Language Acquisition, 2022

Discourse with Few Words: Coherence Statistics, Parent-Infant Actions on Objects, and Object Names

Hadar Karmazyn-Raz[1] and Linda B Smith[1,2]

[1]*Indiana University, Bloomington, US*

smith4@indiana.edu

[2]*University of East Anglia, Norfolk, UK*

Acknowledgements:

This research was supported by the National Science Foundation grant BCS-1842817. The authors thank Daniel Yurovsky who inspired the experimental context and collected the data from 7 of the participating dyads as a pilot study, as well as the comments and criticisms of Elizabeth Clerkin, Erin Anderson, Jeremy Borjon, and Richard Betzel.

Corresponding author:

Linda B. Smith

Discourse with Few Words: Coherence Statistics, Parent-Infant Actions on Objects, and Object Names

Abstract

The data for early object name learning is often conceptualized as a problem of mapping heard names to referents. However, infants do not hear object names as discrete events but rather in extended interactions organized around goal-directed actions on objects. The present study examined the statistical structure of the *nonlinguistic* events that surround parent naming of objects. Parents and 12-month -old infants were left alone in a room for 10 minutes with 32 objects available for exploration. Parent and infant handling of objects and parent naming of objects were coded. The four measured statistics were from measures used in the study of coherent discourse: (1) a frequency distribution in which actions were frequently directed to a few objects and more rarely to other objects; (2) repeated returns to the high-frequency objects over the 10-minute play period; (3) clustered repetitions, continuity, of actions on objects; and (4) structured networks of transitions among objects in play that connected all the played-with objects. Parent naming was infrequent but related to the statistics of object-directed actions. The implications of the discourse-like stream of actions are discussed in terms of learning mechanisms that could support rapid learning of object names from relatively few name-object co-occurrences.

Discourse with Few Words: Coherence Statistics, Parent-Infant Actions on Objects, and Object Names

Infants reliably show receptive knowledge of some common object names by their first birthday and before they say many if any words. We know this because infants near their first birthday look to pictures of infant-common objects upon hearing the name of the object (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012; see also, Kartushina & Mayor, 2019; Steil, Friedrich & Schild, 2021). This looking behavior shows, at minimum, a remembered link between a heard name and the visual properties of category members. The field has little understanding of the experiences that support this early learning (e.g., Clark, 2019; Smith, Suanda, Yu, 2014; Yurovksy, 2018). Prior work has focused on how infants upon hearing a name manage to map the name to the right referent and on the statistics of the name-object co-occurrences that might support that learning (see Smith et al, 2014 for review). The difficulty is that the contexts in which infants with limited productive vocabularies hear object names—contexts such as mealtimes, taking a bath, toy play ---are not organized around a lot of talk and the evidence shows that parents offer names of objects in these contexts at best infrequently (Tamis-LeMonda, Custode, Kurchirko, Escobar & Lo, 2019; Tamis-LeMonda, Kurchirko, Luo, Escobar, & Bornstein, 2017). We propose that the solution to this seemingly insufficient input lies in the extended-in-time structure of the whole interaction, in the *coherence of the non-linguistic events,* that surround occasional naming of objects by parents.

Developmental research has long pointed to parent-infant play with objects as a form of conversation and shared reference between parent and infant (e.g., Bateson, 1975, see also Monroy, Chen, Houston & Yu, 2021). However, little is known about the extended temporal coherence of these action "conversations." We used the literature on the coherence of discourse (Clark, 2019; Tyler, 1994; Frank, Tenenbaum & Fernald, 2013; Schwab & Lew-Williams, 2016) as guidance to the potentially relevant statistics of the parent-infant object-directed actions that surround parent naming.

Behavior in Time

Our hypothesis and analytic approach begins with the well-documented natural statistics of human behavior (Barabasi, 2010; Zipf, 1949): A wide variety of human-generated events – the words in spoken and written language (e.g., Altmann, Pierrehumbert, & Motter, 2009; Piantadosi, 2014; Montag, Jones & Smith, 2018; Serrano, Flamini, & Menczer, 2009); the places people visit (Krumme, Llorente, Cebrian & Moro, 2013; Boyer & Solis-Salas, 2014), and the object categories encountered (Clerkin, Hart, Rehg, Yu & Smith, 2017; Li, Su, Lim & Fei-Fei, 2010; Salakhutdinov, Torralba , & Tenenbaum , 2011) show characteristic frequency and temporal distributions. The possible types –different words, places, object categories –show frequency distributions that are extremely skewed; a very few types are highly frequent, but there are many additional types that can occur and do so but infrequently. The occurrence of specific types is also context dependent, so the likelihood of a type occurring is much higher if it has just occurred. This statistical property yields bursty close-in-time repetitions of the same type. Context dependency is also evident in the different types that occur close together in time (e.g., Hills, 2013; Oliva & Torralba, 2007; Tamis-LeMonda et al, 2019). For example, at breakfast, the word *cup* is more likely to be said close in time to the word *bowl* than to the word *elephant*; a person is more likely to visit Boston on the way to Cape Cod than to visit Chicago; and one is more likely to find the objects cup and bowl together in the cabinet than to find a cup next to a toy elephant. In his seminal work on human behavior, Zipf (1949) proposed that these statistics are pervasive across phenomena and domains because they reflect fundamental properties of human motivation and cognition.

Although conceptualized differently, these same statistics are also well-known to characterize discourse and conversations. There are a few highly frequent topics that occur in clusters of continuity and are returned to over the course of the whole conversation (Dijk, 2010; Tannen, 2007; Tyler, 1994). Conversations also include less frequent topics, sometimes characterized as comments on those topics (Chafe, 1987), that connect in meaningful ways to the more frequent topics (Gernsbacher, 2013; Halliday & Hasan, 1976; Tyler, 1994). Theory (e.g., Gernsbacher, 2013), experiments (e.g., McKinley, Brown-Schmidt & Benjamin, 2017;Carvalho, Chen & Yu, 2021; Benitez, Zettersten & Wojcik, 2020), and computational models (e.g., Folz, 2007; Pulver, Kording, Griffiths & Tenenbaum, 2006; Salakhutdinov, 2011) all indicate that these statistics of natural discourse play a strong role in communicative efficiency (Bohn & Frank, 2019) through the maintenance of common ground (Clark & Benicot, 2008), in the integration of new and old information (Keenan, MacWhinney & Mayhew, 1977), and in the learning of words (e.g., Hendrickson & Perfors, 2019; Kurumada, Meylan, & Frank, 2013; Lavi-Rotbain, & Arnon, 2021).

The present study

Our empirical goal was to characterize the extended-in-time coherence of the parent-infant object-directed actions that surround parents' rarer naming of objects. Because the study of discourse is specifically about characterizing time-extended coherence, we report on four statistics suggested by that literature: (1) a *selective focus* on a small set of objects that is characterized by (2) patterns of *returns* to the same objects throughout the interaction and (3) by clusters of *continuity* (4) with the whole interaction showing systematicity in patterns of transitions among different objects. The contribution of the present study lies in documenting the statistics of the non-linguistic events that surround *one-sided* parent talk to non-talking infants. We measure these coherence statistics in the context of parent-infant object-directed actions and parent-object naming during free-flowing toy play with limited demand characteristics or experimenter-induced constraints. Parents and their 12-month-old infants played in a room with no experimenters present and in which there was a haphazard collection of 32 objects available for exploration; naming of the objects or an interest in early word learning was not mentioned to the parents.

Method

Participants

The participants were 16 infant-parent dyads. All infants were within +/- 2 weeks of their first birthday and half were female. The sample of infants was broadly representative of Monroe County Indiana (75 % European American, 6% African American, 6% Asian American, 6% Latino, 6% Mixed race) and consisted of predominantly working- and middle-class families. All research was approved by the Human Subjects and Institutional Review Board at Indiana University. The predominant language in the home of all infants was English.

Toys, playroom and instructions to parents

The 32 toys were a haphazard selection of the variety of toys that end up in the toy boxes of homes and waiting rooms (and were specifically selected from the toys in the waiting room of the laboratory). The toys are listed in the Appendix. We selected the toys in this way: (1) to offer choice to parents and infants about play and thus to limit experimenter-imposed demand features and (2) to offer toys that were typical of the kinds of toys encountered by infants this age. For each participating dyad, the 32 toys were scattered on the floor before the infant and parent entered the toy-room and thus the spatial placement differed for each dyad.

Parents and their infants were invited to play in a toy room (3 by 4 meters) for 10 minutes without an experimenter present. Parents were told that we were interested in how infants and played and they should play with their infant as they would when playing with toys at home. Parents were asked to keep their infant engaged for the entire 10-minute session. There was no mention of an interest in object names or language.

Audio and Video Recording

Because the only environmental events that matter for learning are the ones that contact the learner's sensory system (Smith, Yu, Yoshida, Fausey, 2015), we used a head-camera and audio recorder worn by the infant to record the interaction. As shown in Figure 1, the head camera was situated low and firmly on the infant's head. The head camera is a commercially available small (22 g) camera (watec) with a 90-degree (diagonal) field of view. Two additional third person cameras (overhead and from the side) also recorded the interaction but were only consulted with if there was an ambiguity with respect to the identity of an object in a participant's hand.

The raw video and audio were coded by human coders using Datavyu (Admin, Gilmore, Seisler, & Soska., 2018). The coders first determined every naming event by a parent. Names were defined as common nouns. Any name used by a parent was counted; for example, if a parent called the toy sheep "dog", "dog" was coded as a name for that referent for that infant. Variants of the same noun ("dog," "doggie," "doggies") were counted as instances of the same type. Three independent coders transcribed object names and two of the three coders had to agree for a name to be registered as produced by the parent. Agreement by all three coders for each naming event exceeded 84%. The images that coincided with a parent naming event were coded for parent and infant handling of objects. The remainder of the images –those that did not coincide with a naming event -- were sampled at .2Hz and were also coded for parent and infant handling of objects. Coders annotated all objects in all hands: whose hand and the specific object. Within a sampled frame, neither participant could be handling an object, one but not the other could be handling an object, both could be handling different objects, one or both could be handling more than one object, or both could be handling the same object. All objects in active contact with a participant's hand were coded as being handled by the participant. Three coders independently coded images with a rule of two out of three agreement on the person and object being handled for the handling event to be counted. Agreement by all three coders per handling event was 89%.

The sampling rate of .2Hz (one frame sampled every 5 seconds) has been shown to yield reliable and valid estimates when compared to more dense sampling rates of 2Hz and 1Hz, (Fausey, Jayaraman & Smith, 2016; Jayaraman, Fausey & Smith, 2015). To ensure this was the case in the present sample, analyses were conducted for 2 dyads with images sampled and coded at 2Hz and 1Hz. These analyses yielded no reliable differences in the reported statistics determined at these three rates for these 2 dyads.

Data Availability

The de-identified data is available [https://osf.io/254yq/].

Results

The analyses focus on the statistics of the nonlinguistic object-directed actions that surround the rarer naming events. However, when possible we also conduct similar analyses on the much less frequent parent naming events. Object handling, as expected, was pervasive throughout the interactions; at least one participant within a dyad was handling at least one object on 76% of all sampled images: 71% of the handling consisted of just *one* of the two partners holding just *one* object. Cases (% all frames with handling) in which one partner held two objects were relatively rare: 8% for infants and 20% for parents. Parents and infants handled objects roughly equally. The mean proportion of sampled frames in which infants handled an object was .54, SD =.11; the mean for parents was .45, SD =.11. Naming was strictly a parent behavior and no infant provided a name for any object during play. On average, parents provided object names 36 times (SD =14) in the 10-minute play period; these were distributed on average among 14 of the 32 toys (SD=3.1). Thus, individual objects were named on average only 2.5 times during play. Because of the multiple measures on the same data, we set *a priori* the alpha level for statistical reliability to .01 for all comparisons.

A Selective Focus on a Few Objects

All dyads were selective, engaging with only some of the available toys. The mean number of the 32 available toys that was handled at least once was 11 for both infants (range 3-16, SD=2.7) and parents (range 3-15, SD=2.8). That is, most of the toys were not handled at all by an individual dyad. This is not because there were just a few toys of interest to all dyads such that the dyads all played with the few same toys and ignored the others. All 32 toys fell in the top 11 toys for at least one dyad and 26 different toys served as one of the top 5 most-frequently engaged toys within a dyad. In a cluttered room with many different items to explore, parents and infant did not sample all of them, but concentrated on just a small number, a substantial reduction of potential competitors for any heard object name.

The selective focus of dyads on just a few toys is magnified by the skewed frequencies with which they handled individual toys. We measured for each participant the frequency with which the participant handled each unique toy. From these data, we determined separately for parent and infant the rank-order frequency distribution of handled toys expressed in terms of the proportion of all of the individual's handling that fell at each rank. We also determined the rank-order frequency distribution at the dyad level, that is, for the combined handling that comprises the interaction: specifically, at each sampled frame, we counted objects handled by any participant without regard to who handled them. The means of these three rank-order frequency distributions of handling -- child separately, parent separately, and the dyad handling-- are shown in Figure 2 A, B, and C. We compared the skew of the distribution of parent and infant handling separately for each dyad. The mean Wilcoxin statistic $V = 63.4$ (range 12-123, SD=32); for all individual dyads, $p>0.01$. In brief, parents and infants generated comparably skewed distributions. Moreover, the Spearman correlations of raw frequency for the handling of specific objects between parent and infant within a dyad, calculated separately for each dyad, yielded a mean of .72; the range for individual dyads was 0.61- 0.99 and for all within dyad correlations, $p < .01$.

Parent frequency of naming individual objects was less skewed as shown in Figure 2D than handling, when compared to both infant handling (average $V = 237$, range 109.5 -321, SD=74.21, $p < 0.01$ for all dyads) and to parent handling (average $V = 244$, range 127-351.5, SD=74, $p < 0.01$ for all dyads). However, the objects most frequently named by parents correlated strongly with the frequency of dyad handling: mean correlation, calculated separately for each dyad was .65 (range 0.44 to 0.92) and was reliable within a dyad, at the $p<.01$ level for 15 of the 16 dyads.

In Figure 2, the dotted lines show the expected proportion of behaviors directed to an object if at each moment participants randomly sampled an object from the set of 32 available toys. Using the dyad handling data, we formed two categories of engaged toys shown in the different colors in Figure 2C: the most frequent 5 toys with mean frequencies that fall above the baseline expected from a random selection of toys and the remaining engaged toys whose frequency did not differ from that expected by chance. We use this frequency partition in the following analyses to compare coherence statistics for the higher and lower frequency items. The coherence statistics could characterize the higher frequency objects, the objects of clear interest within a dyad, but not the lower frequency objects. However, if the entire temporal stream of object-directed actions during the interaction has an *overall* coherence then similar and/or different statistics for the higher and lower frequency objects may be revealing about that structure. We used the same partition for all dyads: the 5 most frequently handled toys versus the remaining toys that were handled.

We borrow the terms "Topic" and "Comment" as labels for the higher and lower frequency toys. These labels are sometimes used to characterize the information structure of discourse (Chafe, 1987). In that literature "Topic" refers to high-frequency items that are the main theme of discourse and thus the "given" information whereas "Comment" refers to low-frequency new information related to the main themes. The 5 Topic objects within a dyad were clearly the focus of the parent and infant object-directed interactions, accounting for 76% (range 58-94%, SD=11%) of all child handling and for 69% (range 50-94%, SD=13%) of all object parent handling. Dyad Topics were also the focus of parent naming, with each Topic object named on average 4.5 times during the play session (SD =1.7, range 1.8 to 8.2); Dyad Comment objects were named very rarely, on average 0.47 times (SD=0.24, range 0.21 to 1.1) during the play session.

In summary, the frequency distributions of parent-and-infant manual actions on objects showed a *shared* pattern of selective focus on a few objects but included infrequent actions directed to other objects. Parent naming was centered on the few high frequency objects.

Returns to topics

What does it mean for a time-extended interaction of any kind to be coherent? Analyses of coherent discourse (e.g., Tyler, 1994) suggest that main themes should play a role throughout the entire discourse. Is the stream of parent and infant object-directed actions coherent in this sense? It could be that parents and infants initially played with one object, became bored, moved to another object until bored, continuing in this way without returns to previously engaged toys. Indeed, this is a pattern that might be expected by some curiosity-based theories of infant exploration (Twomey and Westermann, 2018) or studies of habituation (Johnson & Brody, 1977). Alternately, play may be holistically coherent; the high-frequency objects may deserve the label of Topics because those are the toys returned to throughout the interaction. Accordingly, we partitioned the 10-minute session into 2.5-minute quarters and examined the frequency of play within each quarter of play. Figure 3A shows the presence of dyad actions directed to Topic and Comment objects in each quarter and uses stacked bars to show the overall frequencies for individual (by rank order) Topic and Comment objects. For Dyad Handling, an ANOVA for repeated measures revealed that Topic objects were more frequently handled than Comment objects ($F(1,15)=193.78$, $p<0.01$); there was no effect of Quarter ($F(3, 45)=0.69$, $p=0.55$) and no interaction ($F(3,45)=2.319$, $p=0.07$). For every dyad, each of the 5 Topic objects was engaged at least once in at least 3 of the 4 quarters, mean number of quarters for engaged = 3.2 , range 1-4, SD=0.6. This result is consistent with the view of Topic objects serving as the organizing themes of play throughout the session.

Actions directed to unique Comment objects, in contrast, typically occurred in only one quarter; mean number of quarters in which each Comment object was engaged = 1.4 , range 1-4, SD=0.75. If we consider only the 5 most frequent Comment objects, there was also limited return over the course of play; mean number of quarters engaged at least once = 2.1, range 1-4, SD=0.78. The Comment objects, interleaved between the multiple returns to the Topics, thus provide a form of new information although this new information likely does not reach the standard of a comment about a Topic object. Our main conclusion from this analysis is this: the entire parent-infant interaction shows a time-extended coherence in repeated returns to the high-frequency objects with interleaved infrequent interactions with other objects between those returns.

Figure 3B shows the frequency of parent naming directed to Topic and Comment objects in each quarter. The ANOVA for parent naming revealed a small effect of Topic versus Comment objects (F(1,15)=4.5, p=0.03) that did not meet our *a priori* threshold for reliability and there was no effect of Quarter (F(3),45=0.910, p=0.43) and no interaction (F(3, 45)=0.45, p=0.68). However, parent naming for each of the five Topic objects within each dyad occurred at least once in at least 3 of the 4 (M = 2.8, range 1-4, SD=0.8) quarters of play. The naming of individual Comment objects averaged fewer than 1 per the entire session and thus there were almost no cases of returns in parent naming of Comment objects that spanned quarters of play (M = 1.12 , range 1-3, SD=0.67).

In summary, across the entire 10-minute interaction, parent-infant handling of objects returned to the same high-frequency objects. The other objects, engaged with low frequency, are interleaved among the main themes but do not systematically recur over the entire play period.

Continuity

One of the hallmarks of discourse statistics is the continuity of topics in time. Frank, Tenenbaum & Fernald (2013) developed a Discourse Continuity metric to quantify successive repetitions of reference to an individual object in parent talk to toddlers. We used a modified version of this metric to quantify the continuity of actions on objects by infants and parents, and the continuity of parent naming. We computed the Probability of handling continuity (P_{HC}), that is, the probability of handling a specific toy given it was handled in the last sampled frame, at the infant, parent, and dyad level. Then for each object, *o*, we defined the handling function $H_t(o)$ as a delta function returning whether or not the object was handled at time *t*. We then define $P_{HC}(o)$, the probability of handling continuity for each object, as:

$$P_{HC}(o) = \frac{\sum_t H_t(o)H_{t-1}(o)}{\sum_t H_t(o)}$$

Following Frank et al (2013), we estimated baseline values for P_{HC} using a permutation analysis. The permutation analysis determines the expected values for P_{HC} if object sampling at each moment is independent of the previous moment. We calculated the baseline at the corpus level by recomputing the continuity values for each ranked object for 10,000 random permutations of the observed object handlings. Similarly, we used this metric to quantify successive repetitions of parent naming of individual objects. The Probability of naming continuity (P_{NC}) was computed as the probability of parent naming of a specific toy given it was named in the last sampled frame. Then for each object, *o*, we defined the naming function $N_t(o)$ as a delta function returning whether or not the object was named at time *t*. We then define $P_{NC}(o)$, the probability of naming continuity for each object, as:

$$P_{NC}(o) = \frac{\sum_t N_t(o) N_{t-1}(o)}{\sum_t N_t(o)}$$

we estimated baseline values for P_{NC} following the same permutation analysis for P_{HC}.

Figure 4 shows the continuity probability of infant, parent, and dyad handling and the continuity probability of parent naming. The dot clouds show these measures at the individual participant and dyad levels. The continuity probability of all infant handling, 0.13, SD=0.05, was significantly higher, $t(15)=6.01$, $p<0.01$, than the baseline, 0.02, SD=0.01, as was all parent handling, 0.15, SD=0.04; $t(15)=12.52$, $p<0.01$. The continuity probability of Dyad handling, 0.3, SD=0.05, was higher than baseline, $t(15)=18.1$, $p<0.01$, and noticeably higher than parent and infant handling alone, as handling by one partner was often followed by handling of the same object by the other partner, a dyad behavior pattern that increases continuity at the dyad level over that at the individual participant level. In brief, the stream of behaviors directed to higher and lower frequency objects, like the stream of words in discourse, is not randomly distributed in time but instead the actions of both participants on the objects create a temporal structure of continuity in which the probability of acting on an object is greater given that an action directed to that object has just occurred. The fact that both the frequent Topic and the less frequent Comment objects show comparable moment-to-moment continuity is intriguing. Comment objects may only be engaged for brief portions of the entire play session (falling in one quarter of the play period) but when they are engaged, they are engaged in a cluster of close-in-time repetitions. This suggests that these are not accidental events but rather that handling of the Comment objects is meaningful to the participants in the moment in which they are handled.

The probability of naming continuity, 0.12, SD=0.06, was also higher than baseline (M= 0.02 SD= 0.08) $t(15)=5.97$, $p<0.01$. Note that continuity can be computed only for items that occur at least twice in the time series and thus the data for the naming of Comment objects shown in Figure 4 is over a very small number of events for each dyad.

Relational Patterns

All the measures thus far are computed at the individual object level. Coherence also implies relations among all the events in the time series of behaviors. Consequently, we ask whether the momentary selection of an individual toy for handling is related the whole times series of handled toys. To answer this question, we used network analyses to measure transitions among object-directed actions to different toys in play. We did not do a similar analysis of parent naming because naming was too sparse to yield meaningful relational patterns among the named objects.

The expectation in performing the network analyses of object-directed actions is not that we will find relational patterns indicative of thematic or taxonomic relations among different objects. The infants– all within 2 weeks of their first birthday – explored the objects in ways that are typical for infants near their first birthday (e.g., Bourgeois, Khawar, Neal & Lockman, 2005; Lockman & Tamis-LeMonda, 2021): they banged objects on the floor, they banged horses with hammers, they banged giraffes with horses, they held up and showed different objects in a series or handed them to their partner, they looked at objects upside down and sideways, they fingered details, they put all kinds of things in the bucket. If parents and infants are creating a form of discourse through play, the discourse appears innovative and infantile. The goal of the network analyses is not to determine the meaning of the "conversation of actions" in any adult sense but to quantify the connective patterns in the transitions from one toy to the next.

To build the networks, we first determined the structure of successive object-directed actions from one timestamp to the next for each dyad and for the permuted baseline time series used in the Continuity analyses. Between one sampled time to the next, there are five possible transitions: (1) Continued handling of the same object (SO); (2) Shift from one handled object to a new object (NO); (3) Continued not handling any object (**), (4) Transition from not handling to handling an object (*-), and (5) transition from handling an object to not handling any object (-*). Figure 5 shows an illustration of these transitions coded at the dyad level as well as the proportion of each possible kind of transition in the observed time series and randomly permutated baseline. Notice that continuity of handling (SO) constitutes a significant proportion of the observed transitions but not the randomly created transitions, showing the same continuity result reported above. The key question we want to answer in this section, however, is not about continuity but about the character of the shifts from one object to a different one.

Accordingly, we focused on shifts from one handled object to a different one, the data listed as New Object (NO) in Figure 5. The whole series of shifts from one object to another (the set of NO transitions) can be represented by a network of nodes, with each node indicating an object and each edge indicating a NO transition from one node to another. Figure 6A shows the observed network created at the corpus level by combining the data across all dyads with nodes representing rank order frequency and edges of all the observed transition. The random network, Figure 6B, represents the data at the corpus level for all NO events in the permutated data. The networks include 24 nodes, all ranks that had at least one edge (one dyad made the rank-to-rank transition at least once). Nodes in maroon indicate the Topic objects and those in orange the Comment objects. The edges connecting the nodes indicate the action transitions from one object-node to another and are non-directional, that is a NO transition from a rank 1 object (e.g., hippo) to a rank 8 object (e.g., bucket) is treated as equivalent to a NO transition from a rank 8 object (bucket) to a rank 1 object (hippo). The darkness and length of each edge illustrates the weights, calculated from the frequency (or strength) of each transition. More frequent action transitions between one node to another are visualized as thicker and shorter edges.

Highly structured networks are ones in which some nodes and some edges are more important or play different roles than others. Within graph theory, centrality measures estimate the relative importance of nodes and edges over others with respect to the connectivity pattern of the whole network. Accordingly, we compared the observed and the random networks using three common measures of centrality: degrees, closeness and betweenness (Freeman,1978; Boccaletti et al., 2006; Rubinov & Sporns, 2010) . The random network instantiates the null hypothesis that all effects are due to the frequencies of the object-directed actions. Thus, the comparison of the observed and random network provides information on the relational structure among different toys that is above and beyond the effects of frequency. The network measures were computed using the centrality function available in Matlab (Math Works, 2020). **Degrees** measure the number of unique edges connected to each node; toy objects that were connected to many different toys (independent of the frequency of the unique connections) have greater *degree.* **Closeness centrality** is measured as a **distance** and is an index of strength, that is, the frequency of individual edges and, in this case, the frequency of individual toy-to-toy NO transitions. More precisely, closeness centrality measures the distance of the nodes in the networks as an inverse sum of the weights (a function of frequency) from each node to all other nodes in the network. Higher values of edge weights specify closer, that is, more frequently NO-connected nodes and lower values specify greater distance or less frequent NO-connected nodes. **Betweenness centrality** measures the number of times a node appears on the shortest path between two other nodes in the network. For example, if the hippo and the horse are never successively connected but are often connected through the bucket in triads of NO events such as hippo-bucket-horse and horse-bucket-hippo, then the bucket would be on the shortest path between the hippo and the horse. For each of these measures, we compared the Observed (O) and Random (R) network (with each node in the network serving as the random variable) using two-tailed t-tests with an *a priori* set acceptable alpha level of .01. These analyses compare the structure of the whole network without regard to Topic or Comment objects. Accordingly, we also show the values of each measure for each individual node in Figure 6.

The mean number of degrees per node was reliably smaller for the Observed, $M = 11.25$, $SD = 5.6894$, than Random network, $M = 15.75$, $SD=5.67$, $t(23) = 8.94$, $p < .001$. This metric indicates that there are fewer –that is, more selective – edges connecting one toy to another in the Observed than Random network; all the possible edges are not realized, an indication that the transitions are selective beyond what would be expected by the frequency distribution alone. This can be seen in the degrees for individual nodes in Figure 6. A.1. The observed data relative to the random network data, shows greater selectivity for both the high-frequency Topic objects and the Lower frequency Comment objects. The mean distance of the nodes (the inverse of closeness) was reliably greater for the Observed, mean = 33.5, $SD = 6.03$, than Random network, mean = 63.5 $SD =8.4$, $t(23) =15.2$, $p < .001$. This result indicates that realized edges were repeated more frequently than would be expected by randomly connecting all the nodes. Notice that this difference holds for individual Topic as well as individual Comment nodes,

Figure A.2. The mean betweenness (the number of short paths that go through each node) is reliably more for the Observed , mean = 6.5, SD = 8.09, than the Random network, mean=3.6 SD= 4.8, t(23) = 2.2, p < .001. This indicates that a few core objects play a greater role in the connectivity pattern than would be expected by frequency alone. The connectivity pattern observed at the individual node level, Figure 6 A.3, shows that both Topic and some Comment objects show greater betweenness.

To further understand the relational role of Comment objects to Topic objects, we also exampled the likelihood of NO transitions that were between a high-frequency Topic and a low-frequency Comment object. Because actions on Topic objects are high-frequency events, transitions to and from topic objects should predominate. The degree to which these transitions depend solely on frequency is determinable by calculating the same statistics for the permutated series. The results show that when dyads shifted from one object to a different one, they shifted to a topic object *less often* than the expected baseline. Overall, when the dyad shifted from acting on an object, they shifted to acting on a topic object in 0.61 of the transitions, (SD=0.19) significantly *less* likely, $t(15) = 4.43$, $p < 0.01$, than the baseline expectation of 0.80 (SD=0.13). Put another way, dyads handle the Topic objects with high frequency but when they transition, they do so to the Comment objects more often than expected by frequency alone. This result suggests that the less frequently engaged objects, the so-called Comments, may play a meaningful role in structuring play.

The corpus network shown in Figure 6 aggregates the data across dyads by rank frequency; this is necessary because different dyads played with different toys. However, the aggregation by frequency also obscures possible dyad-specific meaningful relations among the different toys. Accordingly, we also examined the individual dyad networks with nodes representing specific toys. Two of these are shown in Figure 7. Overall, the individual networks and the two examples in Figure 7 have connectivity patterns similar to the observed corpus network. Topics are generally interconnected to many objects. Comment objects are more peripheral but a few are particularly well connected to other objects. For example, in Dyad 1's network, the horse (perhaps the protagonist of the narrative of actions) transitions directly (one step) to the ball, tiger, hammer, giraffe, baby doll, ring and cow, and thus the horse lies in the shortest path of all these objects to many other objects. But some comment objects –cow, sheep, ring –are each related to multiple topic objects. Dyad 2's network has a more complicated structure with the topic objects forming a strongly interconnected core, but comment objects also form interconnected groups. Just as conversations do not reflect a mere sampling of words from the frequency distribution of words in a language and are innovative, parent and infant actions on objects show a structure beyond mere frequency preferences for specific objects. Although the actions that connect different toys may not be particularly sophisticated they nonetheless create a coherence across the time-span of the whole interaction, a coherence that is evident in the returns to objects in the course of play, in the continuity of actions on individual objects and in the transitions from one toy to another. The infrequent naming of objects by parents is embedded in a rich and highly structured system of objected-directed actions.

Discussion

Early object name learning is often conceptualized as learning through ostensive definition. Given a heard name, meaning is found by mapping the name to a referent being attended to by the infant when the name is heard. Within this framework, the data for learning initial object names is a series of discrete name-object mapping events.   A large body of theoretical, experimental and observational research has been conducted within this framework focused on the quantity and quality of name-object co-occurrences (see Smith et al, 2014, for review). Despite all this effort, there is no accepted explanation for how –before they produce or know many words -- infants learn object names. There are three barriers to explanation. First is the sparsity of any individual object name in parent talk as object names, even common ones, are on the long tail of infrequent words in child-directed input (Sandhofer & Smith, 2000; Lui et al., 2008) and surprisingly infrequent in many everyday activities (e.g., Tamis-LeMonda et al, 2017). Second are the well documented immaturities in infant attentional and memory systems (see, Vlach, 2019; Wojcik, 2013, for reviews). Third are the many nameable entities present at moments when names are heard and thus likely spurious name-object mappings (e.g., Yurovsky, Smith & Yu, 2013). In the present study, we examined the statistics of the non-linguistic events that surround object naming in a continuous and extended social interaction. Figure 8 illustrates how isolated naming events are embedded in a coherent temporal series of actions which we propose engage and support learning mechanisms that make robust learning of object names from minimal experiences possible. We next discuss the implications of the temporal coherence of the surrounding nonlinguistic events with respect to the barriers for explanation and for infant learning and remembering of object names.

The statistics of coherence and mechanisms of attention and memory

Parent and infant actions on objects within a free-flowing and many-minute interaction show inter-related statistical properties: a skewed frequency distribution in which a few objects are frequently acted on and other objects are engaged rarely; temporal statistics in which actions on individual objects occur in clusters of continuity and the theme objects are returned to multiple times over the whole interaction; and a *coherence* across the whole interaction in the structure of transitions between different objects. When the partners shift from one object to a different one, there are repeated, selective and thus *predictable* patterns. As illustrated in Figure 8, if the dyad has just handled the cow, they are more likely to handle it again and, for the illustrated dyad, if they have handled the cow but shift to a different object, they are more likely to shift next to the hammer than to the giraffe. The observed statistics in parent-infant toy play share some similarities with the statistics of coherent discourse and effective communication. There is a beginning, middle, and an end with the same topics that are repeated and hold the whole together, and there is new information interleaved and temporally connected to the main topics. The observed statistics are also characteristic of human generated behavior more generally and of many natural phenomena in the world (Barabasi, 2010).

These are also statistical properties likely to support learning by immature cognitive systems. Selective repeated engagement with a very small set of available information and continuity reduces ambiguity as to the topic of current interest and thus spurious connections between words and referents. Recent research shows that the correlated handling of objects by parent and infant supports joint attention and common ground (Chen, Houston & Yu, 2021). The coordinated handling has also been shown to support sustained attention by the infant (Suarez-Rivera, Yu & Smith, 2019), an established factor supportive of robust infant learning (see Fisher, 2019, for review). Continuity and connectivity in the dyad transitions from action on one toy to another may also help organize attention to the right object at the right time serving as predictive cues as to meaning and what comes next (Adamson, Bakeman, & Deckner, 2004; Howard & Woodward, 2019; Masek et al, 2021; Sonne, Kingo, & Krøjgaard, 2018; Suanda, Smith & Yu, 2016). In brief, coherence statistics support and constrain attentional processes coordinating the thoughts and attention of both partners.

These coherence properties may also support durable memory formation and the learning of individual object names from just a few naming events. Recent advances (for reviews see, Hebscher, Wing, Ryan, & Gilboa, 2019, McClelland, 2013; Manohar, Zokaei, Fallon, Vogels, & Husain, 2019; Olivers & Roelfsema, 2020) indicate that robust memories may be established from quite limited experiences when the to-be-learned items emerge from goal directed actions within a social interaction. Theoretical analyses suggest that these factors – temporal streams of multimodal information within a social context– are potent because they lead to predictive relations in which each moment is constrained by the just previous events and biases the next moment, because these predictive patterns lead to persistent internal activation of selected contents, and because the temporal patterns include close-in-time repetitions that are repeated multiple times within the stream of events, causing the to-be-learned material to be retrieved and reactivated. The extant evidence (see Hebscher et al, 2019; Olivers & Roelfsema, 2020) also indicates that persistent activation or reactivation of one component of an association supports the rapid integration of novel information into that memory.

We propose that the coherent series of goal-directed actions of parents and infants creates selective and strong activation of representations of individual objects that enables the rarer naming event to be durably bound to that representation with minimal repetitions. This new hypothesis clearly requires a direct empirical test. However, we note that an increasing number of studies of infant memory are reporting the role of the implicated factors – social context, goal-directed actions, timing of repetitions, and story-like narratives – in infant formation of durable memories, including memories of words and referents (e.g., Kuhl, Tsao, & Liu, 2003; Vlach 2019; Sonne, Kingo, & Krøjgaard, 2018, Masek et al, 2021; Guillory & Kaldy, 2019; Thiele, Hepach, Michel, Gredebäck, & Haun, 2021). In brief, the statistics of parent-infant actions on objects may align with factors that support the efficient establishment of memories that bind the perceptual properties of objects to heard names from minimal repetitions.

The statistics of nonlinguistic coherence and discourse

To understand the statistics of the nonlinguistic behaviors that surround parent naming, we sought guidance from the coherence properties studied in discourse. We borrowed terms and measures liberally. In so doing, we are not suggesting an equivalence nor that parent-infant object-directed actions constitute true narratives. However, human language *is* a behavior and one that evolved in the service of other meaningful behaviors. In this way, the shared statistics may be fundamental to how infants the break into language.

We defined Topics and Comments solely by their frequency which differs from their more formal definition in terms of their informational functions. But the statistical properties of Topics, *above and beyond that expected by frequency*, implicate their role in organizing the whole play session and parent naming – recurrence throughout the session, continuity, temporal relatedness to all the toys in play. The statistical properties of the less frequent Comment objects—their continuity and temporal relations to the Topic objects, *above and beyond that expected by frequency,* also implicate an informational role for these objects. Although our use of the labels “Topic” and “Comment” is generous with respect to formal notions of these terms, their use in the current context may be developmentally meaningful. An expansive literature in cognitive science on the interleaving of information –first one kind of thing and then another – has shown that this interleaving leads to comparison of adjacent types and the discovery of similarities and differences (e.g., Carvalho & Goldstone, 2014; Carvalho, Chen, & Yu, C., 2021; Eimas, Quinn & Cowan, 1994; Goldenberg & Johnson, 2015; Namy & Gentner, 2002). By hypothesis, different Comment objects temporally adjacent to the Topic objects may lead to increased information about the perceptual properties of the Topic object and thus richer representations (see Carvalho & Goldstone, 2014; Goldenberg & Johnson, 2015).

This paper focused on objects and their names and the nonlinguistic experiences that may support early learning of object names. However, there is much more going on in these infant-parent interactions and much more being learned by the infants about language than just object names. The infant and parent activities involve inherently relational actions with goals and purposes. Parents talk is often about these actions: "he fell over", "hug, the baby", "we can put it in here". The extant evidence indicates that by the first birthday, infants are learning action terms (e.g., Bergelson & Swingley, 2013) and the meaning implications of the syntactic patterns in the speech they hear (e.g., Kedar, Casasola, Lust & Parmet, 2017). The coherence structure of parent and infant interactions around objects may be foundational to this learning and may connect the early learning of object names to the coherent relational structure of language more generally.

Conclusion

The observed coherence of the nonlinguistic events that surround parents' naming of objects suggests that the field needs to look beyond name-object co-occurrences and beyond heard language in seeking explanations for how infants break into word learning. Even if our current hypotheses about durable memory formation are not quite right, Figure 8 illustrates the theoretical point. Naming events for very young learners who do not talk emerge in a stream of sensory-motor-social activities. The internal activations that are our thoughts, expectations, and working memory representations at any moment are not discrete independent events. New information has its effects on mind and brain as perturbations of ongoing internal activity (Byrge, Sporns & Smith, 2014). This ongoing activity is determined by the coherency and statistics of experiences *in time*. More generally, all human behaviors – language, social interactions, actions on the physical world – occur in time and are characterized by statistical properties – skewed frequency distributions, continuity, and structured transitions – that likely have constrained the evolution of early learning mechanisms as well as the language we use to talk about the world and our behaviors in that world.

References

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

Admin, D., Gilmore, R.O., Seisler, A.R. & Soska, K. (2018). Databrary & Datavyu Help Videos. *Databrary*. Retrieved September 14, 2021 from http://doi.org/10.17910/B7.686.

Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2009). Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLOS one*, *4*(11), e7678.

Barabási, A. L. (2010). *Bursts: the hidden patterns behind everything we do, from your e-mail to bloody crusades*. Penguin.

Bateson, M. C. (1975). Mother-infant exchanges: the epigenesis of conversational interaction. *Annals of the New York Academy of Sciences*, *263*(1), 101-113.

Benitez, V. L., Zettersten, M., & Wojcik, E. (2020). The temporal structure of naming events differentially affects children's and adults' cross-situational word learning. *Journal of Experimental Child Psychology*, *200*, 104961.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253-3258.

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants.

Cognition, *127*, 391-397. *127*(3

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, *1*, 223-249.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics reports*, *424*(4-5), 175-308.

Bourgeois, K. S., Khawar, A. W., Neal, S. A., & Lockman, J. J. (2005). Infant manual exploration of objects, surfaces, and their interrelations. *Infancy*, *8*(3), 233-252.

Boyer, D., & Solis-Salas, C. (2014). Random walks with preferential relocations to places visited in the past and their application to biology. *Physical review letters*, *112*(24), 240601.

Byrge, L., Sporns, O., & Smith, L. B. (2014). Developmental process emerges from extended brain–body–behavior networks. *Trends in cognitive sciences*, *18*(8), 395-403.

Carvalho, P. F., Chen, C. H., & Yu, C. (2021). The distributional properties of exemplars affect category learning and generalization. *Scientific reports*, *11*(1), 1-10.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & cognition*, *42*(3), 481-495.

Chafe, W. (1987). Cognitive constraints on information flow. *Coherence and grounding in discourse*, *11*, 21-51.

Chen, C. H., Houston, D. M., & Yu, C. (2021). Parent–Child Joint Behaviors in Novel Object Play Create High-Quality Data for Word Learning. *Child Development*.

Clark, E. V. (2019). Research on First Language Acquisition: A brief history. In *International Handbook of Language Acquisition* (pp. 3-19). Routledge.

Clark, E. V. (2018). Conversation and language acquisition: A pragmatic approach. *Language*

Learning and Development, *14*(3), 170-185.

Clark, E. V., & Bernicot, J. (2008). Repetition as ratification: How parents and children place information in common ground. *Journal of child language*, *35*(2), 349-371.

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160055.

Dijk, T. A. (2010). *Studies in the Pragmatics of Discourse* (Vol. 101). Walter de Gruyter.

Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology*, *58*(3), 418-431.

Fisher, A. V. (2019). Selective sustained attention: a developmental foundation for cognition. *Current opinion in psychology*, *29*, 248-253.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101-107.

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 and Development*, *9*(1), 1-24.

Foltz, P. W. (2007). Discourse coherence and LSA. In *Handbook of latent semantic analysis* (pp. 179-196). Psychology Press.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215-239.

Gernsbacher, M. A. (2013). *Language comprehension as structure building*. Psychology Press.

Goldenberg, E. R., & Johnson, S. P. (2015). Category generalization in a new context: The role of visual attention. *Infant Behavior and Development*, *38*, 49-56.

Guillory, S. B., & Kaldy, Z. (2019). Persistence and Accumulation of Visual Memories for Objects in Scenes in 12-month-old Infants. *Frontiers in psychology*, *10*, 2454.

Halliday, M.A.K. & Hasan, R. (1976) *Cohesion in English.* English Language Series No. 9.

Hendrickson, A. T., & Perfors, A. (2019). Cross-situational learning in a Zipfian environment. *Cognition*, *189*, 11-22.

Hills, T. (2013). The company that words keep: comparing the statistical structure of child- versus adult-directed language. *Journal of child language*, *40*(3), 586-604.

Howard, L. H., & Woodward, A. L. (2019). Human actions support infant memory. *Journal of Cognition and Development*, *20*(5), 772-789.

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PloS one*, *10*(5), e0123780.

Johnson, D., & Brody, N. (1977). Visual habituation, sensorimotor development, and tempo of play in one-year-old infants. *Child Development*, 315-319.

Kartushina, N., & Mayor, J. (2019). Word knowledge in six-to nine-month-old Norwegian infants? Not without additional frequency cues. *Royal Society open science*, *6*(9), 180711.

Kedar, Y., Casasola, M., Lust, B., & Parmet, Y. (2017). Little words, big impact: Determiners begin to bootstrap reference by 12 months. *Language Learning and Development*, *13*(3), 317-334.

Keenan, J. M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of verbal learning and verbal behavior*, *16*(5), 549-560.

Krumme, C., Llorente, A., Cebrian, M., & Moro, E. (2013). The predictability of consumer visitation patterns. *Scientific reports*, *3*(1), 1-5.

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096-9101.

Kurumada, C., Meylan, S. C., & Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition*, *127*(3), 439-453.

Lavi-Rotbain, O., & Arnon, I. (2021). Visual statistical learning is facilitated in zipfian distributions. *Cognition*, *206*, 104492.

Li, L. J., Su, H., Lim, Y., & Fei-Fei, L. (2010, September). Objects as attributes for scene classification. In *European conference on computer vision* (pp. 57-69). Springer, Berlin, Heidelberg.

Liu, S., Zhao, X., Li, P., Love, B. C., McRae, K., & Sloutsky, V. M. (2008, July). Early lexical development: A corpus-based study of three languages. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1343-1348). Austin, TX: Cognitive Science Society.

Lockman, J. J., & Tamis-LeMonda, C. S. (2021). Young Children's Interactions with Objects: Play as Practice and Practice as Play. *Annual Review of Developmental Psychology*, *3*, 165-186.

Masek, L. R., McMillan, B. T., Paterson, S. J., Tamis-LeMonda, C. S., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Where language meets attention: How contingent interactions promote learning. *Developmental Review*, *60*, 100961.

Math Works, Inc.(2020). *MATLAB* (version R2020a).Natick, Massachusetts, United States.

Monroy, C., Chen, C. H., Houston, D., & Yu, C. (2021). Action prediction during real-time parent-infant interactions. *Developmental science*, *24*(3), e13042.

Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and diversity: Simulating early word learning environments. *Cognitive science*, *42*, 375-412.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: young children's use of comparison in category learning. *Journal of experimental psychology: General*, *131*(1), 5.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in cognitive sciences*, *11*(12), 520-527.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, *21*(5), 1112-1130.

Purver, M., Körding, K. P., Griffiths, T. L., & Tenenbaum, J. B. (2006, July). Unsupervised topic modelling for multi-party spoken discourse. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics* (pp. 17-24).

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*(3), 1059-1069.

Sandhofer, C. M., Smith, L. B., & Luo, J. (2000). Counting nouns and verbs in the input: Differential frequencies, different kinds of learning?. *Journal of child language*, *27*(3), 561-585.

Salakhutdinov, R., Torralba, A., & Tenenbaum, J. (2011, June). Learning to share visual appearance for multiclass object detection. In *CVPR 2011* (pp. 1481-1488). IEEE.

Schwab, J. F., & Lew-Williams, C. (2016). Repetition across successive sentences facilitates young children's word learning. *Developmental psychology*, *52*(6), 879.

Serrano, M. Á., Flammini, A., & Menczer, F. (2009). Modeling statistical properties of written text. *PloS one*, *4*(4), e5372.

Sonne, T., Kingo, O. S., & Krøjgaard, P. (2018). Meaningful memory? Eighteen-month-olds only remember cartoons with a meaningful storyline. *Frontiers in psychology*, *9*, 2388.

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*(3), 407-419.

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in cognitive sciences*, *18*(5), 251-258.

Steil, J. N., Friedrich, C., & Schild, U. (2021). Failure to show robust noun-referent associations in German-learning infants.

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.

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental psychology*, *55*(1), 96.

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child development*, *90*(6), 2135-2152.

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental science*, *20*(6), e12456.

Tannen, D. (2007). *Talking voices: Repetition, dialogue, and imagery in conversational discourse* (Vol. 26). Cambridge University Press.

Thiele, M., Hepach, R., Michel, C., Gredebäck, G., & Haun, D. B. (2021). Social interaction targets enhance 13-month-old infants' associative learning. *Infancy*, *26*(3), 409-422.

Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432-444.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172-175.

Tyler, A. (1994) The role of repetition in the perception of discourse coherence. *Journal of Pragmatics,* 21, 671-688.

Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology*, *50*, 73-79.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental science*, *16*(6), 959-966.

Vlach, H. A. (2019). Learning to remember words: Memory constraints as double-edged sword mechanisms of language development. *Child Development Perspectives*, *13*(3), 159-165.

Wojcik, E. H. (2013). Remembering new words: Integrating early memory development into word learning. *Frontiers in psychology*, *4*, 151.

Zipf, G.K. (1949) *Human behavior and the principle of least effort.* New York: Addison-Wesley.

Figure Captions

Figure 1. Infant wearing head camera and infant perspective images captured by the head camera.

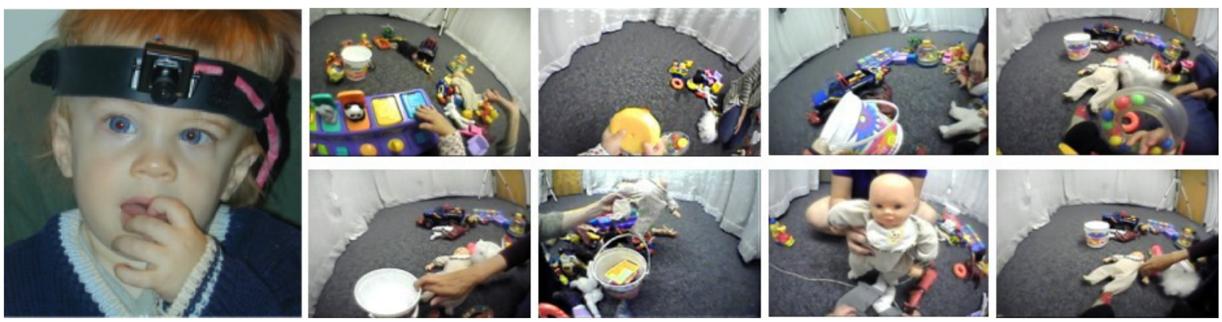


Figure 2. Mean proportion and standard deviations of handled and named objects by their rank order frequency. A. Infant handling. B. Parent handling. C. Dyad Handling. D. Parent naming. Maroon indicates the 5 most frequent handled objects and orange indicates the remaining infrequently handled objects. Dark blue indicates the 5 most frequently named objects, the naming, and green indicates the less frequently named or objects. The dashed lines for handling indicate the expected frequency of handling if actions were equally distributed across all objects. The frequency of Dyad Handling was used to define two categories of objects: Higher Frequency Topics (maroon) and Lower Frequency Comments (orange). See text for clarification.

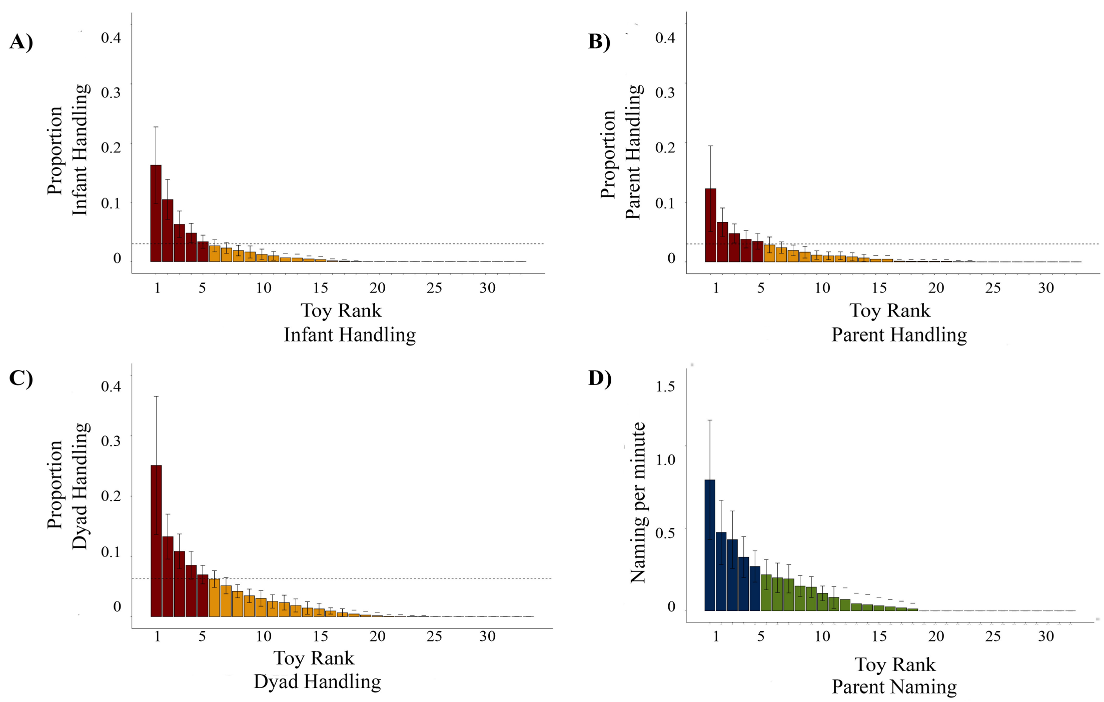


Figure 3. Frequency of dyad handling (A) and parent naming (B) as a function of the four (2.5 minute) quarters of the interaction. Values for Topic objects are shown in maroon for dyad handling, and blue for parent naming and values for Comment objects are shown in orange for dyad handling and green for parent naming.

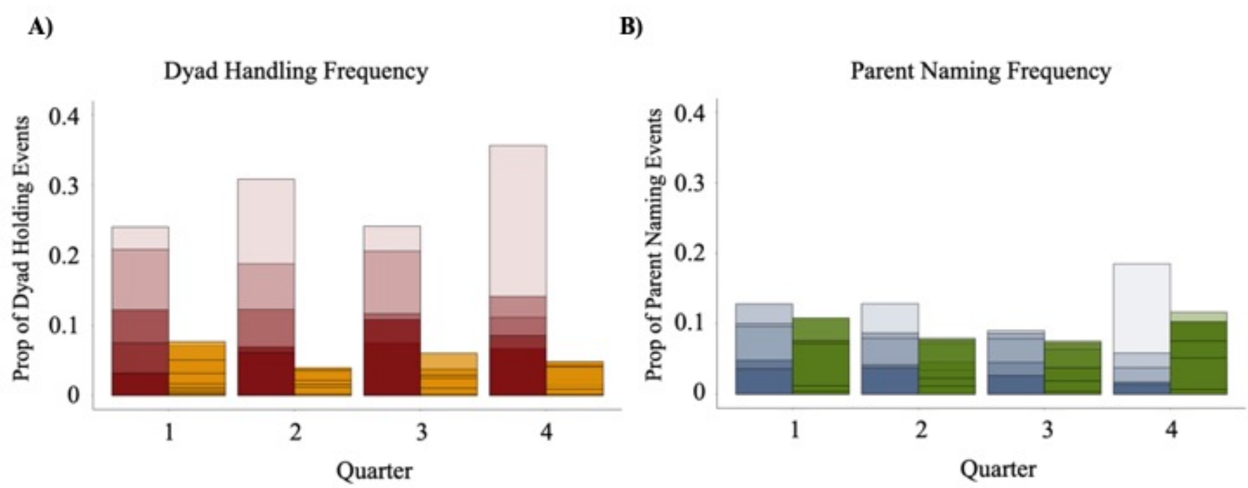


Figure 4. Probability of continuity (see text) for infant, parent, and dyad handling of Topic (maroon) and Comment (orange) objects and for parent naming of Topic (blue) and Comment (green) objects. The bar graphs show continuity computed at the corpus level; the dots show continuity computed for each dyad separately.

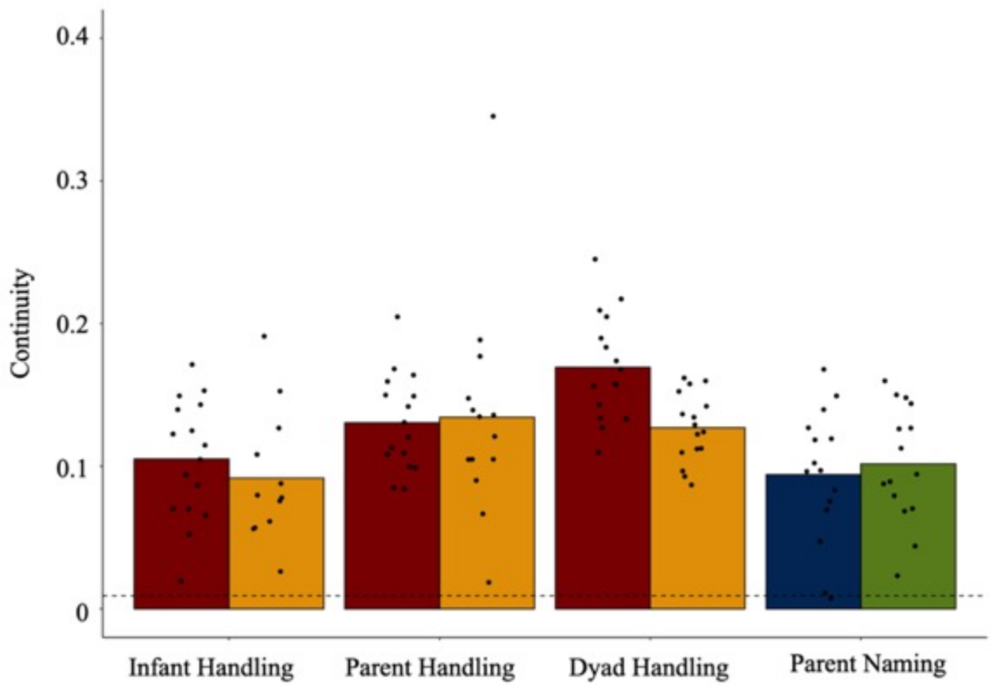


Figure 5. Five kinds of transitions: from handling to no handling (*-), no handling to no handling (**), no handling to handling an object (*-), handling the same object (SO), and handling a new object (NO). The transitions are coded at the dyad level without regard to whose hands. Therefore, in the transition from parent handling the giraffe to parent handling the cow and infant handling the hammer, there are two transitions, both NO, giraffe to cow, and giraffe to hammer. The bar graph shows the proportions of observed transitions (top) for the corpus and the expected transitions (from the randomly generated permutations, see text).

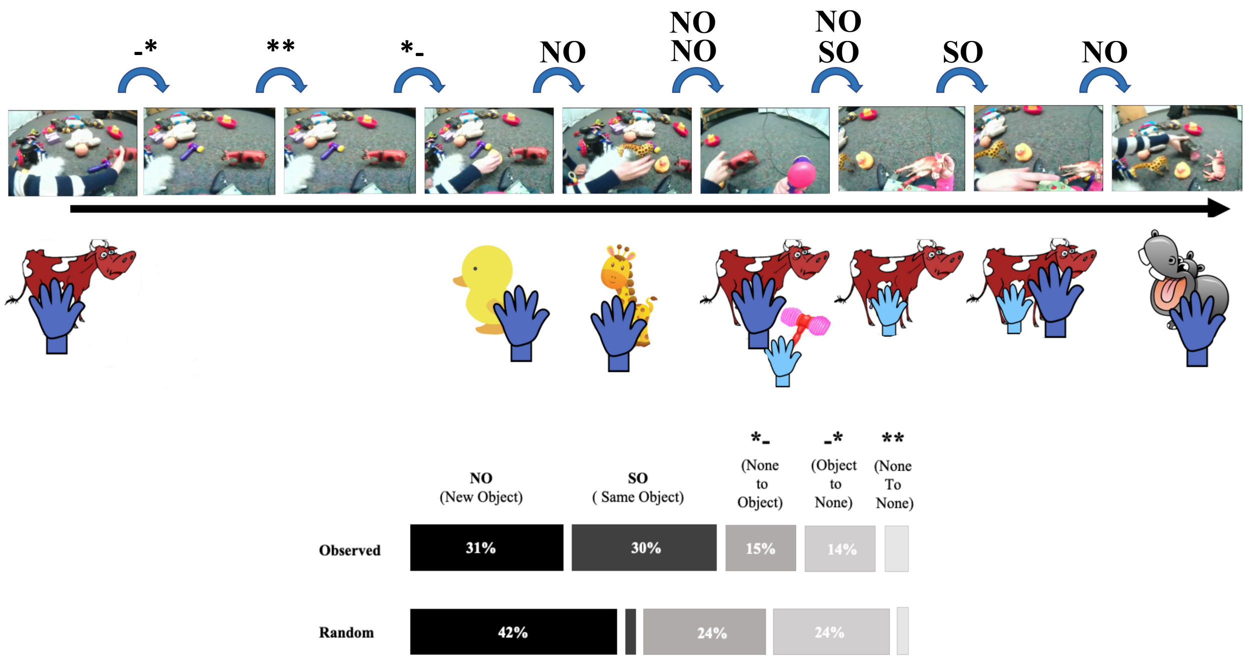


Figure 6. Nondirected corpus-level networks of NO transitions for the Observed (A) and Random (B) networks. The nodes represent individual objects at each rank of dyad handing and the edges indicate successive transitions from one object to another. The darkness of the edges indicates the frequency weight of the edge. Topic objects are indicated in maroon and comment objects in orange. The measures of degrees, closeness, and betweenness (see text for definition) for each node in the network are shown below each network. Topic objects are indicated in maroon and Comment objects in orange.

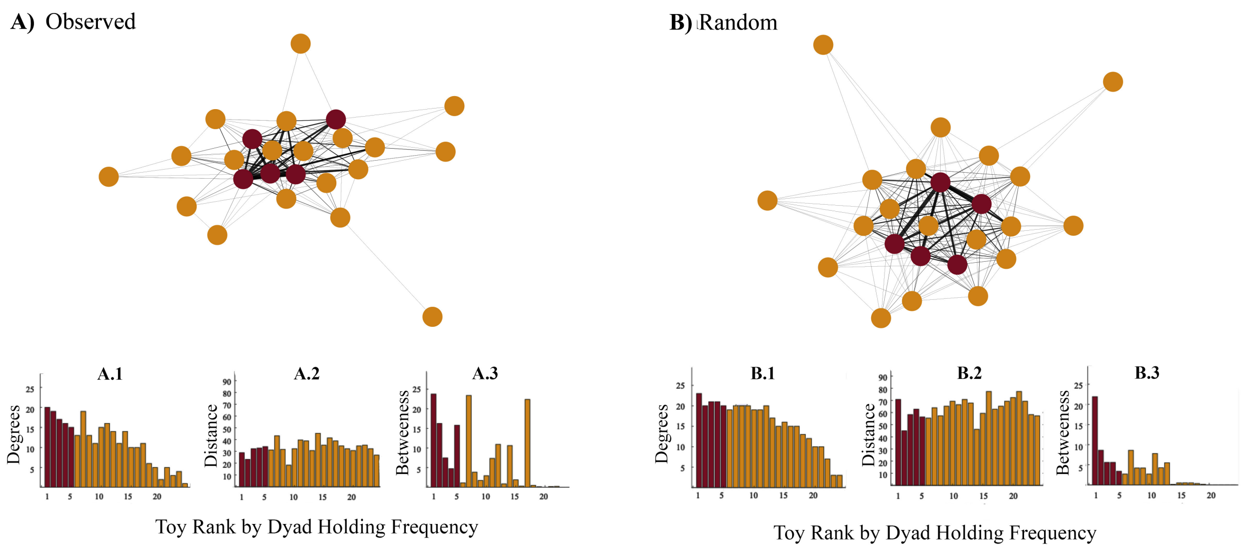


Figure 7. Networks of the NO transitions for two individual dyads (differential edge weights are not shown for ease of seeing the structure). Topic objects are maroon and Comment objects are orange. Nodes are specific objects indicated by name.

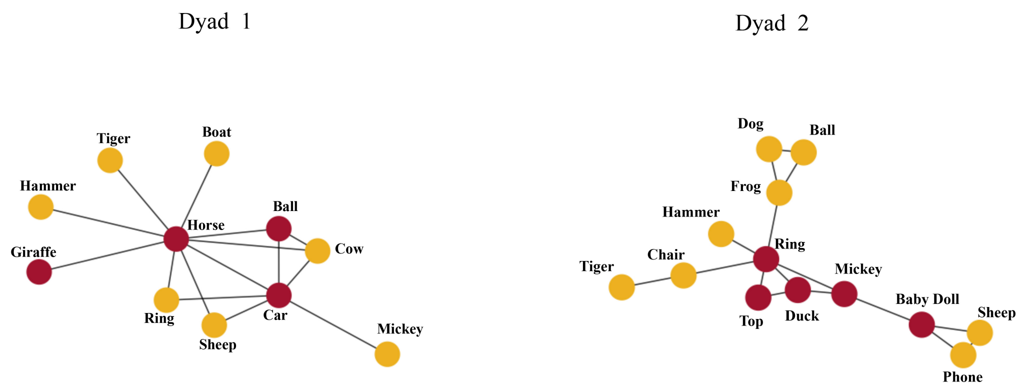
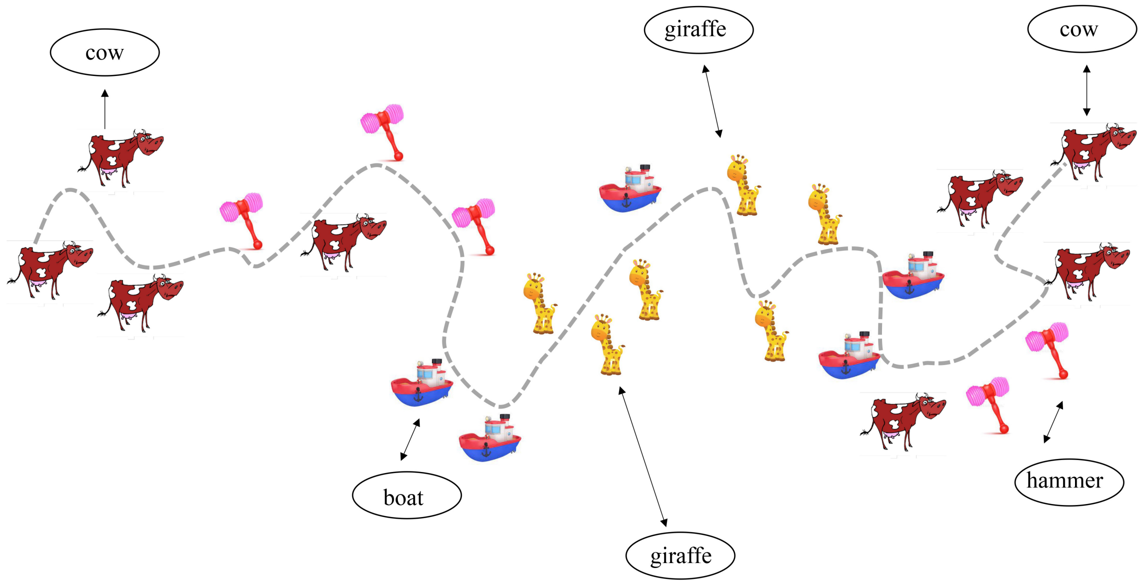


Figure 8. Illustration of how relatively infrequent name-object co-occurrences are surrounded by the coherent structure of object handling.



Appendix