More than Words: The Many Ways Extended Discourse Facilitates Word Learning

Sumarga H. Suanda, Linda B. Smith, Chen Yu
{ssuanda, smith4, chenyu}@indiana.edu
Department of Psychological and Brain Sciences, Indiana University
Bloomington, IN, 47405 USA

Abstract

Child-directed speech is often temporally organized such that successive utterances refer to the same topic. This type of *extended discourse* on the same referent has been shown to possess several verbal signatures that could facilitate learning. Here, we reveal multiple non-verbal correlates to extended discourse that could also aid learning. Multimodal analyses of extended discourse episodes reveal that during these episodes, toddlers and parents exhibit greater sustained attention on objects, and greater coordination between their behaviors. The results indicate the interconnections between multiple aspects of the language-learning environment, and suggest that parents' speech may both shape and be shaped by non-verbal processes. Implications for understanding how the learning environment influences development are discussed.

Keywords: language acquisition; word learning; discourse development; child-directed speech; joint attention.

Introduction

Children acquire language in an environment rich with structure and regularities. One of its noticeable structures is its temporal structure, with adjacent utterances frequently referring to the same conversational topic:

Mother: *oh there's a super car?*
Mother: *you like cars don't you?*
Mother: *what are you going to do with it?*
Mother: *are you going to make it go?*
(Messer, 1980)

These extended episodes of verbal discourse on the same referent – what we will call *extended discourse* for short - could facilitate learning in multiple ways. Repeated utterances to one topic give children multiple opportunities to identify the focus of parents' speech, and allow children to deploy comprehension of one utterance in the service of comprehending subsequent utterances (Frank, Tenenbaum, & Fernald, 2013; Messer, 1980; Sullivan & Barner, 2016). Additionally, the repetition of utterance properties (i.e., words, sentence structures) common in extended discourse, has been proposed to aid speech perception (Bard & Anderson, 1983), word segmentation (Onnis, Waterfall, & Edelman, 2008), and syntax learning (Hoff-Ginsberg, 1986).

Previous research thus provides evidence for the idea that the *linguistic* features of extended discourse facilitates learning. Here, we take a different perspective on why extended discourse aids learning. Our perspective is based on the idea - and data - that the language-learning environment is inherently multi-modal and multi-dimensional (Roy, Frank, DeCamp, Miller, & Roy, 2015), and that the verbal discourse children hear is intricately tied to its nonverbal perceptual and social contexts (Adamson & Bakeman, 2006). If this is true, then extended episodes of verbal discourse likely co-occur with, and may even be driven by, extended episodes of sustained attention on the part of the child and extended episodes of joint engagement on the part of the child-parent dyad. Since both of these processes have been linked to healthy language development (e.g., Adamson, Bakeman, & Deckner, 2004; Salley, Panneton, & Colombo, 2013), we suggest then that extended *verbal* discourse may facilitate language learning in part through its underlying *nonverbal* components. In the current study, we test the hypothesis that extended discourse possesses important nonverbal features, including: (1) enhanced child and parent attention to the talked-about object, which means parents' speech is more *referentially transparent* (Cartmill et al., 2013), and (2) a greater degree of joint engagement between children and their parents (Adamson et al., 2004; Tomasello & Farrar, 1986).

The role of referential transparency and joint engagement on word learning is well established. For example, Cartmill and colleagues found that children who heard more referentially transparent speech as toddlers had larger vocabularies as preschoolers (Cartmill et al., 2013). Similarly, in a series of recent studies using mini head-cameras worn by toddlers as they played with novel objects with their parents, Smith, Yu, and colleagues found that the referential transparency of parents' object naming (measured by the visual dominancy of the named object over competitor objects) was predictive of toddlers' object name learning (Pereira, Smith, & Yu, 2014; Yu & Smith, 2012). With respect to joint engagement, Adamson and colleagues (2004) found that the time parents and toddlers spent in joint visual attention was linked to toddler vocabulary development (see also Adamson et al., 2004). Other non-visual forms of joint engagement, such as fluid turn-taking of head and hand action in play, has also been linked to word learning (Pereira, Smith, & Yu, 2008).

In the current study, parents and toddlers were observed as they played with, and as parents talked about, a set of objects (see Figure 1). From these free-flowing object-play interactions, we identified moments of extended verbal discourse, as well as moments of short verbal discourse. We measured, via head-mounted eye tracking, moment-by-moment gaze patterns of parents and their toddlers. We also

measured parents' and toddlers' manual actions. Of interest was whether parents' speech inside extended discourse (compared to short discourse) would be marked by more sustained and greater referential transparency (indexed by gaze and action patterns on the referent object), as well as greater joint engagement (indexed by the coupling of gaze and manual actions between toddlers and their parents).

Methods

Participants

Fifty-two parent-toddler dyads participated (*Mean* toddler age = 17.9mos, *SD* = 4.3mos). Twenty-three toddlers were girls. Data of 17 dyads, were part of a previous report on joint attention (Yu & Smith, 2013).

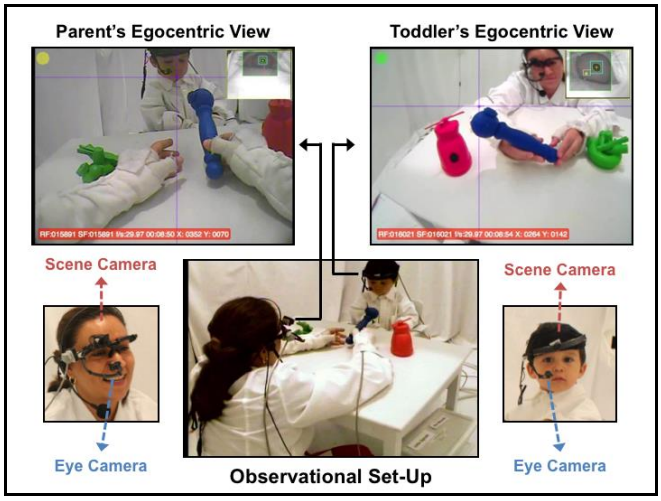

Figure 1. *The observational set up: toddlers and parents played on a table-top in the lab and were equipped with head-mounted eye trackers (bottom row), which produced egocentric views and estimates of gaze direction (top row).*

Apparatus

Figure 1 depicts the set-up. Toddlers sat in a chair across from their parents who sat on floor cushions. Toddlers and parents wore light-weight head mounted eye trackers from Positive Science[1]. As seen in Figure 1, these eye trackers consisted of one outward-facing camera that records the observer's first-person, egocentric views and one inward-facing camera that records the observer's eye movements (for more details, see Franchak, Kretch, Soska, & Adolph, 2011). Both cameras recorded at a temporal resolution of 30 Hz and a spatial resolution of 720x480 pixels. Parents wore a headset equipped with a microphone.

Stimuli included two sets of three novel objects. All objects had a single main color, were similar in size, and were small enough for toddlers to handle. Each object was paired with a novel disyllabic word (e.g., "*habble*", "*tema*").

[1] For 23 dyads, parents were equipped with the Wearcam eye-tracking system. All the current results did not vary as a function of eye-tracking system.

Procedure

After toddlers and parents were fitted with the recording equipment, we placed a set of three objects on the table and instructed parents to play with their children as they normally would, leading to a free-flowing interaction with no constraints on how parents or their children should play, or on what parents should say, with one exception. Prior to play, we told parents that when talking about the objects, to use the names we provided.

The play session consisted of a series of brief trials, each lasting between 1-2 minutes long. On each trial, dyads played with one of two object sets. Object sets were swapped between trials to keep toddlers engaged. Depending on toddlers' compliance, the play session lasted between 2 and 4 trials (*M* = 3.12; *SD* = 1.00); total play duration lasted on average about 5 minutes (*M* = 4 min 57s; *SD* = 89s).

Coding: Parent Speech

Parents' speech during play was fully transcribed. The unit of transcription was the utterance, defined as a string of speech between two periods of silence lasting at least 400ms. Utterances containing reference to one of the objects were marked as *referential utterances*, which included utterances when parents named an object (e.g., "that's a *habble*"), employed a pronoun referring to an object (e.g., "can you push *it*?"), or used an alternate concrete noun referring to an object (e.g., "don't throw *the toy*"). For each referential utterance, we coded the referent by watching the video. On average, there were 95.2 utterances per dyad (*SD* = 36.3), 49.5 of which were referential (*SD* = 23.4).

We then classified each referential utterance as either part of an *extended discourse* or a *short discourse* (Figure 2). To be counted as part of an extended discourse, utterances had to satisfy two criteria. First, utterances had to be part of at least three consecutive utterances referring to the same object. Second, adjacent utterances had to occur within ten seconds of each other. On average, 28.9 utterances per dyad were classified as part of an extended discourse (*SD* = 19.4); 15.7 were classified as part of a short discourse (*SD* = 7.2). An additional 4.7 utterances (*SD* = 4.4) referred to multiple objects and were excluded from analysis.

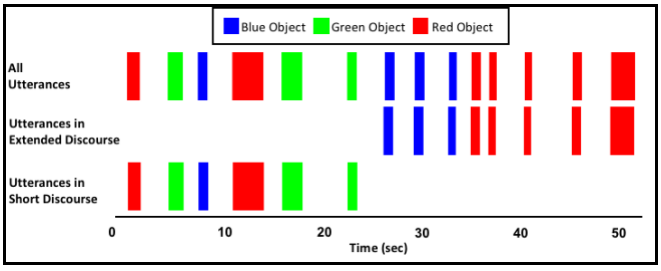

Figure 2. *Representative time series of parents' utterances (top row), including utterances that were part of extended discourse (middle row; see text for criteria) and utterances that were part of short discourse (bottom row).*

Coding: Parent and Child Sensorimotor Behaviors

Gaze Coding. At the end of play, we calibrated both toddlers' and parents' eye trackers by having them fixate to known locations on the table (see Yu & Smith, 2013 for a detailed explanation of calibration). Based on the calibration data, eye-tracking software produced frame-by-frame point-of-gaze estimates, as indicated by the cross-hairs in Figure 1. Eye tracking spatial accuracy is about 3^0 once calibrated (see Franchak et al., 2011).

Using footage from the first-person scene camera (with cross-hair superimposed) and the eye camera (which depicts moment-by-moment eye movements), coders manually annotated the whole session frame-by-frame for parents' and toddlers' target of gaze. For each frame, gaze was coded as one of five possibilities: each of the three toy objects, the partner's face, or elsewhere.

Manual Activity Coding. Coders also watched the session from multiple angles (first-person scene cameras, third-person view cameras) and annotated the session frame-by-frame for moments when toddlers and parents touched each of the three objects. Figure 3 depicts a representative time series of gaze and holding behavior of toddlers and parents over the course of a trial.

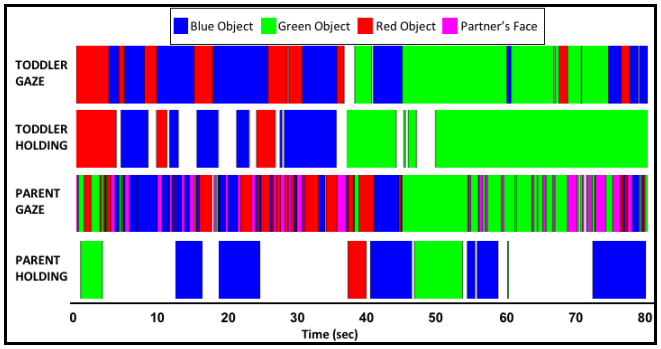

Figure 3. *Representative time series of the sensorimotor behaviors of one toddler (top two rows) and one parent (bottom two rows).*

Results

Referential Transparency in Extended vs. Short Discourse

Is speech in extended discourse more referentially transparent than speech in short discourse? We considered parents' speech to be referentially transparent to the extent that the talked-about (target) objects, and not the distractor objects, were looked at and held by toddlers and parents (see also Frank et al., 2013). Figure 4 depicts the temporal profiles of gaze and holding around utterances that were either part of extended discourse or part of short discourse.

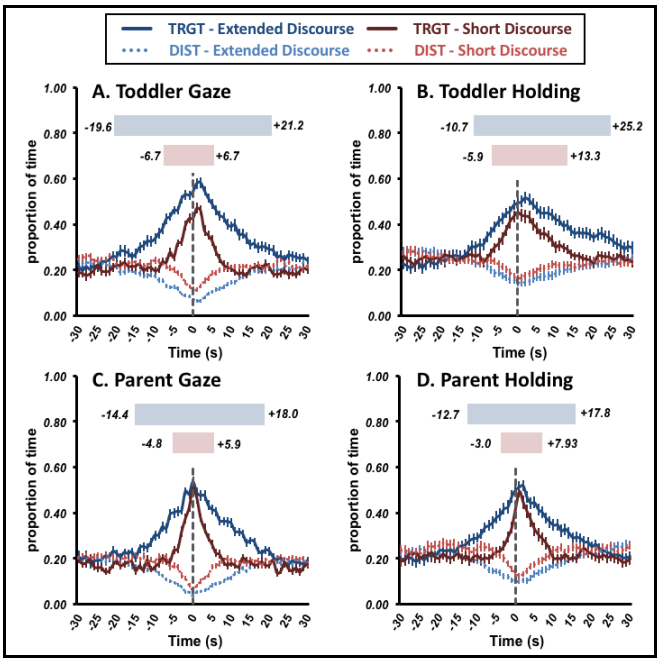

Figure 4. *Temporal profiles of visual and manual attention to target and distractor objects around utterances of different discourse lengths. Vertical dotted lines mark the onset of parents' utterances. Horizontal bands (and adjacent numbers) illustrate the point at which looking (or holding) to target vs. distractors deviates statistically (blue band: extended discourse; red band: short discourse).*

As the figure shows, both utterance types exhibited some referential transparency since both were characterized by more looking and holding to the target than the distractor objects. However, two noticeable patterns distinguished utterances in extended discourse from utterances in short discourse. First, the period of time in which the referent of parents' speech was transparent was longer for utterances in extended discourse than for utterances in short discourse. For utterances in extended discourse, we determined via frame-by-frame *t*-tests that the point in time in which toddlers' gaze at the target object first diverged from the distractor objects was 19.6s prior to utterance onset. Looking time continued to be different until 21.2s after utterance onset. For utterances in short discourse, the period was much shorter: 6.7s both before and after utterance onset. This pattern of more enduring referential transparency in extended discourse was consistent for all toddler and parent measures (see Figure 4).

Second, the figure also shows that even around the moment of utterance, attention to target and distractor objects was more divergent in utterances of extended discourse than in utterances of short discourse. When our analyses honed in on the moments of parents' utterances, which regardless of discourse length are moments when parents are talking about the target object, gaze and holding of the target object was greater in utterances that were part of extended discourse than in utterances that were part of

short discourse (see Table 1). This finding suggests a correlation between the *degree* of toddlers' and parents' nonverbal attention to an object and the duration of verbal discourse on that object. We return to the implications of this correlation in the General Discussion.

Table 1: *Mean gaze and holding of the talked-about object and not talked about object during referential utterances.*

	Extended		Short		Sig.
	TRGT	DIST	TRGT	DIST	
Child Gaze	.56 (.14)	.07 (.05)	.44 (.14)	.12 (.07)	***
Child Holding	.50 (.18)	.15 (.12)	.45 (.19)	.17 (.11)	**
Parent Gaze	.52 (.16)	.04 (.04)	.49 (.16)	.07 (.05)	***
Parent Holding	.52 (.18)	.09 (.08)	.48 (.19)	.13 (.09)	**

Note: **TRGT**: Target object being referred to; **DIST**: Distractor objects not being talked about; **Sig:** significance of paired-samples *t*-tests on the proportion of all object looking or holding that was directed to the target object for utterances in extended vs. short discourse; ***p <= .001, **p <= .01.

Joint Engagement in Extended vs. Short Discourse

To examine the social correlates underlying extended verbal discourse, we asked whether toddlers' and parents' act in a more coupled manner during extended discourse. We considered toddlers and parents to be coupled if their looking and/or holding were synchronously directed towards the same object. As Figure 5 indicates, there are multiple ways dyads could be coupled. Although these forms of coupling are not mutually exclusive (e.g., toddlers might simultaneously be looking at what their parents are holding *and* what their parents are looking at), we chose to analyze them individually because the relevance of each form of coupling has been suggested in the literature (see Yu & Smith, 2013).

For each dyad, we measured frame-by-frame whether or not the dyads were coupled. We then examined the proportion of time that dyads were coupled within each utterance. We considered both the time that dyads were coupled on the talked-about object, or *target coupling*, and the time that dyads were coupled on any object, or *total coupling*[2]. Figure 6 compares the degree of coupling within utterances of extended discourse to coupling within utterances of short discourse. As the figure illustrates, across all forms of dyadic coupling (and across both measurements of coupling), we observed greater coupling in extended discourse than in short discourse. These findings demonstrate that extended discourse co-occurs with social interactions that are richer in joint attention and joint action, illustrating yet another nonverbal reason why participating in extended discourse may benefit children's development.

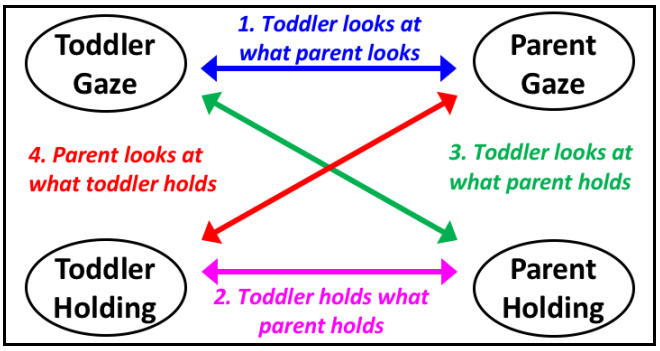

Figure 5. *Four non-mutually exclusive ways toddler and parent behaviors could be coupled.*

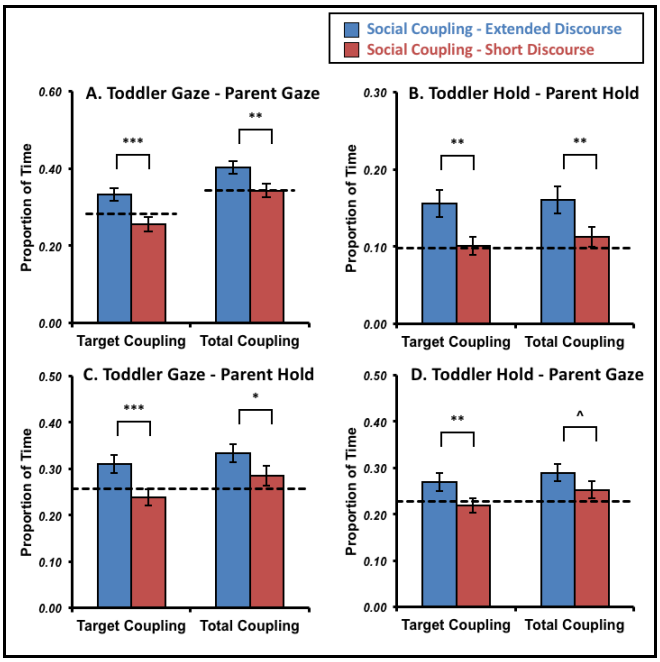

Figure 6. *The degree of social coupling (both to the target object and to all objects) within utterances across discourse types. Horizontal dotted lines reflect baseline coupling across the entire interaction.*

To what extent might the observed greater coupling be due simply to the fact that in extended discourse, toddlers and parents may have been more attentive and active in general (e.g., looking at and manipulating objects more frequently)? If this were the case, greater degrees of coupling in extended discourse than in short discourse would be expected by chance alone. To address this concern and provide a more rigorous test of coupling, we adopted a signal detection analysis that controls for base rate levels of toddlers' and parents' behaviors. Briefly, we classified each

[2] In computing the total coupling between toddler and parent gaze, we considered time spent looking at each other, or *mutual gaze*, as coupled.

frame into *hits* (e.g., toddlers look at an object that parents look at), *false alarms* (toddlers look at an object that parents do not look at), or *misses* (toddlers do not look at an object that parents look at). The *F-score* from this classification scheme served as our primary measure of coupling. The *F-score* is a combination (the harmonic mean) of *Precision* (hits / hits + false alarms), which in our case controls for toddler behavior, and *Recall* (hits / hits + misses), which in our case controls for parent behavior. We measured coupling for each dyad in three contexts: (1) in extended discourse utterances, (2) in short discourse utterances, and (3) across the entire interaction, which reflects baseline degrees of coupling. As Table 2 indicates, toddlers and parents were more coupled during episodes of extended discourse than during episodes of short discourse. Additionally, the degree of coupling in extended discourse was greater than baseline levels of coupling. Thus, this more stringent analysis confirms that episodes of extended verbal discourse are also episodes with rich joint engagement.

General Discussion

In the current study, we show that during the course of parent-toddler object play, multiple aspects of toddlers' language environment converge in real-time, generating segments of interaction that are potential gold mines for word learning. These segments are extended episodes of verbal discourse. Although previous research (e.g., Frank et al., 2013; Messer, 1980) has uncovered linguistic features that make these episodes valuable for learners, we revealed attentional and social correlates to these episodes that would also benefit learning. We found that within extended discourse, toddlers' and parents' attention were focused on the referent object for an extended period of time, making the topic of parents' speech especially clear and transparent. Previous observational (Yu & Smith, 2012), experimental (Hollich et al., 2000; Tomasello & Farrar, 1986), and individual-difference (Cartmill et al., 2013) research provide converging evidence that transparency in parents' speech is important for toddlers' word learning. Thus, extended discourse about an object may facilitate learning through an increase in the referential transparency of parents' speech.

We also observed that within extended discourse, toddlers and their parents displayed great degrees of social coordination. The importance of social coordination, or joint attention, for multiple facets of development is widely accepted and well established (e.g., Moore & Dunham, 1995). Although research in this area typically focuses on the role of joint *visual* attention, recent research suggests the relevance of other forms of joint attention as well. For example, Yu and Smith (2013) found that in the complex contexts of parent-toddler interactions, attending to partner manual activity may be the key way by which they share attention. In a similar vein, Deak and colleagues (2014) argue that following parents' manual actions may be a stepping stone for learning to follow parents' gaze. The fact that in extended discourse we observed heightened rates of many forms of joint attention suggests then that extended discourse may facilitate toddlers' development in part by offering redundant pathways to attention sharing, and by providing fertile grounds for training socio-cognitive development.

The current data do not only show a correlation between the duration of verbal discourse and the *duration* of attention and social coupling. The data also demonstrate a correlation between the duration of verbal discourse and the *intensity* of attention and social coupling. When our analyses honed in on individual utterances, utterances that were part of extended discourse were characterized by *more* toddler attention, *more* parent attention, and *more* social coupling. Considering the correlational nature of these findings, we cannot speak to their precise causal underpinnings. It is possible for example that more focused and sustained toddler attention to an object leads to parents' extended verbal discourse about that object. Alternatively, extended talk may actually play a role in focusing and sustaining toddlers' attention on the talked-about object (see Baldwin & Markman, 1989). And of course it is also possible that the influence is multi-directional; toddlers' attention, toddler-parent coupling and extended verbal discourse may bootstrap each other, producing the temporal

Table 2. *Mean coupling scores (F-Scores) across discourse length and at baseline.*

	Target Coupling			*Total Coupling*		
Form of Social Coupling	Extended	Short	Baseline	Extended	Short	Baseline
Child Gaze – Parent Gaze	.59 (.13)	.52** (.16)	.46*** (.09)	.47 (.12)	.42** (.13)	.42* (.08)
Child Hold – Parent Hold	.30 (.15)	.21*** (.13)	.14*** (.07)	.22 (.15)	.15** (.10)	.14*** (.07)
Child Gaze – Parent Hold	.56 (.13)	.50* (.16)	.37*** (.11)	.47 (.13)	.40** (.15)	.37*** (.11)
Child Hold – Parent Gaze	.50 (.15)	.45* (.14)	.35*** (.10)	.42 (.15)	.36** (.12)	.35* (.10)

Note: Asterisks reflect significance tests between coupling in extended discourse utterances and short discourse utterances, and between extended discourse utterances and baseline; * $p < .05$, ** $p < .01$, *** $p < .001$

dynamics and inter-relations that we observed. Either way, the results highlight a tight link between the verbal and non-verbal aspects of toddlers' language learning experience.

Conclusion

One view of early word learning is that the complexity of toddlers' learning environment makes it an especially challenging task. From a sea of information to which they are exposed - numerous words, several objects, and various social signals - they must figure out how the words they hear relate to the world they see. To make matters worse, toddlers are learning words while they are still figuring out the regularities of their perceptual world and the quirks of their social world. A different view of word learning is that the sea of information provides helpful information and that the multi-tasking makes word learning easier not harder. The current data provide evidence in line with this latter perspective, revealing a learning environment rich with redundancies and correlations between its linguistic, perceptual and social dimensions. Although much more work is needed, it may be that it is through extracting the latent structures from these interconnected dimensions and through solving multiple, mutually-constraining tasks that toddlers come to learn words as effortlessly as they do.

Acknowledgments

We thank many members of the Computational Cognition and Learning Laboratory, especially Maddie Bruce, Danielle Rosenstein and Jessica Steinhiser, for their assistance in this research. This research was supported in part by the National Science Foundation (BCS 092428) and the National Institutes of Health (R01-HD074601, R21-EY017843, K99-HD082358).

References

Adamson, L. B., & Bakeman, R. (2006). Development of displaced speech in early mother-child conversations. *Child Development, 77*, 186-200.

Adamson, L. B., Bakeman, R., & Deckner, D. F. (2004). The development of symbol-infused joint engagement. *Child Development, 75*, 1171-1187.

Baldwin, D. A., & Markman, E. M. (1989). Establishing word-object relations: A first step. *Child Development, 60*, 381-398.

Bard, E. G., & Anderson, A. H. (1983). The unintelligibility of speech to children. *Journal of Child Language, 10*, 265-292.

Cartmill, E. A., Armstrong III, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 11278-11283.

Deak, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science, 17*, 270-281.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development, 82*, 1738-1750.

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning & Development, 9*, 1-24.

Hoff-Ginsberg, E. (1986). Function and structure in maternal speech: Their relation to the child's development of syntax. *Developmental Psychology, 22*, 155-163.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H., … Rocroi, C. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development,* 65(3, Serial No. 262).

Messer, D. J. (1980). The episodic structure of maternal speech to young children. *Journal of Child Language, 7*, 29-40.

Moore, C., & Dunham, P. J. (1995). *Joint Attention: Its Origins and Role in Development*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition, 109*, 423-430.

Pereira, A. F., Smith, L. B., & Yu, C. (2008). Social coordination in toddler's word learning: interacting systems of perception and action. *Connection Science, 20*, 73-89.

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

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 12663-12668.

Salley, B., Panneton, R. K., & Colombo, J. (2013). Separable attentional predictors of language outcome. *Infancy, 18*, 462-489.

Sullivan, J., & Barner, D. (2016). Discourse bootstrapping: preschoolers use linguistic discourse to learn new words. *Developmental Science, 19*, 63-75.

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

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition, 125*, 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*, e79659.