Parents Adaptively Use Anaphora During Parent-Child Social Interaction

Jasmine Falk[1], **Yayun Zhang**[2], **Matthias Scheutz**[1], **Chen Yu**[2,3]

jasmine.falk@tufts.edu, yayunzhang@utexas.edu, matthias.scheutz@tufts.edu, chen.yu@austin.utexas.edu

[1]Human-Robot Interaction Laboratory, Tufts University, Medford, MA, USA
[2]Department of Psychology, The University of Texas at Austin, Austin, TX, USA
[3]Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, USA

Abstract

Anaphora, a ubiquitous feature of natural language, poses a particular challenge to young children as they first learn language due to its referential ambiguity. In spite of this, parents and caregivers use anaphora frequently in child-directed speech, potentially presenting a risk to effective communication if children do not yet have the linguistic capabilities of resolving anaphora successfully. Through an eye-tracking study in a naturalistic free-play context, we examine the strategies that parents employ to calibrate their use of anaphora to their child's linguistic development level. We show that, in this way, parents are able to intuitively scaffold the complexity of their speech such that greater referential ambiguity does not hurt overall communication success.

Keywords: anaphora; child-directed speech; linguistic tuning

Introduction

Adult language use is characterized by striking a fine balance between linguistic efficiency and referential ambiguity. Speakers often use ambiguous expressions instead of a full referential phrase in the interest of brevity (Grice, 1975), thereby relying on their listener's ability to infer the intended meaning. Indeed, the use of anaphora instead of repeated explicit naming in written sentences actually appears to improve comprehension speed for adult and older child readers, a well-documented effect known as the *repeated name penalty* (Eilers, Tiffin-Richards, & Schroeder, 2019; Gordon, Grosz, & Gilliom, 1993). But how does this ability to handle ambiguous expressions first develop? More specifically, how do young children learn to resolve anaphoric references?

A growing body of work suggests that answering how children handle referentially ambiguous expressions will require us to consider not only a child's learning ability, but also their language learning environment (Leung, Tunkel, & Yurovsky, 2019; Tamis-LeMonda, Kuchirko, & Song, 2014; Yurovsky, Doyle, & Frank, 2016). Parents have been shown to intuitively scaffold the complexity of their speech in consistent patterns to match the changing needs and language abilities of children as they mature cognitively, a phenomenon known as **linguistic tuning** (Yurovsky et al., 2016). A multitude of studies demonstrates that linguistic tuning occurs across morphological, phonological, syntactic, and lexical aspects of child-directed speech (CDS), and thus children receive speech input that has been already adjusted to suit their current language capabilities, guiding them towards learning more complex grammars and words (Bellinger, 1980; Kitamura & Burnham, 2003; Kunert, Fernández, & Zuidema, 2011; Messer, 1978; Roy, Frank, & Roy, 2009; Schroer, Smith, & Yu, 2019; Snow, 1972).

For instance, Roy et al. (2009) found that parents and caregivers increased utterance length and lexical diversity, and decreased the proportion of single-word utterances as their child aged. Similarly, Kunert et al. (2011) measured speech complexity via utterance length, word length, number of word types, and number of consonant triples to find that parents' modifications in each of these metrics was contingent on the age of their child. Kitamura and Burnham (2003) have also shown that parents modify the mean fundamental frequency and pitch range of CDS in accordance with their child's age.

In this paper, we examine linguistic tuning as it applies to anaphora use in CDS, providing further evidence for the critical role parents play in a child's linguistic development. We investigate parental use of anaphora in free-play parent-child interactions, and discuss the implications for future work aimed at developing detailed models of anaphora resolution acquisition.

Anaphora

The ability to refer to different objects is an essential part of daily conversation, and with every reference, a speaker must decide how explicit to be in order to allow their listener to pick out the intended referent (Arnold, 2010). One way in which speakers often balance efficacy and efficiency when referring is through the use of **anaphora**. Anaphora are words or phrases that reference an entity mentioned or indicated elsewhere in discourse, such as "it" or "she" (Sukthanker, Poria, Cambria, & Thirunavukarasu, 2020). Anaphora usage allows interlocutors to avoid redundancy when referring, enabling more efficient communication, with the trade-off being the risk of introducing more ambiguity to an utterance.

Because the meaning of an anaphor depends entirely on the context in which it is used, the process of determining what an anaphoric expression refers to, known as **anaphora resolution**, requires inference; that is, an anaphor cannot be understood without first determining its antecedent (Palmović, Matić, & Kovačević, 2018). Anaphora resolution therefore involves inference and increases the computational complexity required to understand an utterance. Given that children have limited processing resources, anaphora usage in CDS is likely to pose a greater risk to effective communication than in adult-adult communication (Sekerina, Stromswold, & Hestvik, 2004).

Despite the risk it may present to communicating effectively with young children, usage of anaphora and other deictic terms is ubiquitous throughout CDS (Laasko & Smith, 2007). A possible explanation as to why this may not hurt communication could be that parents calibrate their use of anaphora in accordance with their child's linguistic development, in the same way that parents scaffold many other features of CDS to facilitate communication with their child. However, it is not well-understand what particular cues, other than overall comprehension difficulties, parents might perceive when their use of anaphora fails (Song & Fisher, 2005). Given this, it is critical to establish that parents are sensitive to their children's anaphora comprehension abilities and are thus able to adjust their use of anaphora such that their overall communication is not hurt.

The overarching hypothesis of our study is that parents fine-tune CDS based on their child's age, where age is used as a proxy for linguistic ability, as is consistent with the literature on linguistic tuning (Bellinger, 1980; Kitamura & Burnham, 2003; Messer, 1978; Roy et al., 2009; Snow, 1972; Yurovsky et al., 2016). Specifically, we examine how parents of infants 12 to 25 months of age use anaphora in free-flowing parent-child interaction. Existing work suggests that children first begin to comprehend anaphora in CDS around 1 to 3 years of age, a developmental period during which children's overall linguistic sophistication increases dramatically (Bellinger, 1980; Moyer, Harrigan, Hacquard, & Lidz, 2015; Nelson, 1975). Parents of children in this age group are thus likely to modify their use of anaphora in CDS rapidly to keep pace with their child's emergent anaphora comprehension abilities.

We compare anaphora use of the parents of older children in the participant group with that of the parents of younger children, hypothesizing that parents would introduce more instances of anaphora and generate fewer accompanying assistive behaviors (e.g. visual cues) the older the child was. Moreover, using eye movement data, we can measure how well children are able to infer the referential intent of the anaphora used in parent speech, given the demonstrated success of applying the visual world paradigm to eye-tracking studies (Sekerina, 2014). If parents are indeed able to successfully adapt their anaphora use in accordance with their child's linguistic development, children should maintain a consistent level of comprehension despite the greater linguistic complexity parents may introduce to their speech as children grow older.

Methods

Data Collection

To study how parents adpatively scaffold their speech to facilitate communication during naturalistic interactions, we collected visual and auditory data from a toy-play context in a home-like environment with 36 parent-child dyads (mean child age = 18.63mos [range: 12.3-25.3]). Parents were instructed to play naturally with their child using 24 toys on a carpeted floor for about 15 minutes (Figure 1). The toys consisted of blocks, animal and insect figurines, dolls, toy cars, and a variety of other play objects.

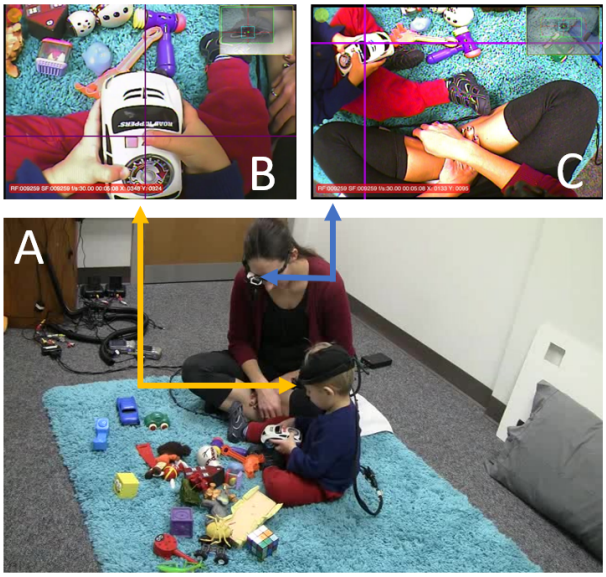

Figure 1: Parent and child were asked to play with a set of toys together in a naturalistic environment while wearing head-mounted cameras (A). Top panel shows child's (B) and parent's egocentric views (C).

Prior to the play session, both parent and child were fitted with head-mounted eye trackers from Positive Science LLC. The eye-tracking system consisted of a scene camera on the wearer's forehead that recorded images from the wearer's perspective with a visual field of 108°, as well as an infrared camera pointed to the participant's right eye to record saccades and fixations. Both cameras sampled at a rate of 30Hz. To capture third-person views of the participants, additional cameras were placed throughout the room.

After placing the eye trackers on the parent-child dyad, the researchers ran a calibration procedure. Once the play session started, the researcher would leave the room and monitor the experiment from an adjacent room, reentering only to adjust and re-calibrate the infant's eye tracker if it was bumped or moved during the play session. All procedures in this study were approved by the Human Subjects and Institutional Review Boards at Indiana University. Written consent forms were obtained from a parent or guardian for each child before data collection.

Coding and Analysis

After the experiment was completed, the eye-tracking videos from the scene cameras and eye cameras were synchronized using a software program that generated cross-hairs to indicate where the participant was looking during each video frame. Using the first-person scene camera with cross-hairs overlaid, the visual gaze of the parent and child were each manually coded by annotating which region of interest (ROI)

Table 1: Descriptive statistics of parent speech metrics

	M	*SD*	Median	Range
Number of utterances	128.36	57.16	121.5	9-272
Number of utterances containing anaphora	40.25	24.44	34	0-116
Number of anaphora	46.31	28.59	39.5	0-133
Number of visually-cued anaphora	24.22	18.49	18	0-86
Number of anaphora with required visual cues	15.58	11.04	14	0-53
Anaphora utterance rate	5.27	2.51	4.82	0.00-11.33
Anaphora use ratio	0.34	0.12	0.33	0.00-0.59
Proportion visually-cued anaphora	0.52	0.16	0.50	0.14-0.89
Proportion required of visually-cued anaphora	0.69	0.20	0.70	0.26-1.00

the cross-hairs overlapped with during a gaze fixation. Each of the 24 toys and the social partner's face were coded as ROIs (25 ROIs in total).

The recordings of the parents' speech were also transcribed. Following standard practice in the field (e.g. Pereira et al. (2014); Suanda, Smith, and Yu (2016); Yu and Smith (2012)), parent speech was objectively coded at the utterance level, defined as strings of speech between two periods of silence lasting at least 400ms. All parent talk and vocal play (such as saying "vroom-vroom") were considered speech.

In the current study, we focused on anaphoric instances spoken by parents to their children. The transcriptions were manually coded for instances of anaphora, as well as contextual information about their usages. We annotated each utterance that contained an anaphoric expression, regardless of whether or not the anaphor referred to an entity that was physically present in the room. As a reliability check, two trained coders independently transcribed utterances of four randomly-selected subjects (about 10% of the total utterances coded) and reached 96% agreement.

For each anaphor, we manually coded for the corresponding noun phrase antecedent(s) that the parent intended the anaphor to refer to. In addition to coding for intended referents, we were also interested in parent use of multimodal cues in conjunction with anaphora.

Intended referent. Each of the 25 ROIs (the 24 toys and the social partner's face) was assigned a unique referent identification number. The coder determined the intended referent by observing video recordings from the parent's first-person scene camera and third-person cameras for visual context of the speech.

Use of multimodal cues. The use of multimodal cues was defined as behavior in which a parent paired the usage of an anaphor with a visual cue, such as pointing to, gesturing at, or handing an object to the child. We coded for the presence or absence of visual cues in conjunction with an anaphor. Parental use of multimodal cues was observed via both third-person and parent first-person scene cameras.

Visual cues paired with anaphora were further categorized into required and supplemental cues. **Required visual cues** were defined as extra-linguistic cues that were necessary to resolve an anaphor whose antecedent could not be determined from speech alone. For instance, throughout the corpus, parents could often be observed asking their child, "What is this?", in conjunction with pointing to or holding up a toy for the child to see. In these cases, the anaphor in question could not be resolved based on the speech, but rather required a visual cue for the listener to determine the intended referent. On the other hand, **supplemental visual cues** were defined as visual cues that were not strictly necessary for a fully linguistically-developed interlocutor to correctly resolve the anaphor, but were used instead to assist the conversation partner in identifying the intended referent.

Results

For this study, the data of interest were the coded transcripts of the parent utterances spoken during the free-play session, as well as the child's anaphora resolution accuracy. Descriptive statistics of the metrics of parent speech can be found in Table 1.

To determine whether parents modified their use of anaphora as their children aged, we conducted two analyses: a correlational analysis to test for linear relationships between child age and metrics of parent anaphora use, and a median split analysis, to test if parents of older children differed in their anaphora use as compared to parents of younger children. For the correlational analysis, we evaluated the strength of the relationship between the child's age and the adjustments parents made to their anaphora usage. For the median split analysis, we divided the subjects into two groups based on the median child age (median = 18.2mos) and compared metrics of anaphora use between parents of children above the median age and parents of children below the median age. Using these methods of analysis allowed us to extract both linear relationship and group difference patterns in the data.

Additionally, we evaluated the communication efficacy of a parent-child dyad based on the child's success at correctly identifying the intended referent of each anaphor using the study's eye-tracking data. We compared the average anaphora resolution accuracy scores of children above the median age to children below the median age.

Frequency of Anaphora Use

To test our first hypothesis, that parents scaffold the frequency of their anaphora usage based on their child's age, we measured frequency of anaphora use via anaphora utterance rate and anaphora use ratio. Both metrics yielded significant results for the correlational and median split analyses, confirming our hypothesis that parents increase their use of anaphora as their children develop greater linguistic knowledge.

Anaphora utterance rate. Anaphora utterance rate was defined as the number of utterances containing at least one anaphor divided by the length of time of the parent-child dyad's free-play session (in minutes). As predicted, results of the Pearson correlation indicated that there was a significant positive association between a parent's anaphora utterance rate and their child's age ($r(34) = .46, p = .005$). Figure 2 summarizes the results. Additionally, parents of children above the median age had higher anaphora utterance rates ($M = 6.18, SD = 2.88$) than parents of children below the median age ($M = 4.35, SD = 1.70$). Results of an independent samples t-test indicated that this difference was significant ($t(34) =$ -2.32, $p = .026$).

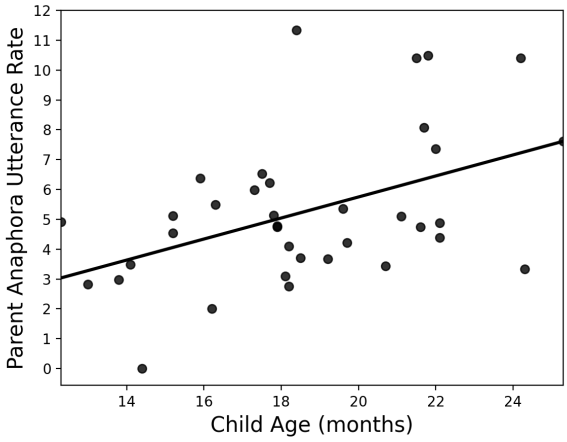

Figure 2: Relationship between child age and parent anaphora utterance rate. $r(34) = .46, p = .005$.

Anaphora use ratio. The anaphora use ratio was defined as the ratio of a parent's anaphora utterance rate to their rate of speech. Results of the Pearson correlation, summarized in Figure 3, also indicated a significant positive association between a parent's anaphora use ratio and their child's age ($r(34) = .49, p = .003$). In comparing the groups split by median age, we again found that parents of children above the median age had higher anaphora use ratios ($M = 0.39, SD = 0.11$) than parents of children below the median age ($M = 0.29, SD = 0.11$). The difference between these groups was confirmed statistically with an independent samples t-test ($t(34) =$ -2.84, $p = .007$).

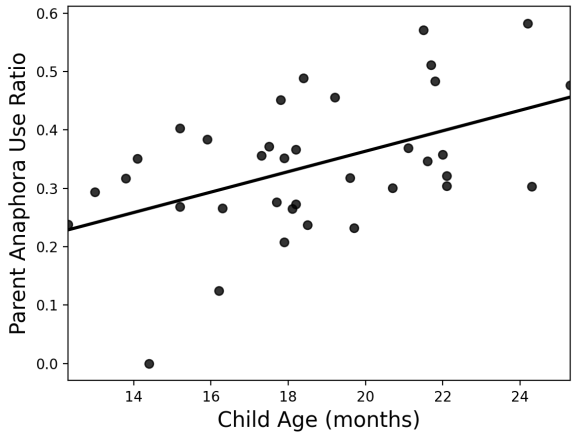

Figure 3: Relationship between child age and parent anaphora use ratio. $r(34) = .49, p = .003$.

Frequency of Visually-Cued Anaphora

Parents sometimes paired an anaphor with a visual cue, such as pointing or gesturing, that aided the child in determining the anaphor's intended referent. Visual cues were either considered required or supplemental. If no visual cue was present in conjunction, the anaphor was considered to be verbally-cued.

To determine whether parents adjusted their use of visual cues when using anaphora according to their child's age, we measured the proportion of anaphora that were paired with a visual cue out of all anaphora the parent used. Although we found that parents of children above the median age used a slightly smaller proportion of visually-cued anaphora ($M = 0.51, SD = 0.11$) than parents of children below the median age ($M = 0.53, SD = 0.21$), the difference between the groups was not statistically significant ($t(34) = 0.39, p = .701$). The results of the Pearson correlation also did not indicate a significant negative association between a parent's use of visually-cued anaphora and their child's age ($r(34) =$ -.113, $p = .517$), as summarized in Figure 4. We had expected that parents would use more visual cues to aid younger children in resolving anaphora accurately and would decrease their use of visual cues when communicating with older, more linguistically-developed children who may not require as much assistance to successfully understand anaphora.

We also calculated the proportion of anaphora with required visual cues out of the total number of visually-cued anaphora. As expected, parents of children above the median age used a smaller proportion of required visual cues ($M = 0.62, SD = 0.16$) than parents of children below the median age ($M = 0.75, SD = 0.22$). The difference between these groups was statistically significant ($t(34) = 2.04, p =$ .049). However, there was no significant correlation between child age and the proportion of required visual cues

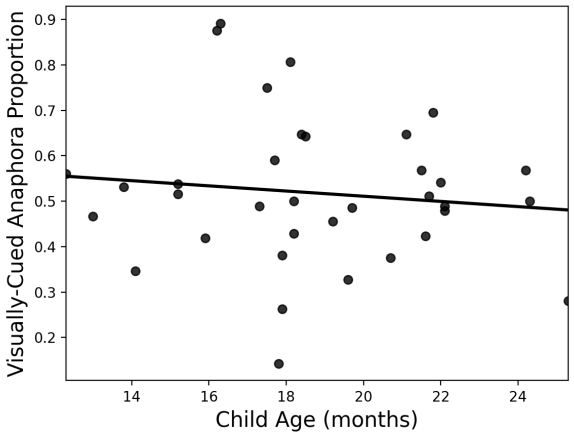

Figure 4: Relationship between child age and the proportion of all anaphora that were paired with a visual cue. $r(34) =$ -.113, $p = .517$.

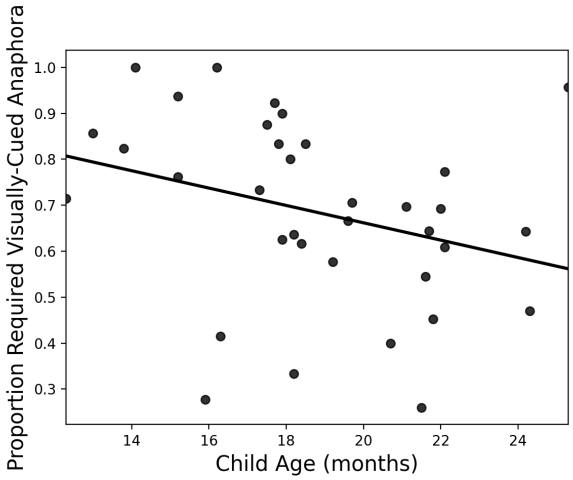

Figure 5: Relationship between child age and the proportion of all visually-cued anaphora that were paired with a required visual cue. $r(34) =$ -.293, $p = .088$.

($r(34) =$ -.293, $p = .088$), summarized in Figure 5. Although stronger results would likely be less sensitive to specific statistical methods, the lack of a significant linear relationship between child age and proportion of required visual cues does not preclude findings of group differences using a median split analysis, given that different approaches extract different patterns in the data and thus demand different interpretations.

These results suggest that rather than scaffolding their use of anaphora simply by decreasing the use of visual cues overall as we had expected, parents appeared to scaffold visual cue use in a more nuanced manner. For older children, parents were more likely to use anaphora that *could* be resolved without a visual cue, but continued to offer assistance through the use of supplemental visual cues, perhaps in order to ensure effective communication levels could be maintained. On the other hand, parents of younger children more frequently used anaphora when the context did not require linguistic knowledge to resolve the anaphor and instead used only visual cues to indicate the intended referent.

Communication Efficacy

To measure the efficacy of a parent's communication with their child, we determined the child's overall anaphora resolution accuracy score. The accuracy score was defined as the proportion of the total number of anaphora used by the parent for which the child correctly identified the intended referent. Scores could range from 0 (indicating the child did not correctly resolve any of the anaphora used) to 1 (indicating that the child correctly identified the intended referents for all anaphora used).

Using an in-house program, parent speech utterances were temporally aligned with child eye-tracking data, allowing us to analyze the child's eye gaze at the utterance level. Because we could not synchronize eye-gaze data with parent speech at the granularity of each word, a child was considered to have correctly identified the intended referent of an anaphor if their eye movements indicated a fixation on the intended ROI at any point during the parent's utterance containing the anaphoric expression. This definition of successful anaphora resolution applied to utterances containing multiple anaphora as well. If a child fixated on any of the intended ROIs during the utterance, the anaphora referring to these ROIs were considered to be correctly resolved. Anaphora that referred to entities that were not physically present in the room or were not one of the 25 defined ROIs were excluded from the calculation of anaphora resolution accuracy scores.

As predicted, children in both age groups had similar average accuracy scores ($M_{\text{above-median}} = 0.60, SD_{\text{above-median}} = 0.14, M_{\text{below-median}} = 0.58, SD_{\text{below-median}} = 0.14$). The anaphora resolution accuracy of children above the median age was not statistically different from the accuracy of children below the median age ($t(33) =$ -0.34, $p = .739$). One parent-child dyad was excluded from this analysis as the parent did not use any instances of anaphora during the free-play session.

These results confirm our hypothesis that the success of communication between parents and their children would remain consistent regardless of any modifications parents made to their anaphora usage. Although parents used anaphora more frequently with older children, both overall as well as in contexts that required greater linguistic capabilities to resolve, parents intuitively scaffolded these adjustments such that the efficacy of their overall communication was not hurt.

Discussion

In the current work, we provide evidence demonstrating that parents are motivated to adjust CDS by a desire for *efficient* communication. Achieving smooth communication requires using speech that balances both efficacy and efficiency, as Grice (1975) formalized in his theory of the Cooperative Prin-

ciple. We demonstrate that the use of linguistic tuning in CDS is ultimately motivated by this Gricean framework of smooth communication. Parents modify their speech when conversing with their children to maximize both success and efficacy of communication given the communicative constraints of a young language learner.

In particular, we focus on how parents adjust their usage of anaphora to achieve efficiency while maintaining a successful level of communication with their children. We show that parents use fewer anaphora when speaking to younger children, and indeed, report a significant positive correlation between child age and frequency of anaphora in parent speech. We also demonstrate that parents of younger children use more anaphora that requires assistive visual cues to resolve than parents of older children. However, we found no significant correlation between child age and the proportion of anaphora requiring visual cues, indicating that the results of the median split analysis may not have been strong.

In that regard, our study may have been limited by the use of age as a proxy for linguistic development. Although age is related to children's communication skills and vocabulary size, it is an indirect measure of knowledge as infants acquire language at varying rates. Examining parent speech patterns in relation to direct measures of child language skills, such as the MacArthur–Bates Communicative Development Inventory, may yield stronger results.

Role of anaphora in early word learning. Beyond demonstrating the critical role linguistic tuning plays in anaphora resolution acquisition in young children, our findings also lend support for the importance of further work in understanding how children acquire the ability to resolve referentially ambiguous expressions.

Given the abundant usage of anaphora in parent speech, understanding the referential intents of anaphora can be a "game-changer" for early word learning. If children are able to infer the referential intents of anaphora, then all utterances containing anaphora can provide direct supervisory signals for word learning (Smith & Frank, 2012). They can be treated as information-rich as labeling utterances containing object names can be. To quantify word learning from parent speech, we need to have a direct measure of children's comprehension of anaphora. In developmental research, there is a long history of linking gaze behaviors with language comprehension. It has been shown that upon hearing speech, infants' looking behavior is driven by their language knowledge (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998).

Towards that end, we could quantify anaphora comprehension by utilizing the moment-by-moment gaze data collected during toy-play. For example, we can measure where children look during an anaphoric event, whether they look at the correct target object or not, and whether they switch attention during or after different types of anaphoric events. We can also link real-time measures of learning with children's age and vocabulary knowledge to further quantify the effectiveness of anaphora usage in parent speech. It is important to take children's contingent responses during word learning moments into account because how parents scaffold their speech could lead to real-time changes in childrens' attention, which could subsequently influence word learning outcomes.

Multimodal communication in parent-child interaction. Naturalistic interactions usually involve learning from multiple information sources. There is evidence showing that CDS is just one example among many for adults to introduce facilitating modifications during parent-child interactions (Brand, Baldwin, & Ashburn, 2002). Take gestural modification as another example beyond speech. Gestures, such as pointing, can highlight the correct referent ostensibly and offer crucial clues for infants to locate the intended referent when facing referential uncertainty (Rowe, Özçalışkan, & Goldin-Meadow, 2008), since deictic gestures provide an easier pathway for infants to identify and integrate audio-visual information (Cook, Mitchell, & Goldin-Meadow, 2008).

However, instead of treating gestures and speech as two separate information sources that learners must process in parallel, researchers have found that parents tailored their use of gestures with a goal of disambiguating the verbal message (Iverson, Capirci, Longobardi, & Caselli, 1999). If the child is able to infer the referents of anaphora, or if the child's attention is already drawn to the target referent, providing additional gestural cues becomes redundant and optional. Alternatively, if the child cannot make a correct inference, then pointing is necessary and required. It is likely that efficient communication between parents and children is influenced by not only one single information source but multiple contributing factors. These factors can each independently and collectively impact early social interaction and human communication in general.

Conclusion

In this paper, we have demonstrated based on data from free-play parent-child interactions that parents introduce more anaphora in their speech interactions with older children. This tendency points to the importance of parent linguistic scaffolding for the development of the child's cognitive inferential capabilities. Additionally, a child's ability to understand more and more forms of anaphora enables their parent to progress from child-directed speech towards more adult-like language usage, in particular, adult-like efficiency of language use supported by potentially ambiguous linguistic devices. Our results thus lend further support to the linguistic tuning hypothesis and can also form the basis for computational models of child language acquisition that incorporate the scaffolding trajectories provided by parents.

Acknowledgments

This work is supported by National Institute of Child Health and Human Development R01HD074601 and R01HD093792.

References

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187-203. doi: https://doi.org/10.1111/j.1749-818X.2010.00193.x

Bellinger, D. (1980). Consistency in the pattern of change in mothers' speech: Some discriminant analyses. *Journal of Child Language*, *7*(3), 469–487. doi: 10.1017/S0305000900002798

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*(1), 72–83.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047–1058.

Eilers, S., Tiffin-Richards, S. P., & Schroeder, S. (2019). The repeated name penalty effect in children's natural reading: Evidence from eye tracking. *Quarterly Journal of Experimental Psychology (2006)*, *72*(3), 403–412.

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.

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.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*(3), 311-347. doi: https://doi.org/10.1207/s15516709cog1703_1

Grice, H. P. (1975). *Logic and conversation*.

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development*, *14*(1), 57–75.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, *4*(1), 85-110. doi: 10.1207/S15327078IN0401_5

Kunert, R., Fernández, R., & Zuidema, W. (2011, 05). Adaptation in child directed speech: Evidence from corpora. *SemDial 2011 (Los Angelogue)*, 112-119.

Laasko, A., & Smith, L. B. (2007). Pronouns and verbs in adult speech to children: A corpus analysis. *Journal of Child Language*, *34*(4), 725–763. doi: 10.1017/S0305000907008136

Leung, A., Tunkel, A., & Yurovsky, D. (2019). Parents calibrate speech to their children's vocabulary knowledge. In *Cognitive Science.*

Messer, D. J. (1978). The integration of mothers' referential speech with joint play. *Child Development*, *49*(3), 781–787.

Moyer, M., Harrigan, K., Hacquard, V., & Lidz, J. (2015). 2-year-olds' comprehension of personal pronouns. In *Online Supplemental Proceedings of the 39th Boston University Conference on Language Development.*

Nelson, K. (1975). The nominal shift in semantic-syntactic development. *Cognitive Psychology*, *7*, 461-479.

Palmović, M., Matić, A., & Kovačević, M. (2018, 09). Resolution of anaphoric expressions in children and adults: Evidence from eye movements. *Suvremena Lingvistika*, *44*, 139-154. doi: 10.22210/suvlin.2018.085.13

Pereira, A. F., Pereira, A. F., Smith, L. B., Smith, L. B., Yu, C., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178–185.

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.

Roy, B. C., Frank, M. C., & Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. *Thirty-first Annual Meeting of the Cognitive Science Society*.

Schroer, S. E., Smith, L. B., & Yu, C. (2019). Examining the multimodal effects of parent speech in parent-infant interactions. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (p. 1015-1021).

Sekerina, I. (2014). Visual world eye-tracking paradigm. In P. J. Brooks & V. Kempe (Eds.), *Encyclopedia of language development* (p. 657-658). Sage Publications, Inc.

Sekerina, I., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language*, *31*(1), 123–152. doi: 10.1017/S0305000903005890

Smith, C., & Frank, M. C. (2012). Zero anaphora and object reference in Japanese child-directed speech. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (p. 989-994).

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*(2), 549–565.

Song, H., & Fisher, C. (2005). Who's "she"? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language*, *52*, 29-57.

Suanda, S. H., Smith, L. B., & Yu, C. (2016). The multisensory nature of verbal discourse in parent–toddler interactions. *Developmental Neuropsychology*, *41*(5-8), 324-341. doi: 10.1080/87565641.2016.1256403

Sukthanker, R., Poria, S., Cambria, E., & Thirunavukarasu, R. (2020). Anaphora and coreference resolution: A review. *Information Fusion*, *59*, 139-162. doi: 10.1016/j.inffus.2020.01.010

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science: A Journal of the American Psychological Society*, *23*(2), 121–126.

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

Yurovsky, D., Doyle, G., & Frank, M. C. (2016). Linguistic input is tuned to children's developmental level. *Cognitive Science*.