapping

Introduction

Shared book reading is widely recognized as a key activity for fostering language development in young children because it combines rich linguistic input with meaningful social interaction. Through this activity, children are exposed to diverse vocabulary, complex sentence structures, and interactive dialogue that go beyond typical everyday conversations. However, despite extensive evidence highlighting the benefits of shared book reading for early language acquisition, little is known about the specific mechanisms by which children learn new words in these interactions. This study addresses this gap by exploring how complexity of parental speech and gesture use during book reading influence children's real-time attention during language learning.

Speech complexity in shared book reading

Children's books expose young children to more advanced linguistic structures than typical child-directed speech. Previous studies (Montag et al., 2015; Montag, 2019) have shown that texts in children's books feature a greater variety of unique words and more frequent use of passive sentences and relative clauses. Similarly, Cameron-Faulkner and Noble (2013) found a higher prevalence of canonical and complex sentence constructions in books, which is rare in everyday child-directed speech. In addition to the linguistic richness found in the text of children's books, shared book reading also fosters extra-textual talk that includes structurally rich language, distinct from the simpler language used during everyday interactions such as mealtime, dressing, or play (Noble et al., 2018). During shared reading, parents often adapt their language to include diverse syntactic structures and vocabulary, exposing children to a broader range of linguistic inputs. This includes introducing new words, using varied verb tenses, and providing detailed explanations. Parents also engage in dialogical reading (Whitehurst et al., 1988) by asking open-ended questions and encouraging active participation, further enriching the linguistic environment (Lingwood et al., 2023). The linguistic richness, both in the book text and extra-textual talk during shared book reading, introduces children to structurally complex language, significantly contributing to their language development.

Despite the known benefits of shared book reading, how children process and learn from such input remains unclear. Research on information processing in other areas shows that stimulus complexity can impact learning. For instance, Kidd et al. (2012) measured 8-month-olds' responses to visual sequences of differing complexity and found that they were most attentive to visual events of intermediate complexity, showing a "Goldilocks effect." Extending this to language, Foushee et al. (2021) studied how 4-6-year-olds process different levels of linguistic complexity. Children listened to two versions of the same story: one with familiar words (SIMPLE) and one with unfamiliar words (COMPLEX). Younger children preferred SIMPLE speech, while older children paid more attention to COMPLEX speech, suggesting that children learn best from input suited to their learning capacity. These findings highlight a connection between speech complexity and how much information children process, raising questions about how infants process language input of varying complexity during parent-child book reading interactions.

Pointing gestures in shared book reading

In addition to rich speech input, non-verbal input, such as pointing gestures, has also been found to play a supportive role in language development during book reading (Murphy, 1978; Rohlfing et al., 2015). A longitudinal study by Rohlfing et al. (2015) found that pointing and labeling occur more frequently during joint book-reading interactions with 14-month-olds than during free play. Moreover, maternal use of gestures is positively correlated with children's receptive and productive vocabulary at 14 months, and mothers' use of gesture-speech combinations when the child is 14 months old predicts the child's productive vocabulary at 24 months. By pointing, parents can draw a child's attention to specific referents, narrowing the range of possible meanings and allowing children to identify word-object mappings more easily (Rowe et al., 2008). Children also use their own pointing gestures to signal readiness for learning, prompting caregivers to provide timely object labels (Church et al., 2004; Valenzeno et al., 2003).

Most research to date has relied on controlled studies using either artificial or naturalistic stimuli; however, we still have limited understanding of how children process complex linguistic input during spontaneous parent-child interactions in real time. To address this gap, we investigate the dynamic interplay between *speech complexity*, *pointing gestures*, *attention*, and *word learning* within the semi-naturalistic context of shared book reading.

Current study

We used a head-mounted eye-tracking paradigm, which allowed us to measure participants' gaze directions in real time, revealing the exact timing and dynamics of attention shifts and how children engage with and process linguistic and gestural input. It has been shown that the ability to attend to the right information is crucial for language learning because it enhances processing at the attended location (e.g., Mundy et al., 2007). For example, using the Preferential Looking Paradigm, Golinkoff et al., 1987 showed that even preverbal infants looked longer at scenes matching spoken sentences, indicating early noun and verb comprehension. Fernald et al., 1998 found that older infants processed speech faster using the Looking-While-Listening paradigm. Recent eye-tracking studies reveal how infants use visual attention to resolve referential ambiguity in word learning (e.g., Yu and Smith, 2013). Differences in gaze patterns between strong and weak learners suggest that real-time attention supports statistical learning and predicts later language outcomes. Therefore, knowing how children selectively attend to objects when hearing object names is critical for understanding how correct word-object mappings may be established for word learning.

Using moment-by-moment gaze and gestural data from both parents and children, as well as parent speech, This study aims to answer the following questions: (1) How linguistically complex and diverse is parent speech during object-naming moments in book reading? (2) How does children's visual attention to the target change in response to naming events of varying speech complexities, and does this effect interact with age and vocabulary size? (3) Do pointing gestures from either the parent or the child influence children's attention to the target during naming moments of different types? By examining the interplay between parent speech, gestures, and child attention, this study could offer valuable insights into the mechanisms that support vocabulary development in naturalistic book-reading contexts.

Method

Participant

The participants were 16 parent-child dyads who resided in Midwest, U.S.A. All children (12 females) were between the ages of 18 and 24 months (*M* = 19.03, *SD* = 1.6, *Min* = 18, *Max* = 24.4). Twelve additional dyads participated but contributed no or limited eye-tracking data due to children's unwillingness to wear the head camera equipment. Procedures in this study were approved by the Human Subjects and Institutional Review Boards at Indiana University.

Materials

We selected five commercially available picture books: *I Went Walking* (1989), *Goodnight, Gorilla* (1994), *Let's Go Visiting* (1997), *Sammy the Seal* (2005), and *I Am a Little Lion* (1994). These books differ in story content and illustration style but share a clear narrative structure centered around a main character. As they are designed for beginner readers, some contain minimal text or none at all. To maintain consistency, we removed all written text and instructed parents to create their own narratives based on the illustrations.

Previous research suggests that when reading to children, parents often go beyond the written text, engaging in dynamic dialogic reading (Whitehurst et al., 1988). Self-generated storytelling introduces a richer vocabulary that extends beyond the book's content to include the child's personal experiences (Deckner et al., 2006; Dickinson et al., 2012; Ninio and Bruner, 1978). We anticipated that this approach would encourage more spontaneous speech, allowing us to examine variations in parent speech and provide a clean baseline for future exploration comparing parent reading with children's books with texts

Experimental setup

During the experiment, the child and parent sat next to each other at a table (61cm x 91cm x 64cm). Infants sat in a customized highchair that supported sitting stability, and parents sat on the floor. A bookstand was used to hold the book at a consistent 60° angle and roughly 10cm away from the edge of the table. This setup allowed parents to freely interact with their children while avoiding displacement of the eye-tracking devices due to voluntary head movements. Both participants wore head-mounted eye trackers from Positive Science, LLC (Franchak et al., 2010; Yu and Smith, 2013). Each eye-tracking system included an infrared eye camera and a scene camera. The eye camera, mounted on the head, recorded the participant's right eye, while the scene camera captured their first-person view with a 90° field of vision. Although narrower than the human visual field (about 180°), it covered the book-page area needed to track gaze. The system recorded both the egocentric-view video and gaze direction at 30 Hz (Figure 1). Parent speech was captured using a microphone built into the parent's eye tracker. Three additional high-resolution cameras were also added on two walls and

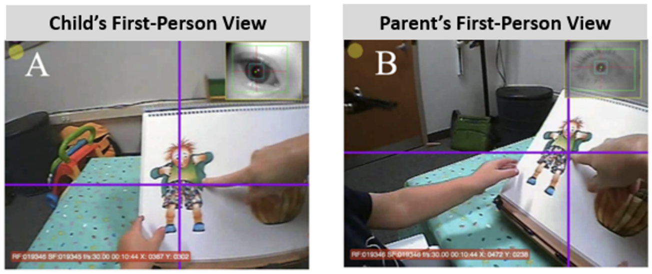

Figure 1: Experimental setup. (A) Child's first-person view and (B) parent's first-person view, with eye images superimposed in the upper-right corner. The purple crosshair indicates the gaze location for each participant.

the ceiling to capture the interaction from three third-person views, allowing accurate coding of pointing behaviors.

Procedure

We first fitted the parent with the eye-tracking gear. Once both the parent's and child's eye trackers were properly positioned, we calibrated the system by instructing them to look at the four corners and the center of a book page. Parents were then instructed to read books to their children as they naturally would for 15 minutes. They were not required to follow any specific order or finish all the books. Parents were asked to place the book on the bookstand while reading and to maintain the original sitting arrangement as much as possible. They were unaware that the study focused on word learning and were not instructed to name objects. Once the reading session began, the experimenters left the room and monitored the interaction remotely. Lastly, parent reports of their children's receptive and productive vocabularies were collected using the standardized MacArthur-Bates Communicative Development Inventory (MCDI, Fenson, 2002).

Corpus

In total, we collected 45 book-reading sessions with good eye-tracking data from 16 parent-child dyads. This equals to 157 minutes of usable video data. On average, each book was read 9 times. Each dyad contributed 2-5 books. Parents and children spent about 3.49 minutes per book. The shortest single-book interaction lasted 1.19 minutes and the longest one lasted 9.38 minutes.

Data processing

We synchronized and calibrated first-person view videos from both the parent and the child. Using calibrated videos with crosshairs superimposed on the videos indicating gaze directions (Figure 1), we manually annotated five variables: child and parent gaze, child and parent gestures and parent speech using the following coding scheme.

Gaze data. We first identified a list of region-of-interest (ROIs) for each book. All ROIs are whole objects on the page that can be named using concrete nouns. The number of objects varies page by page. The average number of objects on a page is 5.45 (*SD* = 2.83, *Min* = 2, *Max* = 15). This shows that book reading creates word-learning moments that are referentially uncertain as there are always multiple objects on a page when naming happens. Coders watched the calibrated first-person view videos and coded these ROIs frame by frame using an in-house program.

Gesture data. During book reading, the most frequent deictic gesture is pointing using hands or fingers. In our coding scheme, only pointing gestures directed at objects on the pages were coded. Coders examined videos from all angles to identify instances where parents or children pointed to objects. The duration of each gesture was calculated based on the time when the object being pointed to was clearly visible in the videos.

Speech data. Coders transcribed the parent's speech using only the audio recordings from the interactions. The speech was then categorized into sentence-level construction types and coded based on the taxonomy outlined by Cameron-Faulkner et al. (2003). Sentences that included at least one label for an object printed on the page (e.g., "What is the duck doing there?") were classified as naming utterances.

Using standard linguistic criteria, each naming utterance was categorized into one of four levels of sentence complexity, ranging from simple to more complex structures (Cameron-Faulkner et al., 2003). The first category, fragments, includes incomplete sentences or noun/nominal phrase constructions, such as "a cat" or "in the basket." The second category, copula sentences, comprises sentences using a copula verb to link the subject to a complement, such as "The boy is happy" or "The cat is black." The third category, subject-predicate sentences, involves simple sentences containing a subject and a predicate, such as "The boy loves cats" or "He puts the cat in the basket." Finally, complex sentences, the most advanced category, include at least two lexical verbs, such as "The boy loves hugging cats" or "I know you love hugging cats too."

We additionally categorized two types of sentences: Questions and Imperatives. Questions were identified as utterances featuring question syntax in the main clause, such as "Where is the cat?" or "Can you say cat?" Imperatives, on the other hand, were subjectless requests that prompted child action, such as "Look at the cat!" or "Find the cat!" All other sentences were categorized as statements.

Results

Complexity of parent naming events

Naming frequency Parents produced approximately 60 speech utterances per reading session (*M* = 59.78, *SD* = 26.57), which translates to about 17.54 utterances per minute (*SD* = 3.25 utterances/min), with each utterance lasting around 1.84 seconds (*SD* = 1.38 seconds). In total, parents produced 1930 object labels, which averaged to about 14 labels per minute (*M* = 13.51 labels/min, *SD* = 3.91 la-

Table 1. Selected CLAN measures.

	Examples	MLU Words	MLU Morphemes	Total type	Total token	TTR	Density
Fragment	Cat! A cat!	2.41 (0.65)	2.58 (0.71)	22.82 (12.87)	38.64 (29.88)	0.67 (0.16)	0.30 (0.11)
Copula	It's a cat. This is a cat.	4.68 (0.59)	5.11 (0.63)	30.25 (11.46)	77.16 (45.97)	0.46 (0.15)	0.36 (0.07)
Subject-predicate	The boy has a cat. Can you say cat?	5.67 (0.62)	6.32 (0.74)	57.78 (19.99)	145.73 (73.86)	0.44 (0.11)	0.39 (0.05)
Complex	The boy enjoys hugging the cat. I know you love hugging cats too.	7.58 (1.16)	8.46 (1.38)	39.18 (17.25)	74.80 (55.33)	0.62 (0.15)	0.45 (0.05)

bels/min), highlighting the fast-paced nature of book reading with frequent object-naming moments. Because the focus of the study was on learning object names, all subsequent analyses were centered around parent object-naming moments.

Naming complexity The plain text files containing the transcribed parent speech were converted to CHAT Transcription Format (.cha) files so that they could be processed using Computerized Language Analysis (CLAN) software (MacWhinney, 2014). The 'mor' function in CLAN was used to lemmatise and generate part-of-speech tagging for all words within the corpus. Assessment through KIDEVAL produces a profile across 41 variables including details on lexicon, morphology, and syntax. See Table 1 for descriptive statistics produced for selected speech properties.

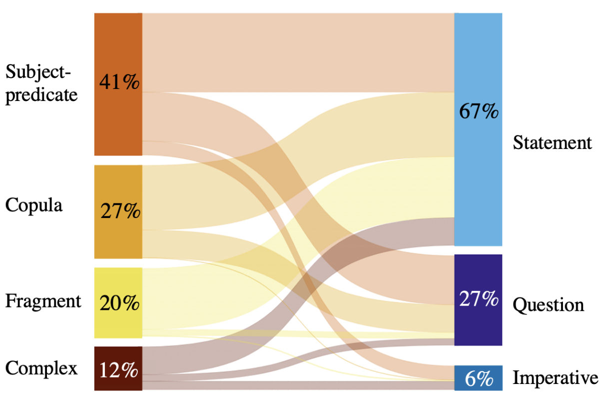

Figure 2: Sankey diagram showing naming instances belonging to different sentence complexity and sentence types.

Two commonly used indices are Mean Length of Utterance (MLU; Brown, 1973) and the Type-Token Ratio (TTR; Templin, 1957). Together, these measures provide valuable insights into a speaker's typical sentence length, complexity, and vocabulary richness. As expected, in parent speech, simpler sentences tend to have smaller MLU. When applying TTR as a measure of lexical diversity, we found that complex sentences were more lexically diverse than copula and subject-predicate sentences. However, we also noticed that TTR for fragments was high. This is likely due to the effect of sample size, as the TTR measure is strongly influenced by the number of tokens. Smaller samples often show greater diversity because fewer tokens are used, resulting in a higher frequency of distinct words.

As shown in Figure 2, parent naming utterances contain 20.4% of fragments, 26.9% of copula, 40.8% of subject predicate, and 11.9% of complex sentences containing two or more lexical verbs. Additionally, 27.6% of naming instances are questions, and 6.4% of naming instances are imperative. As expected, parent speech is quite linguistically diverse, with a mixture of long and short sentences and various grammatical structures and types. This diversity likely supports children's language development by introducing them to a broad array of vocabulary and sentence structures, strengthening their language skills over time.

Effect of speech properties on attention

Given that parent speech is linguistically complex, how do naming sentences of different properties guide the child's attention to the named target? To explore this, we analyzed the child's attention to the target during the naming window, spanning from the onset to the offset of the naming utterance. Within this window, we calculated the proportion of time the child focused on the target and compared the target-looking time across different types of naming events.

Overall, children spent roughly 27% (SD = 38%) of time looking at targets. To formally test the effect of speech complexity, we fit a mixed-effects logistic regression predicting *target looking time* from *speech complexity* while also taking into account the random intercepts for each subject (`lmer(target_look~complexity + (1|subj))`. For complexity, fragment was coded as 1, copula was coded as 2, subject predicate was coded as 3 and complex was coded as 4. As shown in Figure 3, the model revealed a significant main effect of complexity ($\beta = -0.02$, $SE = 0.01$, $p < 0.05$), suggesting that higher speech complexity was associated with a decrease in children's target looking time. Simpler sentences might offer the right level of difficulty for getting children to attend to target, whereas complex sentences could be too hard for them to process at this young age.

To further explore whether the amount of information children process depends on their individual learning capacity (Foushee et al., 2021), we used each child's MCDI score as an indicator of their current language capacity and examined whether speech complexity interacts with this capacity in predicting attention to the target (`lmer(target_look~complexity + vocab + complexity:vocab + (1|subj))`). Neither complexity ($\beta = 0.03$, $SE = 0.02$, $p = 0.16$) nor vocabulary size ($\beta = 0.002$, $SE = 0.002$, $p = 0.26$) significantly predicted attention. The interaction was also non-significant ($\beta = -0.0006$, $SE = 0.0004$, $p = 0.54$), suggesting complexity effect on attention did not depend on vocabulary size. Similarly, there was no interaction effect between age and complexity ($\beta = 0.003$, $SE = 0.005$, $p = 0.51$). Therefore, the complexity of speech does not depend on individual language capacity or age in predicting attention to the target.

We ran another linear mixed-effects model with sentence type as a predictor of target look, accounting for individual differences with a random subject effect. We found that questions led to significantly shorter target looking time compared to statements ($\beta = -0.05$, $SE = 0.02$, $p = 0.02$), while imperatives had no significant effect ($\beta = 0.05$, $SE = 0.04$, $p = 0.20$). This pattern may arise when parents ask questions about unfamiliar targets, where the child might be uncertain about the correct name of the object, leading to difficulty in identifying and attending to the correct referent.

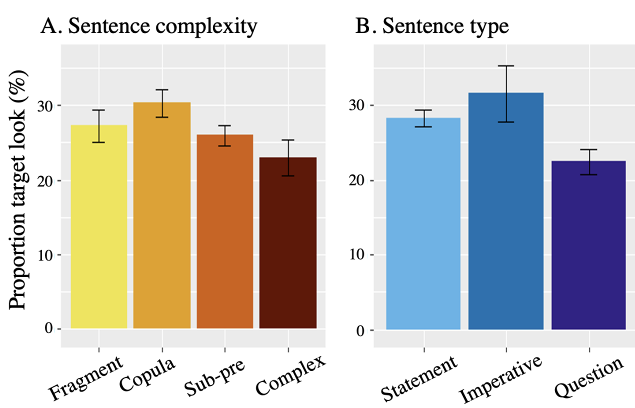

Figure 3: Children's proportion of target looking time during naming events of varying sentence complexity and type.

Effect of speech properties and pointing on attention

During book reading, both parents and children pointed to objects on book pages regularly (Figure 4). As expected, naming and gestures were highly coupled (72%), especially in action-oriented and attention-guiding imperative sentences. To further investigate the role of pointing gestures, we divided each type of naming events into three categories: with parent gesture, with child gesture, and with no gesture. We then explored how the combination of speech properties and pointing impacted children's attention to the target.

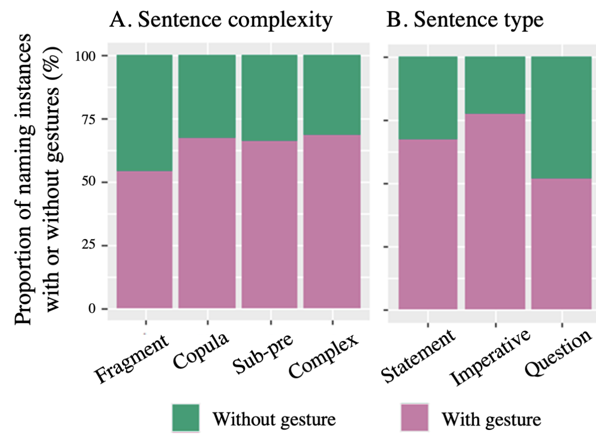

Figure 4: Proportion of naming instances paired with gestures across different sentence complexity (A) and type (B).

Sentence complexity and gestures Using naming events with and without gestures, we first examined how sentence complexity influenced children's attention to the target. The results revealed a significant main effect of complexity ($\beta = -0.02$, $SE = 0.01$, $p = 0.04$) when naming was paired with a gesture, indicating that increased complexity was associated with reduced attention. No significant complexity effect was found when there was no gesture ($\beta = -0.01$, $SE = 0.01$, $p = 0.27$). Further analysis, splitting the data by the provider of the gesture (the child or the parent), showed that complexity significantly predicted attention only when the naming event involved a gesture from the child ($\beta = -0.05$, $SE = 0.02$, $p = 0.03$), but not from the parent ($\beta = -0.02$, $SE = 0.01$, $p = 0.11$). These findings suggest that child gestures play a crucial role in modulating the effect of sentence complexity on children's attention during naming events.

Sentence types and gestures A series of similar linear mixed-effects models were conducted to examine the effect of sentence types on children's attention to the target during naming events with and without gestures. Questions did not significantly predict attention, regardless of whether the naming event was paired with a gesture or not (child gesture: $\beta = -0.08$, $SE = 0.05$, $p = 0.11$; parent gesture: $\beta = -0.05$, $SE = 0.03$, $p = 0.09$; no gesture: $\beta = -0.03$, $SE = 0.03$, $p = 0.29$). However, imperatives were associated with increased attention both when the naming event was paired with a child gesture ($\beta = 0.22$, $SE = 0.08$, $p = 0.004$) and when no gesture was present ($\beta = 0.13$, $SE = 0.06$, $p = 0.04$), but not when paired with a parent gesture ($\beta = -0.03$, $SE = 0.04$, $p = 0.36$). These results suggest that imperative sentences are more effective in directing children's attention, particularly when accompanied by child gestures, but less effective when paired with parent gestures. This is likely because imperatives with parent gestures often require the child to switch attention (e.g., parents point and say "Look at the cat over

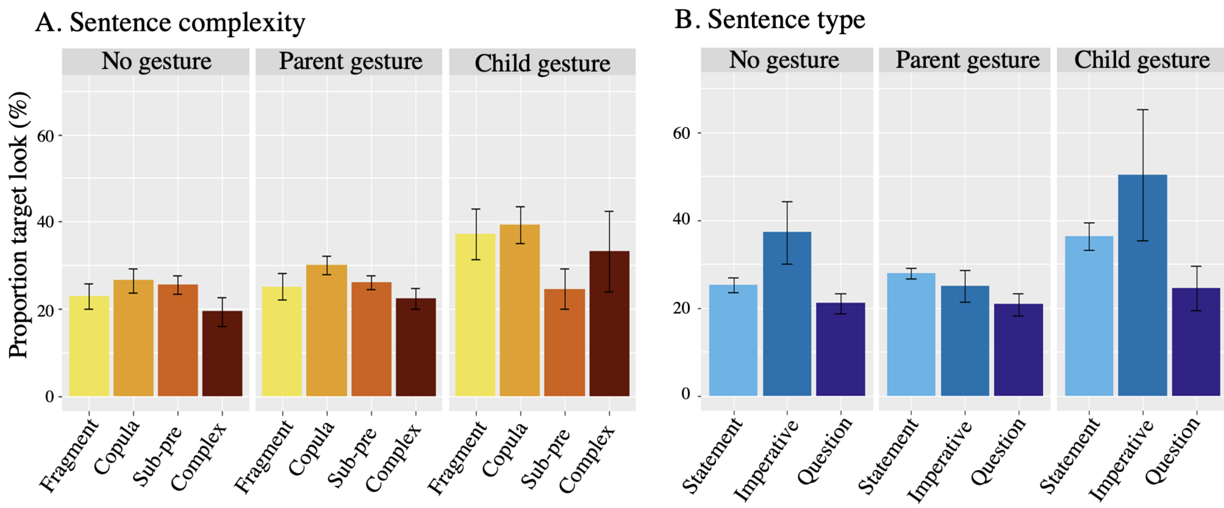

Figure 5: Effect of speech complexity on attention during naming events paired with parent, child, and no gestures.

here!"), while child gestures may prompt the parent to name the object that is already the focus of the child's attention.

Discussion

Our study investigated how speech properties and gesture use influenced children's real-time attention during word learning moments in shared book reading. Using children's moment-by-moment gaze data and parents' spontaneous speech, we discovered three major findings.

First, in line with many previous studies (e.g., Cameron-Faulkner and Noble, 2013), we found converging evidence that shared book reading exposes children to a rich variety of vocabulary and sentence structures, enhancing language acquisition. This linguistic diversity helps children build a multi-layered framework of language, incorporating varied vocabulary and grammatical structures and sentence types.

Second, despite the complex linguistic environment of shared book reading, children in our study showed a preference for simpler sentences, likely due to their relatively low language capacity at this stage. This pattern aligns with findings from Foushee (2021), suggesting that simpler sentences are more accessible to younger children. Although we did not observe a significant vocabulary or age effect on attention, this may be due to the narrow age range (18–24 months) of participants tested. Because most children were at a similar developmental stage, there was limited variability in vocabulary size, making it difficult to detect such an effect.

Third, the child's pointing gestures are key in linking words to their referents, particularly in simpler sentences and attention-getting imperatives. Pointing helps ground abstract speech in observable contexts, making it easier for children to map words to referents. This could be because when children point, caregivers typically label what children point (Goldin-Meadow, 2007; Colonnesi et al., 2010). For instance, if a child points to a picture of a dog, the parent might say, "Yes, that's a dog!" This type of follow-in labeling has been found to be more effective for supporting language learning, as it capitalizes on the child's active and existing engagement with the object to be named (Baldwin, 1991; Yu and Smith, 2012). Additionally, follow-in labeling is more likely to facilitate joint attention, a key component of language learning, as both the parent and child focus on the same object or event (Tomasello and Farrar, 1986). Surprisingly, parent gestures were not effective in guiding child attention in our book-reading context. This is possibly because these gestures redirected the child's attention to an object the child was not yet focused on. When children are not actively engaged with an object or concept, redirective labeling may require more cognitive effort to process and may be less successful in fostering word learning (Tomasello and Farrar, 1986).

Although we did not find that complex sentences capture the attention of children in our tested age group, it is possible that as children grow older, they become better equipped to process more complex linguistic input. In such cases, introducing more complex linguistic information after attention to the target has been established could potentially further support learning. This readiness to engage with more complex speech is likely facilitated through ostensive cues, such as pointing. Pointing helps direct children's attention and highlights important aspects of the learning environment. By following the child's points and using pointing to direct the child's attention, caregivers may enhance the child's ability to process more complex language input more efficiently.

Understanding the quality of learning input during shared book reading is essential for gaining insight into the mechanisms behind the well-documented positive effects of books on language acquisition. Multimodal input, combining verbal and nonverbal cues, plays an important role in resolving referential ambiguity, thereby enhancing language learning. Our study demonstrates how speech and gestures work together to support these learning mechanisms, fostering a more effective language learning environment.

References

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*(5), 875–890.

Brown, R. (1973). A first language: The early stages. *Cambridge, MA: Harvard University*.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive science*, *27*(6), 843–873.

Cameron-Faulkner, T., & Noble, C. (2013). A comparison of book text and child directed speech. *First Language*, *33*(3), 268–279.

Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education and Bilingualism*, *7*(4), 303–319.

Colonnesi, C., Stams, G. J. J., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352–366.

Deckner, D. F., Adamson, L. B., & Bakeman, R. (2006). Child and maternal contributions to shared reading: Effects on language and literacy development. *Journal of Applied Developmental Psychology*, *27*(1), 31–41.

Dickinson, D. K., Griffith, J. A., Golinkoff, R. M., & Hirsh-Pasek, K. (2012). How reading books fosters language development around the world. *Child development research*, *2012*.

Fenson, L. (2002). *Macarthur communicative development inventories: User's guide and technical manual*. Paul H. Brookes.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231.

Foushee, R., Srinivasan, M., & Xu, F. (2021). Selective attention to spoken language in preschoolers based on speech complexity and learning rate.

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., & Adolph, K. E. (2010). Head-mounted eye-tracking of infants' natural interactions: A new method. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 21–27.

Goldin-Meadow, S. (2007). Pointing sets the stage for learning language—and creating language. *Child development*, *78*(3), 741–745.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(1), 23–45.

Lingwood, J., Lampropoulou, S., De Bezenac, C., Billington, J., & Rowland, C. (2023). Children's engagement and caregivers' use of language-boosting strategies during shared book reading: A mixed methods approach. *Journal of Child Language*, *50*(6), 1436–1458.

MacWhinney, B. (2014). *The childes project: Tools for analyzing talk, volume i: Transcription format and programs*. Psychology Press.

Montag, J. L. (2019). Differences in sentence complexity in the text of children's picture books and child-directed speech. *First language*, *39*(5), 527–546.

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological science*, *26*(9), 1489–1496.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child development*, *78*(3), 938–954.

Murphy, C. M. (1978). Pointing in the context of a shared activity. *Child development*, 371–380.

Ninio, A., & Bruner, J. (1978). The achievement and antecedents of labelling. *Journal of child language*, *5*(1), 1–15.

Noble, C. H., Cameron-Faulkner, T., & Lieven, E. (2018). Keeping it simple: The grammatical properties of shared book reading. *Journal of child language*, *45*(3), 753–766.

Rohlfing, K. J., Grimminger, A., & Nachtigäller, K. (2015). Gesturing in joint book reading. In *Learning from picturebooks* (pp. 99–116). Routledge.

Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First language*, *28*(2), 182–199.

Templin, M. C. (1957). Certain language skills in children; their development and interrelationships.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*(2), 187–204.

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental psychology*, *24*(4), 552.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS one*, *8*(11), e79659659.