Title
Distributional Learning in Context: How Social Embedding Structures Infant-Directed Speech

Author
Chang, Lucas Moraes

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Distributional Learning in Context: How Social Embedding Structures Infant-Directed Speech

A dissertation submitted in partial satisfaction
of the requirements for the degree Doctor of Philosophy

in

Cognitive Science

by

Lucas Moraes Chang

Committee in charge:

Professor Gedeon Deák, Chair
Professor Andrew Kehler
Professor Rachel Mayberry
Professor Eran Mukamel
Professor Federico Rossano

2020

The dissertation of Lucas Moraes Chang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

DEDICATION

This dissertation is dedicated to all children between the ages of 4 and 30 months, regardless of whether they are learning American English or have been diagnosed with a developmental disorder.

EPIGRAPH

Lo cierto es que vivimos postergando todo lo postergable; tal vez todos sabemos profundamente que somos inmortales y que tarde o temprano, todo hombre hará todas las cosas y sabrá todo.

The truth is that we live out our lives putting off all that can be put off; perhaps we all know deep down that we are immortal and that sooner or later all men will do and know all things.

Jorge Luis Borges. Funes el Memorioso, 1942.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Gedeon Deák, for his supportive mentorship. He gave me the independence to explore novel ideas and methods, but also paid attention to detail when needed. I would also like to thank Federico Rossano, Eran Mukamel, Rachel Mayberry and Andy Kehler for their helpful suggestions and for serving on my dissertation committee. The late Jeff Elman also provided valuable feedback during his time on my committee.

I would like to thank Kaya de Barbaro, whose research in the lab laid the foundations for my projects, and she provided valuable advice for navigating grad school.

Sean Trott and Arthur Semenuks provided valuable comments on draft manuscripts. Mike Goldstein and Jen Schwade introduced me to research on language development and invested a remarkable amount of effort in mentoring me during my time as an undergraduate research assistant.

Generous funding for my graduate education was provided by the Katzin Prize and by the National Science Foundation Graduate Research Fellowship program. Funding for my research was provided by the Temporal Dynamics of Learning Center.

The observational studies in Chapters 1 and 2 required a massive team effort. Vivian Zhang, Maria Kremliovsky, Kimberli Ng, Christen Buetz, Erin Maxwell, Sofia Weiss, Alissa Gutierrez, Esther Yoon, Scott Ragland, Yunjia Yao, Mandy Natsuhara, and Shirlene Wade assisted with video coding and transcription. Jordan Danly oversaw the home observational sessions.

Finally, I would like to thank Helen Wang for all her love and support.

Chapter 1, in full, is a reprint of the material as it appears in: Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology.* The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in: Chang, L. M., & Deák, G. O. (2018). Maternal discourse continuity and infants' actions organize 12-month-olds' language exposure during object play. *Developmental science,* e12770. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication of the material as it may appear in: Chang, L. M., & Deák, G. O. (2020). Adjacent and Non-Adjacent Word Contexts Both Predict Age of Acquisition of English Words: A Distributional Corpus Analysis of Child-directed Speech. Cognitive Science. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is currently being prepared for submission for publication of the material, and is co-authored with Gedeon Deák. The dissertation author was the primary investigator and author of this paper.

VITA

2013 Bachelor of Arts, Cornell University

2016 Master of Science, University of California San Diego

2018 Data Scientist, Accenture

2020 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Chang, L. M., & Deák, G. O. (2018). Maternal discourse continuity and infants' actions organize 12-month-olds' language exposure during object play. Developmental science, e12770.

Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology.*

Luna, A., Karac., E. I, Sunshine, M, Chang, L, Nussinov, R., Aladjem, M. I., and Kohn, K. W. (2011). A Formal MIM Specification and Tools for the Common Exchange of MIM Diagrams: an XML-Based Format, an API, and a Validation Method. *BMC Bioinformatics,* 12:167

ABSTRACT OF THE DISSERTATION

Distributional Learning in Context: How Social Embedding Structures Infant-Directed Speech

by

Lucas Moraes Chang

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2020

Professor Gedeon Deák, Chair

Infants and toddlers typically hear words accompanied by a variety of direct and indirect cues to their meaning. To name just a few, words are embedded in frequently repeated linguistic constructions, they tend to co-occur with specific objects that they refer to, and they tend to be used in different social-interaction routines and activity contexts. Whereas children are capable of detecting several different types of cues and using them to facilitate word learning, we are only beginning to uncover the developmental processes by which words come to be embedded in multimodal, dynamic contexts that mark them as items to be learned and help children to discover their meaning.

In this dissertation I address two broad questions. First, how do infants and caregivers co-construct interaction sequences in which words are accompanied with useful cues? In a series of observational studies of infant-mother dyads observed longitudinally from the age of 4 months to 12 months, I describe how infants' increasing motor abilities enable them to elicit

contingent caregiver responses containing object-naming and predictable sequences of other informative utterances.

Second, what can we learn about what matters for word learning by using the contextual distributions of words to predict how early and in what combinations they will be learned? In these studies, I use a corpus of child-directed speech to construct a representation of each word's distribution over syntactic and thematic usage contexts. Then, using a large open dataset of children's parent-reported word production and comprehension, I show that both types of context distribution contribute over and above previously described factors in predicting both the age of acquisition of words and the degree to which word pairs tended to be learned together versus randomly.

Taken together, the studies in this dissertation support a view of early word learning in which (1) multiple layers of social, linguistic, and sensorimotor contextual cues jointly facilitate word learning, (2) infants learn to participate actively in the responsive interactions that produce high-quality word exposures, and (3) although these processes are too complex to be replicated with full experimental control, they leave identifiable traces in the structure of children's lexicons.

INTRODUCTION

Language acquisition takes place in an environment of remarkable richness and complexity. Of particular interest are the ways in which the child's environment supports the early stages of word learning. Researchers have long noted that the task of mapping word forms with meanings presents numerous difficulties for learners. The sets of candidate word forms and meanings are unknown, word-referent associations are noisy and ambiguous, and word meanings can be multiple, complex, and overlapping. Nevertheless, much of the existing research on word learning in infants and toddlers treats word-meaning (often, word-object) associations as the "atom" of the lexicon. In light of the phenomenon of rapid vocabulary acquisition (Benedict, 1979; Goldfield & Reznick, 1990), several theories have been proposed to explain how some particular factor or set of factors constrain and support word learning. Three main trends can be identified:

Social theories, based on observational studies and individual differences research, treat the environment as a set of social experiences in which infants' active participation and parents' responsiveness are central. Studies of this type have identified patterns of infant-parent interaction that are associated with individual differences in language development, but do not provide explicit models of how these patterns of interaction modulate the learning process itself.

Information-oriented theories, based on corpus analyses and statistical learning studies, treat the environment as an input stream of linguistic information, and emphasize the learnability of linguistic patterns. These studies have demonstrated the feasibility of acquiring linguistic structures by various algorithms, but rarely address the social coordination required to provide the learning algorithms with the necessary inputs.

Knowledge-oriented theories, based on studies of the implicit biases in children's word learning process, as well as studies relating the set of words in a child's existing vocabulary to further vocabulary growth, emphasize the structured knowledge available at a given point in developmental time as a key factor in further word learning. Interdependencies among words and generalized strategies for learning words in a particular class provide evidence that word learning is best understood in the context of an existing lexicon.

Each of the three approaches has provided robust findings that give valuable insights into the mechanisms of language acquisition. However, existing studies typically demonstrate how some particular factor facilitates word learning above mere word-referent association. Instead of viewing social, informational, and knowledge theories as competing conceptualizations of the role of environmental support for word learning, the growing body of evidence for each suggests that a full understanding of word learning requires studying the mutual effects of all three sets of factors. Word learning events occur in a context involving social support, probabilistic information, and past experiences with language, and the role and effectiveness of each set of factors likely depends on the others. Each instance of word usage is multiply embedded across perceptual, cognitive, and behavioral modalities.

The studies described in this dissertation take a step toward integrating our understanding of the multiple interacting environmental supports for word learning. The first series of studies integrates social and information-oriented factors. Using observational microbehavioral analysis of infant-parent interaction, we describe how parental contingent responses modulate cross-modal correlations between parental speech content and infants' visual and haptic experience at different ages. We then expand this analysis to slightly older infants and describe how different types of parental speech acts are differentially structured by parental contingent responsiveness and by predictable multi-utterance sequences. The second series of studies integrates information- and knowledge-oriented factors. Using large-scale corpora of transcribed infant-directed speech, we derive distributional representations of the embedding of words into local linguistic contexts. Then, using a database of children's vocabularies, we characterize how the probability of learning each word is determined both by its usage and its relation to other known words in the child's lexicon.

In the remainder of this introduction, I will first review previous research on social, information-oriented, and lexicon-oriented theories of language acquisition. I will then outline several key open questions regarding how multiple factors identified by different theoretical approaches to word learning relate to each other, and how they are addressed by the studies in this dissertation.

Social theories: Parent-infant interaction, responsiveness, and shared attention

Studies of the influence of the social environment on language acquisition are rooted in the bioecological model of development (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 2006). This framework posits that development takes place within nested environmental systems. The most distal systems include broad cultural contexts and values. These shape the specific environments and situations infants experience, which in turn determine their language development. Most proximal to the mechanisms of language acquisition are social environmental factors relating to parent-infant interaction--when, how, and how much language is used with infants (Hart & Risley, 1995; Hoff, 2003). A large body of research indicates that parents respond contingently and appropriately to infants' communicative actions and focus of attention. Parental responsiveness to infant actions, both communicative and non-communicative, begins at an early age: for example, infants and parents mutually coordinate their looking, smiling, and vocalizing by 1 to 3 months (Kärtner et al., 2008; Kaye & Fogel, 1980). Mothers respond contingently to infants' object exploration (Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008, Wu & Gros-Louis, 2014, Yu & Smith, 2012, Tamis-Lemonda, Kuchirko, & Tafuro, 2013) and gesture (Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007), and take turns vocalizing (Snow, 1977). Moreover, this social embedding of language appears to "gate" acquisition (Kuhl, 2007). Children exposed to a second language through television do not acquire that language (Snow et al., 1976), and neither do children in the rare circumstance of exposure to a first language only by television (Sachs, Bard, & Johnson, 1981). In a more controlled setting, English-learning infants exposed to Mandarin via live storytelling sessions retained the ability to discriminate Mandarin phonetic distinctions, whereas those exposed via pre-recorded sessions did not (Kuhl, Tsao, & Liu, 2003). Thus, it seems that responsive interaction is necessary for infants to learn from at least some kinds of linguistic input.

However, comparisons of live versus televised language models conflate a number of differences, making causal inferences difficult. Another approach is to test how infant-parent interaction styles that vary in their degree of responsiveness relate to infants' rates of language acquisition. In several studies, parents' contingent and appropriate responses to infants' actions have been counted and summarized as a quantitative measure of "responsiveness." Although precise definitions vary from study to study, parent responsiveness by various metrics is typically a strong positive predictor of children's language development (Tamis-LeMonda, Kuchirko, & Song, 2014, Tamis-LeMonda, Bornstein, and Baumwell, 2001, McGillion et al., 2013, Gros-Louis, West, & King, 2014, Wu & Gros-Louis, 2014).

Furthermore, several related lines of evidence suggest that infants are sensitive to temporal contingency. At two months of age, infants notice when a mobile moves in response to their own head movements, and begin to direct social actions toward the mobile (Watson, 1972). Ten-month-olds readily detect the contingent responses of a robot and use this information to determine its orientation of "attention" based on the direction it faces when contingently responding to the infant (Movellan & Watson, 1987). In the linguistic domain, Goldstein and Schwade (2008) found that adults' contingent, but not random, responses to infants' vocalizations, caused infants' vocalizations to become more advanced and incorporate phonological elements from the adult's responses. Twelve-month-olds' object name learning was also facilitated by presenting object names contingently on infants' vocalizing while looking at the object (Goldstein, Schwade, Briesch, & Syal, 2010).

A related hypothesis about the nature of "responsiveness" is that, rather than temporal contingency between infant actions and parent responses, reponsiveness primarily reflects coordination of joint attention between the infant and parent. Tomasello & Farrar (1986) studied mothers' speech to 15- and 21-month-old children during episodes of joint attention. They found that mothers' references to objects that were the children's target of attention before the utterance positively predicted vocabulary scores. In the same paper, an experiment showed that 17-month-olds were more likely to learn words presented in response to their current focus of attention than words presented in an attempt to redirect their attention. As early as 9 months, caregivers' contingent comments about objects that were the focus of the child's attention predict infants' subsequent language comprehension (Rollins, 2003). Similarly, at 13 months, mothers' directive language following, but not leading, the child's focus of attention correlated positively with later vocabulary (Akhtar, Dunham, & Dunham, 1991). Moreover, infants' ability to follow an experimenter's gaze or gestures to establish joint attention predicts their later receptive vocabulary (Mundy & Gomes, 1998).

Recently, a growing number of researchers who focus on social theories have recognized that children's experience with language arises from mutual interaction between the child and the physical and social world (Hockema & Smith, 2009). The primary linguistic data are not merely an input to the developmental system; rather, children's exposure to language is both a cause and an effect in the multi-person developmental system. Thus, measures of social interaction, such as responsiveness, simultaneously capture variance in parents' tendency to respond to children's behaviors, and children's production of behaviors that readily elicit contingent responses. This latter, endogenous variance can have clinically relevant consequences: for example, infants at high risk for autism spectrum disorder (ASD) produce fewer speech-like vocalizations, fewer consonant types, and fewer canonical syllables than low-risk controls, and these differences predict the later appearance of ASD symptoms (Paul, Fuerst, Ramsay, Chawarska, & Klin, 2011). Thus, deficits in language and attention to social contingencies in ASD might be partially caused by affected infants' initially impoverished vocal repertoire, which in turn gives them fewer opportunities both to learn from contingent responses and to experience their reward value. In addition to differences between individuals and over time in vocal behavior, gross motor skills are related to language development (Oudgenoeg-Paz, Volman, & Leseman, 2012; Walle and Campos, 2014; He, Walle, & Campos, 2015; Libertus and Violi, 2016). Motor behaviors shape children's language environment (Iverson, 2010). At 13 months, infants who carry an object to their mother by walking elicit more action directives than do infants who do so by crawling (Karasik, Tamis-LeMonda, & Adolph, 2014), and 8-month-olds who have begun crawling hear more prohibitions than non-locomoting infants (Zumbahlen, 1997). Furthermore, infants who have recently begun walking show changes in the types of social interactions they engage in (Clearfield, 2011; Karasik, Tamis-LeMonda & Adolph, 2011). Onset of walking is also associated with increases in infant following of parents' joint engagement cues, parents' perception of the infant as an individual, and receptive and productive language (Walle, 2016). Thus, the characteristics of the child's learning environment such as quality of speech input and responsiveness cannot be considered independently of individual differences and developmental changes in the child's motor repertoire and interactions between the child, parent, and physical environment.

Overall, researchers have identified a broad range of factors that account for individual differences in word learning. The common thread is that infants benefit from linguistic input that is embedded in rich, responsive and meaningful exchanges with the social and physical world. However, the pathways by which social environmental factors impact and influence children's word learning at the cognitive level are poorly understood. The next two sections review research on environmental influences on word learning from perspectives with more cognitive detail.

Information-oriented theories: Corpus analysis and statistical learning

In contrast to the social influences reviewed above, information-oriented theories of language acquisition focus on specific learnable regularities in the linguistic environment. These theories emphasize that language acquisition is fundamentally a learning problem; the main questions therefore revolve around what information is available in the environment to be learned, and what types of information (e.g. statistical structure in a sound sequence) children are able to detect. Research on the statistical patterns in the environment typically involves corpus-level analyses of natural infant-directed speech; by contrast, research on children's capacity to learn those patterns typically involves experiments on children's statistical learning, conducted on small-scale "toy" tasks in the laboratory.

Any mechanistic theory of word learning, and language acquisition more generally, must consider the nature of the input: the range of sets of experiences from which children will eventually learn their native language. If children have powerful general-purpose statistical learning abilities that are tuned–through early social and perceptual learning, or intrinsic biases, or both–to process incoming linguistic information, it is essential to know what types of patterns could hypothetically be learned. This has been addressed by analyzing corpora of child-directed speech, transcribed at various levels of granularity, to extract statistical regularities.

One of the most basic tasks faced by a language learner is to identify words. Is there sufficient information in fluent speech to identify word boundaries, which are typically not marked with audible pauses? To test this, a connectionist model was trained to discover word boundaries based on a corpus of phonetically transcribed child-directed speech, using the phoneme sequence, stress patterns, and utterance boundaries as cues. The model segmented word boundaries with over 70% accuracy and completeness, far from perfect but nevertheless showing that a meaningful segmentation of speech into words can be computed from bottom-up processing of the statistical structure in natural child-directed speech (Christiansen, Allen, & Seidenberg, 1998). A more recent word segmentation model shows that a model that assumes that words predict subsequent words learns a more complete segmentation than one that assumes that words are independent (Goldwater, Griffiths, & Johnson, 2009). This last finding illustrates a more general point: additional knowledge or assumptions about linguistic structure might enable the extraction of more patterns than are learnable by the most basic models.

Once words are identified, learners must discover that words have categorical structure: some words are nouns, some are verbs, and some are closed-class function words. Several studies have examined the degree to which a word's phonological properties predict its lexical class; e.g. in English, nouns tend to have stress on the first syllable, whereas verbs tend to have stress on the second (Kelly & Bock, 1988). Shi, Morgan, and Allopenna (1998) classified Turkish and Mandarin words as closed-class or open-class based on a combination of phonological features.

Another source of information about word class comes from distributional properties of words--that is, which other words a given word tends to co-occur with. Distributional information in the form of counts of a word's co-occurrence with other common words provide a reasonably strong cue to word class, especially for nouns (Redington, Chater, & Finch, 1998). Combining phonological and distributional cues leads to even higher performance, as the two sets of cues provide complementary information: distributional cues are more useful for frequent words and for identifying nouns, whereas phonological cues are more predictive of the class of infrequent words and for identifying verbs (Monaghan, Chater, & Christiansen, 2005; Monaghan, Christiansen, & Chater, 2007). Distributional cues also provide fine-grained information within a word class, e.g., the types of relations that can be expressed by a verb can be determined from the context words it occurs with (Hare, McRae, & Elman, 2004). Notably, it may be possible in some cases to learn word categories from distributional information even with a small sample of child-directed speech. Mintz (2003) performed an analysis of child-directed speech in terms of frequent frames: patterns of the form *you X the*, in which a variable word is flanked by a specific pair of words. Frequent frames generally consist of function words, and words that appear within them generally belong to the same class. Using corpora derived from single parent-child dyads– contrasting with studies using a million or more words pooled over many recordings –Mintz showed successful identification of nouns and verbs. Furthermore, frequent frames generalize to French and outperform similar analyses using contextual elements consisting of two words preceding or following the target word (Chemla, Mintz, Bernal, & Christophe, 2009). Cameron-Faulkner, Lieven & Tomasello (2003) found that 51% of utterances directed to 2-year-old children started with one of 52 common phrases), suggesting that a relatively manageable number of specific item-based constructions constitute a large fraction of child-directed speech. Thus, early language acquisition could uncover abstract structure despite operating on relatively small samples of specific elements.

Modeling the learnability of linguistic structures from corpus data has also yielded demonstrations of the learnability of specific syntactic structures, including some previously used as justification for the Poverty of the Stimulus argument (Reali & Christiansen, 2005), although the learning model was not completely unconstrained. Nevertheless, the level of constraint imposed by these and similar models is consistent with the model of language-brain coevolution proposed by Christiansen and Chater (2008). According to their language evolution hypothesis, languages evolve over generations of transmission to adapt themselves to the peculiarities of the brains they inhabit. Computer simulations of the coevolution of languages and brains suggest that the higher rate of change of language structure dominates the relatively slow rate of change of brain structure. The result of this diachronic process of language shaping is that languages possess apparently arbitrary structure that is well-matched to the domain-general cognitive abilities human brains possess. Thus, researchers who select models whose constraints ensure good performance may be implementing something analogous to the mechanisms by which human children come to have an apparent predisposition to detect and correctly learn linguistic patterns.

A second thread of research has used small-scale experimental studies to demonstrate that infants have the ability to detect a wide range of probabilistic regularities relevant to language. These processes are referred to collectively as statistical learning. At the phonetic level, prelinguistic infants can learn boundaries between two phonemic categories from the distribution of a linguistic feature: bimodal distributions suggest the presence of two categories to be distinguished, whereas unimodal distributions suggest a single category, the members of which should be treated alike (Maye, Werker, & Gerken, 2002). Another influential finding shows that 9-month-old infants can track conditional transition probabilities between syllables (Saffran, Aslin, & Newport, 1996, Aslin, Saffran, & Newport, 1998). Because the conditional probability of a syllable given the previous syllable is usually high within words and low across word boundaries, this type of learning is a potential mechanism by which children could learn to parse a continuous stream of sound as a sequence of words.

In addition to the low-level elements described above, infants also use statistical or distributional information to acquire higher-order elements of language structure, i.e. "grammar." In one study, 12-month-olds were exposed to an artificial language consisting of strings conforming to the patterns aX and bY, where X and Y elements were distinguished both by phonetic features and by their respective co-occurrence with a and b elements. At test, infants heard strings that either conformed to the familiar pattern or the opposite pairing. Infants discriminated familiar from unfamiliar patterns even when the X and Y elements were novel at test (Gómez & Lakusta, 2004). Using patterns from natural language as stimuli, Gerken, Wilson, and Lewis (2005) found that 17-month-olds were capable of distinguishing grammatical from ungrammatical combinations of Russian noun stems and gender endings after brief exposure to a set of Russian words in which gender was correlated with phonetic properties of the stem.

Statistical learning has also been studied in the context of learning word meanings. The problem of assigning words to their correct referents is in principle difficult and highly indeterminate (Quine, 1960, Trueswell et al., 2016). Nevertheless, children do succeed in acquiring specific vocabulary items. Word learning may rely at least in part on children's integration of word-world associations that are individually ambiguous and probabilistic. To demonstrate this, Smith & Yu (2008) exposed 12- and 14-month-old infants to training trials in which they viewed pairs of objects and heard pairs of words, such that the specific word-object pairings were ambiguous on each trial. Nevertheless, infants were capable of learning the word-object pairings, as evidenced by above-chance performance on a looking-time task.

It should be noted that these and similar studies exposed infants to the statistical regularities for brief periods ranging from a few minutes to tens of minutes. In the real world, accumulated weeks and months of experience could, in principle, provide a much stronger signal, although infants may be less sensitive to dispersed everyday experience with statistical regularities compared to the concentrated blocks of input they receive in typical statistical learning studies.

Regarding whether infant statistical learning is a domain-general or a language-specific mechanism, some studies have been conceptually replicated using non-language-like stimuli, e.g. pictures of dogs (Saffran, Pollak, Seibel, & Shkolnik, 2007) and shapes (Kirkham, Slemmer, & Johnson, 2002), although performance on non-language stimuli can be facilitated by prior learning of the same patterns instantiated in speech (Marcus, Fernandes, & Johnson, 2007). Similarly, 4- and 5-year-olds are able to form associations between objects and facts or pictograms as well as or better than words (Deák & Toney, 2013). Statistical learning has also been used to argue against the modularity of the language faculty: individuals with "specific language impairment" perform worse than controls on both linguistic and non-linguistic tests of sequential-statistical learning (Evans, Saffran & Robe-Torres, 2009, Tomblin, Mainela-Arnold & Zhang, 2007), and language ability is also correlated with statistical learning performance in normal undergraduates (Misyak & Christiansen, 2012).

Overall, information-oriented theories of language acquisition have been useful in that they have made it clear that an abundance of structured information is available in children's experience, and that children have mechanisms for detecting at least some of that structure. Nevertheless, these theories provide an idealized and incomplete picture of the real processes of language acquisition, which take place neither in controlled lab contexts nor solely by means of decontextualized transcribed corpora. It is also unclear which of the many available patterns in the input are causal drivers of acquisition. Furthermore, statistical learning is a cumulative process where the knowledge accumulated as a result of previous learning events forms the basis of future ones. Past experience can influence the learning process by helping children to focus on relevant aspects of the input, and by providing conceptual structures to incorporate newly learned items. Explaining these processes requires viewing language acquisition as a prolonged developmental process in which children progressively acquire a lexicon and grammar, and world knowledge, with the different developing parts inextricably linked. This is the subject of the next section.

Knowledge-oriented theories: biases, lexicons, and inter-word dependencies

Children are not simply general-purpose learning machines who encounter structured information in the language input. Rather, their inferences about word meaning are enabled by knowledge about how words tend to be used in general, and by generalizing from already-known words. Intuitive support for this view comes from the fact that word learning is initially slow and laborious: although infants' recognition of common words is first detectable around 6 to 9 months (Bergelson & Swingley, 2012; 2015; Tincoff & Jusczyk, 2012), their vocabulary growth is initially slow until the "vocabulary spurt," a dramatic acceleration of word learning that typically occurs in the second year (Bloom, 1973; Ganger & Brent, 2004; Bloom, 2000).

Researchers have proposed several conceptual or representational advances that might account for the vocabulary spurt. For example, the "naming insight," where infants discover that word-forms are names for object types, has been suggested as an explanation (Dore, Franklin, Miller & Ramer, 1976; Reznick & Goldfield, 1992; Kamhi, 1986). A related hypothesis is that infants shift from an associationist strategy (matching words and meanings) to a referential strategy (inferring what people intend to communicate) around the time of the vocabulary spurt (Nazzi & Bertoncini, 2003). Armed with this basic insight, it becomes possible for toddlers to learn an array of further principles that guide word learning. Toddlers have been found to disambiguate words on several such grounds. For instance, the taxonomic constraint implies that words apply to objects of the same *type*, rather than those that are thematically related – even though thematic relations are highly salient to children (Markman & Hutchinson, 1984). Children also assume that a novel label is more likely to apply to a category without a known name than one with a known name (Markman & Wachtel, 1988; but see Deák & Maratsos, 1998). One study suggests that this last effect depends on the degree to which the child's existing lexicon approximates a one-to-one correspondence between words and meanings, as it was observed in 18-month-old Chinese-English bilingual children who knew few translation equivalents, but not those who knew many translation equivalents (Byers-Heinlein & Werker, 2013).

Moving up the conceptual hierarchy, children use higher-level knowledge about relations among words in an utterance to disambiguate meaning. For instance, children assume that a count noun should refer to a class of objects defined by a common shape, whereas a mass noun should be a category defined by a common substance (McPherson, 1991; Smith, 1999). According to the *syntactic bootstrapping* hypothesis, children infer the meanings of novel verbs by assigning meanings that fit the set of grammatical roles in the verb phrase (Gleitman, 1990). Naigles (1990) obtained experimental evidence for this effect by presenting 2-year-old children with either transitive (e.g. *the duck is gorping the bunny*) or intransitive (*the duck and the bunny are gorping*) sentences accompanied by videos demonstrating both transitive and intransitive actions. In a subsequent test phase, the children were prompted to look at "gorping" and were shown separate videos of the two actions, and looked longer at the action consistent with the phrase structure they heard during training, even though both sentences contained the same set of focal nouns and verbs. Another example is the distinction between count nouns and proper names. In one paradigm, children are shown a novel doll that is introduced either as a count noun (e.g. *this is a Zav*) or as a proper noun (e.g. *this is Zav).* The children are then shown two identical dolls, one of which is the same exemplar from the labeling phase, and are asked to either find "a Zav" or "Zav," respectively. 24-month old children, but not 20-month-olds, preferred to select the same individual in the proper name condition (Hall, Lee, & Bélanger, 2001). Crucially, in all these situations the cues to meaning are not inherent to the naming event, but rather require children to have preexisting knowledge of linguistic usage. Furthermore, two-year-old children's inferences can also operate across utterance boundaries: children prompted by an utterance such as *I'm thirsty* subsequently were more likely to assign a novel word to the drinkable item in an array (Sullivan & Barner, 2016).

Nativistic theories represent a different perspective that explains word learning as a result of structured knowledge. For instance, Pinker (1994a) proposes that children's learning of word meanings is facilitated by constraints on the range of possible meanings that can be represented in the brain's language-specific circuitry, which he calls Universal Lexical Semantics. Pinker and other theorists suggest that the inductive abilities reviewed above are best explained by a combination of children's inherent linguistic endowment and rational inferences based on limited, ambiguous inputs, rather than by learning probabilistic structure from large amounts of previous language exposure (Pinker, 1994b; Bloom, 1999; Markman, 1994). Attributing specific knowledge items to inherent constraints *as opposed to* learning is conceptually problematic, as learning mechanisms might play important roles in most or all developmental trajectories, and all learning processes are biologically canalized to varying degrees (Deák, 2000). Nevertheless, researchers with different theoretical orientations agree that syntactic and semantic knowledge are important factors in word learning.

Finally, the knowledge context that children bring to word learning can be described in terms of the set of words they already know. Known words can be relevant if they are either phonologically similar or semantically related to the target word. Most simply, a word can be characterized in terms of its *neighborhood density*, i.e. the number of known related words. On the phonological side, neighbors are typically defined as word forms that differ in only one phoneme. 14-month infants can discriminate such pairs, but are inhibited from mapping them to different objects (Stager & Werker, 1997), and this inhibitory effect persists at least up to 18 months of age (Werker, Fennell, Corcoran, & Stager, 2002; Hollich, Jusczyk, & Luce, 2002; Swingley & Aslin, 2007). Nevertheless, words with higher phonological neighborhood densities are learned earlier *on average* (Storkel, 2009), suggesting that population-level vocabulary trends fail to capture the interdependence of items within an individual child's developing lexicon. On the semantic side, knowledge of semantic neighbors (i.e. same-category nouns) is associated with better performance in word learning and lexical and sentence processing (Borovsky, Ellis, Evans, & Elman, 2016a, 2016b). Whereas neighborhood density effects describe only the influence of a word's immediate neighbors, higher-order local and global relationships can be studied using network models, unlocking a further layer of complexity.

In a network, nodes represent words and edges represent phonological or semantic connections between words. Network measures can provide more complex measures of a word's local "neighborhood," which can then be evaluated as a predictor of word learning. Hills, Maouene, Maouene, Sheya, & Smith (2009) used word co-occurrences to define a semantic network, and then used normed age-of-acquisition data to model the normative growth of that network with age. They tested the ability of three local growth models to explain the trajectory of vocabulary acquisition: *preferential attachment* (enhanced learning of words that connect to well-connected known words), *preferential acquisition*, (enhanced learning of words with many connections to other words in the language), and *lure of the associates* (enhanced learning of words that connect to many known words). The preferential acquisition model best explained the growth of the normative network, a finding that was recently replicated using data from ten different languages (Fourtassi, Bian, & Frank, 2018). However, a recent study (Beckage & Colunga, 2019) applied the same models to networks constructed from individual children's vocabularies over time and found that lure of the associates was the strongest predictor; that is, children tend significantly to learn words that are related to groups of words they already know.

Networks can also be described in terms of their global structural properties, which can be related to individual differences in vocabulary development. Notably, vocabulary networks exhibit *small-world* structure (Watts & Strogatz, 1998; Ferrer i Cancho & Solé, 2001). Networks using adult free association to link words also show *scale-free* structure, in which a few “hub” words have many neighbors, and the overall number of neighbors per node (degree) follows a power law distribution (Steyvers & Tenenbaum, 2005). Following these descriptive findings in normative vocabulary networks, researchers have begun investigating global network properties as potential markers of individual variability in learning mechanisms. For example, in one study children whose lexical networks were more centered around hub words were more likely to show a "mutual exclusivity" assumption about novel word meanings (Yurovsky, Bion, Smith, & Fernald, 2012), suggesting that network structures might emerge in parallel with and as a signature of particular word learning strategies. In another study, late-talking children (i.e. children with small productive vocabularies for their age) were found to show less small-world structure in their vocabulary networks as compared to typically developing children (Beckage, Smith, & Hills, 2011). However, this pattern failed to replicate in a larger sample of children (Jimenez & Hills, 2017). So far, whereas network measures of vocabulary structure have shown promise as sensitive measures of an individual child’s learning history, more work remains to be done toward resolving the apparent contradictions by understanding how network measures depend on the particular methods used to construct and describe each child’s network.

Overall, knowledge-oriented theories provide a powerful framework to analyze lexical development as a process that occurs within a cognitive system and over a developmental timescale. Although the research reviewed above lacks a unified theoretical framework, the various studies identify areas of complexity in the developing mind that control how children can make use of their word-learning opportunities.

Integrating perspectives

The preceding sections demonstrate that social interaction, statistical learning, and structured knowledge each provide relevant support for word learning. For each type of theory, a range of evidence supports its role in facilitating the task of word learning to make it tractable and guide its trajectory along normative paths. To date, however, explanations of word learning have only just begun to address the mechanisms by which multiple factors interact to support word learning in complementary and potentially mutually reinforcing ways. Of particular interest is the possibility of complex dynamics in which infants first learn about language by association from environmental statistics, but gradually construct a suite of word-learning processes that all act to promote and constrain each other's development (Hockema & Smith, 2009). This type of dynamics might explain how relatively simple learning mechanisms guided by coarse, ambiguous inputs could lead to remarkably powerful and robust lexical acquisition. To begin understanding word learning at this level, it is therefore necessary to understand how social, informational, and knowledge factors work in combination.

Social and information-oriented theories could be unified if learnable patterns in language input emerge out of the structure of infant-parent interaction. Social interaction might facilitate language acquisition by combining speech with correlated cues in other modalities to form cohesive multimodal "packages," which might be more salient and provide cues to meaning. Parents in multiple cultures embed object-naming utterances in such multimodal packages of synchronous speech and object motion (Gogate, Bahrick, & Watson, 2000; Gogate, Maganti, & Bahrick, 2015; Matatyaho & Gogate, 2008). In one study, mothers taught 6- to 8-month-old infants two novel object words. Only infants who attended to their mothers' synchronized speech and object motion showed a higher proportion of anticipatory or first looks to the named object after a 3-minute teaching period (Gogate, Bolzani, & Betancourt, 2006). Speech about actions is also packaged by temporal alignment with the onsets and offsets of those actions (Meyer, Hard, Brand, McGarvey, & Baldwin, 2011). Thus, parents provide a parallel stream of information that might help constrain and supervise learning from the stream of speech sounds. Social interaction might also generate structured inputs for children by guiding their focus of attention. Microbehavioral analyses of infant-parent toy play show that infants strongly prefer to look at handled objects compared to faces or other objects, and this tendency leads to episodes of sustained joint attention (Yu & Smith, 2013; Deák, Krasno, Triesch, Lewis, & Sepeta, 2014), which in turn might contain sustained verbal reference to specific objects (Frank, Tenenbaum, & Fernald, 2013). Thus, rather than a static property of the ambient language, statistical properties of the language input can be understood as a result of interactive social and embodied processes (Smith, Jayaraman, Clerkin, & Yu, 2018).

Efforts to integrate information-oriented and knowledge-oriented theories involve models of statistical learning that incorporate linguistic knowledge rather than assuming a strictly associationist "blank slate." Even within a single experimental session, infants can use a previously learned statistical pattern to learn a new one. For instance, Saffran and Wilson (2003) exposed infants to an artificial language, presented as a continuous stream of syllables in which transitional probabilities over syllables defined the words, and a finite-state grammar on the words defined grammatical sentences. Although the grammar was only evident after words were identified, infants successfully discriminated grammatical from ungrammatical sentences.

Similarly, previous statistical learning might also help infants learn statistical patterns that are analogous to those they already know. In one study, 12-month-olds who were pre-trained on an artificial language with adjacent word dependencies were subsequently able to learn non-adjacent dependency rules (i.e. patterns of the form aXb, where element a predicts element b with a variable intervening element X) that infants typically cannot detect until 15 months of age (Lany & Gómez, 2008). Learning one pattern can thus facilitate learning of similar but more complex patterns, suggesting a cumulative mechanism for statistical learning. Within the domain of word learning, Yu (2008) used a computational model to show how initial knowledge of word-object mappings helps constrain the search space for new mappings. In this study, the model was trained on naturalistic data derived from recordings of parents reading books to their 20-month-old children. The basic model learned words by forming associations based on co-occurrences between words and objects. After the first round of learning, the model was augmented using the results of the preceding session: previously identified function words (i.e. frequent words that occurred with all objects) were ignored, previously learned mappings were excluded from the input, and weakly associated word-object pairs were given a higher prior probability. This cumulative model learned more correct mappings per session as it used its existing knowledge to narrow the search space for new words. Overall, these studies show that statistical learning is a cumulative process in which the learner's construal of new inputs depends crucially on past exposure to related inputs. So far, however, we have just scratched the surface of how structured knowledge might interact with statistical learning.

Finally, social and knowledge-oriented theories interact because children's knowledge relevant to word learning extends beyond knowledge about language itself. Rather, children learn how to expect people to behave and this impacts their ability to use learning opportunities. As a simple example, 18-to-20-month-old infants selectively map novel words to objects that are consistent with the speaker's gaze direction (Baldwin, 1995). Infants form expectations about actions long before first words. By 4 months, infants learn to expect turn-taking during "peek-a-boo" games such that infant gazing and smiling triggers an adult turn (Rochat, Querido, & Striano, 1999). Again at 4 months of age, infants engage in smiling exchanges in which the timing of infants' similes is optimized to elicit smiling reactions (Ruvolo, Messinger, & Movellan, 2015). At 3 months of age, parents accompany speech acts with overt bodily actions during diaper-changing, a routine and stereotyped activity (Nomikou & Rohlfing, 2011). Associations between dyadic action patterns and speech could facilitate learning by "bootstrapping" in either direction: linguistic patterns might help infants discover structure in social interaction, and patterned interactions might help infants discover structure in language.

In sum, multiple contextual factors interact to form the context of children's word learning opportunities. The research trends reviewed above provide initial examples of how favorable conjunctions of factors emerge during everyday interactions, and how developmental changes in one domain or modality might potentiate or accelerate the development of a different domain.

Moving forward, our understanding of how children solve the problem of word learning must rely on explaining the interactions between diverse contextual factors.

The studies in this dissertation explore children's word learning with an emphasis on these interactions. In one set of studies, we used a micro-level observational approach to investigate infants' exposure to object-labeling utterances within the context of dynamic infant-parent interaction. These studies help illustrate some types of multimodal, interactive contextual supports that we hypothesize might underlie some of the predictive value of word usage patterns. For example, we found that object-naming utterances were embedded in characteristic discourse sequences, and that parents produced them in response to infants' object exploration. By embedding informative utterances in predictable and interactive protoconversational sequences, parents thus supported not only infants' mapping of the correct words to the correct meanings, but also the *act* of meaning itself. By experiencing (and helping to construct) many word-learning opportunities within a similar interactive context, children might gradually become more efficient learners of groups of words whose usage patterns share a family resemblance.

In a second set of studies, we used a macro-level, data-mining approach to search for effects of words' usage patterns on the order in which children learn them. To do this, we used a corpus of child-directed speech to characterize the distributional usage of each word, and a separate database of children's vocabulary data to investigate regularities in the order in which children acquire words. If multimodal, interactive contextual embedding is important for word learning, we would expect that words that are used in similar contexts should have similar learning profiles beyond those predicted by e.g. frequency or shared semantic features. Consistent with this, we find that words that occur with similar context words and in similar sentence frames tend to have similar normative ages of acquisition, and they tend to be learned together more often than expected by chance. We also found that children with large vocabularies for their age learned sets of words with different network structure than did children with small vocabularies. Thus, macro-level patterns in children's word learning could be predicted based on the specific usage contexts of each word.

These two lines of research hint at a synthesis in which early word learning is facilitated by the statistical structure of the inputs generated through children's contingent and embodied interaction with the physical and social world, and is constrained and biased by principles and patterns induced from the child's past experience. In this way, parent-child dyads might gradually develop fluency in enacting habitual routines in which novel words can be embedded, contributing to the observed acceleration in children's word learning and leaving traces in the structure of each child's lexicon.

Outline of the dissertation

Infants' and toddlers' linguistic environments are rich in learnable associations and co-occurrences, but these are not limited to purely linguistic patterns, or to word-referent links. Instead, the entire stream of linguistic and paralinguistic behavior to which a child is exposed is a source of probabilistic structure, and the child's access to and interpretation of the stream is conditioned on their developmental stage, learning history, and prior knowledge. To understand the computational learning problems faced by infant language learners, it is necessary to characterize the multiple layers of contextual embedding of each word learning event: the linguistic context of the target word, the multimodal interactive behavioral context of the episode, and the conceptual context of the child's relevant knowledge. The studies in this dissertation each inform our understanding of word learning by characterizing aspects of this multiple embedding.

Chapters 1 and 2 use detailed behavioral observations of infants and parents engaging in a structured play task with toys to elucidate learnable patterns of probabilistic co-occurrence of speech events with linguistic and non-linguistic contexts, and how they emerge from responsive social interactions. In Chapter 1, we address how infants' developmental status and parents' contingent responsiveness jointly determine infants' exposure to word learning opportunities. In a longitudinal sample of infants aged 4 to 9 months, we show how, as infants' object handling becomes more developmentally advanced, they elicit both more object-naming utterances from parents, as well as more opportunities for parents to produce those object-naming utterances contingently on infant actions. In Chapter 2, we investigate how different types of speech content in infant-directed speech are embedded in both linguistic contexts (multi-utterance sequences) and nonlinguistic contexts (cross-modal associations between speech, visual attention, and manual object exploration). Together, these studies inform our understanding of the learnable patterns in children's language exposure not as a given input, but as an emergent product of interacting developmental systems.

Chapters 3 and 4 use data mining techniques to study how the large-scale usage patterns of words predict children's observed patterns of word learning. In Chapter 3, we introduce a novel distributional semantic representation of words. In our method, words are represented based on their patterns of usage in a large corpus of child-directed speech. From these patterns, we extract two semantic vectors: one based on a word's adjacent, ordered co-occurrences with context words ("frames"), which reflects primarily the role of the word in the sentence; and one based on a word's non-adjacent, unordered co-occurrences with context words, which reflects primarily thematic information. We use the principal components extracted from these distributional-semantic vectors to predict the normative age of acquisition of each word. The results reveal how embedding of words both in utterance contexts and activity contexts supports learning, and extend previous studies using simpler and more subjective features to predict age of acquisition. In Chapter 4, we investigate the tendencies of certain words to be learned together more than expected under independence. We measure these word-word dependencies on a large sample of children's vocabularies and use them to define networks of related words. We show that children who are precocious talkers have vocabularies with more related words and more small-world structure than do late talkers. We also show that overlap in the distributional-semantic representations introduced in Chapter 3 are highly correlated with the tendency of words to be learned together. Together, these studies show in detail how word learning involves multiple layers of context, including the utterance, the activity context, and the rest of the child's lexicon.

References

Akhtar, N., Dunham, F., & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of child language*, *18*(1), 41-49.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological science*, *9*(4), 321-324.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In Moore, C., & Dunham, P. (Eds.), *Joint attention: Its origins and role in development* (pp. 131-158), Hillsdale, NJ: Lawrence Erlbaum Associates.

Beckage, N. M., & Colunga, E. (2019). Network Growth Modeling to Capture Individual Lexical Learning. *Complexity*, 2019.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS one*, *6*(5), e19348.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of child language*, *6*(2), 183-200.

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. (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development*, *11*(4), 369-380.

Bloom, L. (1973). *One word at a time: The use of single-word utterances before syntax.* The Hague: Mouton.

Bloom, L. (2000). The Intentionality of Word Learning: How to Learn a Word, Any Word. In *Becoming a word learner: A debate on lexical acquisition* (pp. 19-50).

Bloom, P. (1999). Language capacities: Is grammar special? *Current Biology*, *9*(4), R127-R128.

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C. S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental psychology*, *44*(3), 867.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental science*, *19*(6), 918-932.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child development*, *87*(6), 1893-1908.

Bronfenbrenner, U. (1979). *The ecology of human development*. Harvard university press.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. *Handbook of child psychology.*

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*(3), 407-416.

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

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using ‘frequent frames’: what cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental science*, *12*(3), 396-406.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and cognitive processes*, *13*(2-3), 221-268.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, *31*(5), 489-509.

Clearfield, M. W. (2011). Learning to walk changes infants’ social interactions. *Infant Behavior and Development*, *34*(1), 15-25.

Deák, G. O. (2000). Hunting the fox of word learning: Why “constraints” fail to capture it. *Developmental Review*, *20*(1), 29-80.

Deák, 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*(2), 270-281.

Deák, G. O., & Maratsos, M. (1998). On having complex representations of things: Preschoolers use multiple words for objects and people. *Developmental Psychology*, *34*(2), 224.

Deák, G. O., & Toney, A. J. (2013). Young children’s fast mapping and generalization of words, facts, and pictograms. *Journal of experimental child psychology*, *115*(2), 273-296.

Dore, J., Franklin, M. B., Miller, R. T., & Ramer, A. L. (1976). Transitional phenomena in early language acquisition. *Journal of child language*, *3*(1), 13-28.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321-335.

Ferrer i Cancho, R., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1482), 2261-2265.

Fourtassi, A., Bian, Y., & Frank, M. C. (2018). Word Learning as Network Growth: A Cross-linguistic Analysis. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society.*

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.

Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental psychology*, *40*(4), 621.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of child language*, *32*(2), 249-268.

Gleitman, L. (1990). The structural sources of verb meanings. *Language acquisition*, *1*(1), 3-55.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child development*, *71*(4), 878-894.

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6-to 8-month-old infants and learning of word–object relations. *Infancy*, *9*(3), 259-288.

Gogate, L., Maganti, M., & Bahrick, L. E. (2015). Cross-cultural evidence for multimodal motherese: Asian Indian mothers' adaptive use of synchronous words and gestures. *Journal of experimental child psychology*, *129*, 110-126.

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of child language*, *17*(1), 171-183.

Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental science*, *10*(6), 778-785.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological science*, *19*(5), 515-523.

Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy*, *15*(4), 362-391.

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*(1), 21-54.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental science*, *7*(5), 567-580.

Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy*, *19*(4), 385-408.

Hall, D. G., Lee, S. C., & Bélanger, J. (2001). Young children's use of syntactic cues to learn proper names and count nouns. *Developmental Psychology*, *37*(3), 298.

Hare, M., McRae, K., & Elman, J. (2004). Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes*, *19*(2), 181-224.

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

He, M., Walle, E. A., & Campos, J. J. (2015). A cross-national investigation of the relationship between infant walking and language development. *Infancy*, *20*(3), 283-305.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological science*, *20*(6), 729-739.

Hockema, S. A., & Smith, L. B. (2009). Learning your language, outside-in and inside-out. *Linguistics*, *47*(2), 453-479.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child development*, *74*(5), 1368-1378.

Hollich, G., Jusczyk, P. W., & Luce, P. A. (2002). Lexical neighborhood effects in 17-month-old word learning. In *Proceedings of the 26th annual Boston University conference on language development* (Vol. 1, pp. 314-323). Boston, MA: Cascadilla Press.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of child language*, *37*(2), 229-261.

Jimenez, E., & Hills, T. (2017). Network Analysis of a Large Sample of Typical and Late Talkers. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Kamhi, A. G. (1986). The elusive first word: The importance of the naming insight for the development of referential speech. *Journal of child language*, *13*(1), 155-161.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child development*, *82*(4), 1199-1209.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental science*, *17*(3), 388-395.

Kärtner, J., Keller, H., Lamm, B., Abels, M., Yovsi, R. D., Chaudhary, N., & Su, Y. (2008). Similarities and differences in contingency experiences of 3-month-olds across sociocultural contexts. *Infant Behavior and Development*, *31*(3), 488-500.

Kaye, K., & Fogel, A. (1980). The temporal structure of face-to-face communication between mothers and infants. *Developmental psychology*, *16*(5), 454.

Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of experimental psychology: human perception and performance*, *14*(3), 389.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental science*, *10*(1), 110-120.

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.

Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*(12), 1247-1252.

Libertus, K., & Violi, D. A. (2016). Sit to talk: relation between motor skills and language development in infancy. *Frontiers in Psychology*, *7*, 475.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological science*, *18*(5), 387-391.

Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua*, *92*, 199-227.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive psychology*, *16*(1), 1-27.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, *20*(2), 121-157.

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word–object relations. *Infancy*, *13*(2), 172-184.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.

McGillion, M. L., Herbert, J. S., Pine, J. M., Keren-Portnoy, T., Vihman, M. M., & Matthews, D. E. (2013). Supporting early vocabulary development: What sort of responsiveness matters? *IEEE Transactions on Autonomous Mental Development*, *5*(3), 240-248.

McPherson, L. M. P. (1991). A little goes a long way: Evidence for a perceptual basis of learning for the noun categories COUNT and MASS. *Journal of child language*, *18*(2), 315-338.

Meyer, M., Hard, B., Brand, R. J., McGarvey, M., & Baldwin, D. A. (2011). Acoustic packaging: Maternal speech and action synchrony. *IEEE Transactions on Autonomous Mental Development*, *3*(2), 154-162.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91-117.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302-331.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143-182.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive psychology*, *55*(4), 259-305.

Movellan, J. R., & Watson, J. S. (1987). Perception of directional attention. In *Infant behavior and development: Abstracts of the sixth international conference of infant studies*.

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant behavior and development*, *21*(3), 469-482.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of child language*, *17*(2), 357-374.

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science*, *6*(2), 136-142.

Nomikou, I., & Rohlfing, K. J. (2011). Language does something: body action and language in maternal input to three-month-olds. *IEEE Transactions on Autonomous Mental Development*, *3*(2), 113-128.

Oudgenoeg-Paz, O., Volman, M. C. J., & Leseman, P. P. (2012). Attainment of sitting and walking predicts development of productive vocabulary between ages 16 and 28 months. *Infant Behavior and Development*, *35*(4), 733-736.

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry*, *52*(5), 588-598.

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, *92*, 377-410.

Pinker, S. (1994). *The language instinct*. New York, NY, US: William Morrow & Co.

Quine, W. V. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, *29*(6), 1007-1028.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive science*, *22*(4), 425-469.

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, *24*(2), 221-234.

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental psychology*, *28*(3), 406.

Rochat, P., Querido, J. G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental psychology*, *35*(4), 950.

Ruvolo, P., Messinger, D., & Movellan, J. (2015). Infants time their smiles to make their moms smile. *PloS one*, *10*(9), e0136492.

Sachs, J., Bard, B., & Johnson, M. L. (1981). Language learning with restricted input: Case studies of two hearing children of deaf parents. *Applied Psycholinguistics*, *2*(1), 33-54.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*(3), 669-680.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273-284.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of child language*, *25*(1), 169-201.

Smith, L. B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. *The emergence of language*, 277-303.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences*, *22*(4), 325-336.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558-1568.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of child language*, *4*(1), 1-22.

Snow, C. E., Arlman-Rupp, A., Hassing, Y., Jobse, J., Joosten, J., & Vorster, J. (1976). Mothers' speech in three social classes. *Journal of Psycholinguistic Research*, *5*(1), 1-20.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41-78.

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of child language*, *36*(2), 291-321.

Sullivan, J., & Barner, D. (2016). Discourse bootstrapping: Preschoolers use linguistic discourse to learn new words. *Developmental science*, *19*(1), 63-75.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive psychology*, *54*(2), 99-132.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child development*, *72*(3), 748-767.

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

Tamis-LeMonda, C. S., Kuchirko, Y., & Tafuro, L. (2013). From action to interaction: Infant object exploration and mothers' contingent responsiveness. *IEEE Transactions on Autonomous Mental Development*, *5*(3), 202-209.

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

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

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, *3*(4), 269-293.

Trueswell, J. C., Lin, Y., Armstrong III, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, *148*, 117-135.

Walle, E. A. (2016). Infant social development across the transition from crawling to walking. *Frontiers in psychology*, *7*, 960.

Walle, E. A., & Campos, J. J. (2014). Infant language development is related to the acquisition of walking. *Developmental psychology*, *50*(2), 336.

Watson, J. S. (1972). Smiling, cooing, and "the game". *Merrill-Palmer Quarterly of Behavior and Development*, *18*(4), 323-339.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, *393*(6684), 440.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1-30.

Wu, Z., & Gros-Louis, J. (2014). Infants’ prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, *34*(1), 72-90.

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language learning and Development*, *4*(1), 32-62.

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

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS one*, *8*(11), e79659.

Yurovsky, D., Bion, R. A., Smith, L. B., & Fernald, A. (2012). Mutual exclusivity and vocabulary structure. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society.*

Zumbahlen, M. R. (1997). The role of infant locomotor onset in shaping mother-infant communication (Doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois.

CHAPTER 1: Contingencies Between Infants' Gaze, Vocal, and Manual Actions and Mothers' Object-Naming: Longitudinal Changes From 4 to 9 Months

Abstract

Infants’ early motor actions help organize social interactions, forming the context of caregiver speech. We investigated changes across the first year in social contingencies between infant gaze and object exploration, and mothers’ speech. We recorded mother–infant object play at 4, 6, and 9 months, identifying infants’ and mothers’ gaze and hand actions, and mothers’ object naming and general utterances. Mothers named objects more when infants vocalized, looked at objects or the mother’s face, or handled multiple objects. As infants aged, their increasing object exploration created opportunities for caregiver contingencies and changed how gaze and hands accompany object naming over time.

Introduction

To acquire language, infants must solve the word-to-world mapping problem, associating the language they hear with concurrent nonlinguistic experiences. Previous work has shown that features of parents’ speech to infants, including total number of word tokens and mean length of utterance, predict their vocabulary as toddlers (Hart & Risley, 1995; Hoff, 2003). In addition to speech quantity, the timing and nonverbal context of caregiver speech are features that correlate with early language development. Critically, infants are not merely passively exposed to a stream of words and percepts. Rather, infants’ self-generated exploratory activity directly and indirectly affects the structure of their experience with language. Specifically, some actions might elicit different amounts or types of caregiver speech. That is, caregivers might help structure the infant’s language input by monitoring infants’ activity and producing speech contingent on infant’s engagement with objects and people. Over the course of development, therefore, infants’ developing motor abilities might influence their language development by shaping the non-verbal context that accompanies caregiver speech or by giving caregivers increasing opportunities to respond to infant actions. In the current report, we examine the temporal dynamics of caregiver speech, and accompanying non-verbal actions, during unscripted mother-infant interactions. By following infants and mothers longitudinally from 4 to 9 months, we describe changes in these dynamics during the first year. Specifically, we examine how infants' actions elicit maternal speech in general, and object naming in particular, because object nouns dominate infants' first word-world associations (Gentner, 1982; but see Tardif, 1996) which for many common objects are established by 9 months (Bergelson & Swingley, 2012). In the following section, we detail what is known from previous studies regarding the features of mother–infant interaction that shape infants' language experience and development.

Contingent responsiveness

An important feature of infants' early interactions with caregivers is the timing and appropriateness of caregiver actions relative to infant actions. This has been described in the literature as caregiver sensitivity, responsivity or contingency. Caregivers respond contingently to infants' behaviors from very early in infancy: for example, infants and caregivers engage in sequences of mutually contingent patterns of looking, smiling, and vocalizing by 1 to 3 months or earlier (Kärtner et al., 2008; Kaye & Fogel, 1980).

Many previous studies have described caregiver responsiveness in qualitative terms, with features such as "appropriateness" or consistency with the infants' ongoing activity or attention (e.g., Kochanska & Aksan, 2004). Others have characterized the frequency and timing of specific responses to infants' actions. For instance, infants and mothers engage in vocal turn-taking (Snow, 1977), and mothers also respond contingently to infants' object-exploratory actions (Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008). By 1 year of age, infants can elicit object naming from parents by object exploration (Wu & Gros-Louis, 2014b, Yu & Smith, 2012) and by explicit gestures (Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007; Olson & Masur, 2011, 2013). At 14 months, Tamis-LeMonda, Kuchirko, and Tafuro (2013) observed that mothers contingently responded to infants' object exploration actions by handling the objects themselves and by producing more referential language. Such specific caregiver responses to infants' actions might impact specific developmental outcomes, including language development (Tamis-LeMonda, Kuchirko, & Song, 2014).

A large body of evidence suggests that infants learn from contingently presented information. Two-month-olds will direct social actions to a mobile that moves in response to their head movements (Watson, 1972), and even by 2 weeks infants can learn to modify their sucking rate around their mother's activity during breastfeeding (Kaye & Wells, 1980). Caregivers' contingent responses to their infants also matter specifically for language learning. For example, Tamis-LeMonda, Bornstein, and Baumwell (2001) found that maternal responsiveness at 9 and 13 months positively predicted toddlers' language development. Similar correlational findings have been reported elsewhere (McGillion et al., 2013, Gros-Louis, West, & King, 2014, Wu & Gros-Louis, 2014a, 2014b). Certain patterns can specifically facilitate prelinguistic vocal production and word learning. For example, Goldstein and Schwade (2008) found that if adults vocally responded to 9-month-olds' vocalizations, the infants' vocalizations became more advanced, incorporating phonological elements from the adults' responses. Moreover, 12-month-olds tended only to learn object names that were presented contingent on the infant vocalizing while looking at the object (Goldstein, Schwade, Briesch, & Syal, 2010). In addition, if mothers respond contingently to infants' manual gestures by "translating" the gestures into words, infants are more likely to learn to produce those words (Goldin-Meadow et al., 2007; Masur, 1982; see also Dimitrova, Özçalışkan, & Adamson, 2016). Other research has focused on caregivers' responses that match or "follow in" to infants' focus of attention (Tomasello & Farrar, 1986). For example, caregivers' tendency to produce speech acts related to infants' focus of attention at 9 months predicted infants' subsequent language comprehension (Rollins, 2003).

In the current study, we investigated potential longitudinal changes in the action contingencies that shape infants' language input across the first year. Bornstein et al. (2008) longitudinally followed mothers' verbal responses to infants' vocalizations and object exploration at 10, 14, and 21 months, and found specificity between infant behaviors and maternal response types, as well as changes in overall frequencies of infant behaviors and maternal response types with age. However, mothers' rates of response to infant behaviors were generally stable across ages. We extended this approach to a younger age group, describing contingencies between infants' gaze, hand, and vocal actions and mothers' speech between 4 and 9 months. Specifically, we investigated the frequencies of different infant behaviors and maternal responses over that period, and any changes in the contingent relations between them.

Infant motor development

Infants' motor activity directly and indirectly affects the structure of their experience with language (Iverson, 2010). Infants' object manipulation and gaze directly influence their sensory experience during caregiver speech (Yoshida & Burling, 2013). Additionally, infants' actions—looking, handling, and vocalizing, among others—can indirectly influence their experience by eliciting caregiver actions. Therefore, advances in infants' motor skills are likely to change their language experience in two ways: by shaping their sensory (especially visual and haptic) experience during speech, and by eliciting different speech from caregivers. For example, 13-month-olds who have begun walking elicit more action directives in response to their object bids than do crawling infants (Karasik, Tamis-LeMonda, & Adolph, 2014). At 8 months, crawling infants elicit more prohibitions than noncrawling infants (Zumbahlen, 1997). During the first year, infants' growing motor proficiency in manipulating objects might also trigger changes in caregiver speech. Coordination of gaze and manual activity changes rapidly during the first year as infants become more proficient at reaching, manipulating objects, and flexibly directing their gaze and hands to things in the world (Fagard & Lockman, 2010; Rochat, 1989; Ruff, 1984; von Hofsten, 1991) and coordinating manual and vocal activity (Iverson & Fagan, 2004). Previous work in our lab has shown that in dyadic toy play interactions, infants decrease their object gaze time and increase object handling across the first year (de Barbaro, Johnson, Forster, & Deák, 2016). Additionally, infants increasingly manipulate two objects simultaneously (de Barbaro et al., 2016) and in more complex ways (de Barbaro, Johnson, & Deák, 2013). We hypothesize that these and other changes in object manipulation and other motor skills will affect the language that infants hear by increasing how often they produce behaviors that elicit contingent speech. Opportunities to learn language through social contingencies might thereby expand as infants' behavioral skills develop and diversify.

Correlates of caregiver speech

In addition to being produced contingently on infant actions, caregiver speech is also "packaged" with regularities in ongoing sensorimotor activity (Meyer, Hard, Brand, McGarvey, & Baldwin, 2011). Infants' experiences unfold in social exchanges in which both partners structure the visual and auditory scene through multiple behavioral modalities, including vocalizations, gaze, and manual activity (Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; de Barbaro et al., 2016; Yu & Smith, 2013). Mapping words to referents is likely to be facilitated by concurrent sensory and motor experiences around the times when objects are named. For example, 18-month-old children's learning of a novel object word is predicted by the size of the object in their visual field when it is named (Yu & Smith, 2012).

Some of the regularities in activity around caregiver speech are provided by caregivers' paralinguistic actions. During infants' first year, caregivers' object-name productions are embedded in multimodal behavioral complexes featuring synchronous speech and object motion (Gogate, Bahrick, & Watson, 2000; Gogate, Maganti, & Bahrick, 2015; Matatyaho & Gogate, 2008). In one study, mothers taught 6- to 8-month-old infants two novel object words. Only infants who attended to their mothers' synchronous speech and object motion showed a higher proportion of anticipatory or first looks to the named object after a 3-minute teaching period (Gogate, Bolzani, & Betancourt, 2006). Thus, production of multimodal communicative actions by caregivers contributes to early word learning. However, such studies of caregivers' object naming with infants under a year of age have not addressed how infants' actions might contribute to the “packaging” of speech with ongoing activity. We therefore investigated, across the age range studied, associations between caregiver speech and either infant or caregiver actions occurring before, during, or after speech. We expected that these associations would increasingly involve infants’ manual activity at the expense of mothers’ as infants became more active participants in the interaction.

Present study

The present study documents changing contingencies between infant exploratory and social behaviors from 4 to 9 months, and their mothers’ vocal responses including utterances about objects, recorded within a longitudinal sample of unscripted toy-play interactions. We focused on the following questions: What infant gaze, hand, and vocal actions predict maternal speech in general, or object naming in particular? Do age-related changes in infants’ object exploration lead mothers to produce more predictable object-naming or overall speech? Across the age range studied, what sequences of infant and mother manual or visual engagement with objects tend to accompany object naming? We predicted that mothers’ verbal responses to infants’ behaviors would be specific and relatively stable. We also predicted that, with increasing age, object naming would become more strongly associated with infants’ own manual activity rather than that of the mother. Finally, we predicted that infants’ increasingly differentiated manual activity would strengthen contingencies by giving mothers of older infants more opportunities to respond.

Methods

Participants

Participants were 42 mother–infant dyads (20 female) from a longitudinal study of infant social development (Deák, Triesch, Krasno, de Barbaro, & Robledo, 2013). Participants were recruited as a sample of convenience from the greater San Diego area. Mothers’ mean age upon recruitment was 32.1 years (range = 21–42) and they had completed a mean of 16.1 years of formal education (range = 12–21). Twenty-nine infants were Caucasian, 2 were Asian, 4 were Hispanic, 5 were other or multiple races, and 2 did not report their ethnicity. None of the infants had any neurological, cognitive, or sensory deficits, according to parental report. Six additional participants dropped out of the longitudinal study. An experimenter visited the participants' home each month between 4 and 9 months, and again at 12 months (participants also visited the laboratory to complete various tasks every month; those data are reported elsewhere, e.g., Deák, 2015; Ellis, Gonzalez, & Deák, 2014). Data from the 4, 6, and 9-month home sessions were analyzed for this study. Participants were observed within 2 weeks of the infant's 4, 6, or 9-month birthday (five sessions had to be rescheduled; these were completed before the infant's next month's birthday. Infants' mean age was 125 days at the 4-month session (range: 113–142), 186 days at the 6-month session (range: 175–211), and 277 days at the 9-month session (range: 260–300). Due to infant fussiness or equipment failure, one dyad did not complete the 4-month session, two did not complete the 6-month session, and one did not complete the 9-month session; however, these dyads' data from the remaining sessions were included in analyses. In addition, for several sessions one or more specific variables could not be coded because of video recording problems or because the mother predominantly used a non-English language. Therefore, the number of participants whose complete data were analyzed was 35 at 4 months, 38 at 6 months, and 39 at 9 months.

Procedure and coding

During each session, infants were seated in a modified walker with a tray, and mothers were seated on a pillow facing the infant (Figure 1.1). This arrangement controlled for differences in postural stability, distance between the participants, and the angle between the infant and mother. In this position, mothers could keep their infant's face and hands within their visual field when facing forward. Three cameras mounted on tripods recorded the interaction from different angles: one was centered on the infant's head and upper body, one on the mother's face and upper body, and one was further away, positioned lateral to the dyad, and zoomed out to capture both participants and their nearby environment. Videos were digitized and synchronized post-production, to facilitate coding (described in the following section).

At each session, the dyad interacted with three novel toys (mothers verified the novelty of all toys before the first session). The toys were the same for all participants, but different at each month. The toys were: a box with buttons, a caterpillar with rings, and a wobbling wolf doll at 4 months; a plush soccer ball, a light-up ring-shaped toy, and a wobbling chicken doll at 6 months; and a plush football, a light-up rattle, and a wobbling doll at 9 months. At the start of the session, two toys were placed in wells at the sides of the walker tray. Mothers were instructed to "play as they normally would" with their infants, but to try to leave one toy on the tray at a time and return the others to the wells (this was intended to facilitate coding); however, toys in the wells were visible to infants and within their reach, and infants could (and did) freely retrieve toys from the wells. Additionally, for purposes of a study to be described elsewhere, mothers were instructed to occasionally draw infants' attention to two targets located out of reach in specific locations; thus, mothers freely chose times (at least once per target) to punctuate the object-play interaction with a brief bid to re-direct the infant's attention. This activity was designed to represent the common situation of a caregiver interrupting an interaction to call an infant's attention to something. Mothers spent an average of 8% of the interaction time engaged in this secondary attention-directing task. Excluding those intervals had no significant effect on our results; therefore, here we report analyses of the entire session including the attention-directing intervals. The first three minutes of each session were analyzed because most sessions included at least three minutes of interaction during which the infant remained attentive. Mean session duration was 171.8 seconds at 4 months, 178.8 seconds at 6 months, and 177.0 seconds at 9 months.

We have previously reported data on changes in infant and mother gaze patterns and hand-object contact in a subsample of 26 mother–infant dyads (randomly selected) from this dataset (de Barbaro et al., 2016). For the current study, we expanded this dataset to include the entire sample (42 dyads). Additionally, for all dyads we coded all utterances, defined as bouts of meaningful speech separated by nonvocalizing periods of ≥200 ms. Nonlinguistic maternal vocalizations (e.g., sound effects, gasps, emotive sounds, etc.) were not included in the current analyses. Utterances were additionally classified as naming (i.e., containing a conventional label for one of the toys) or non-naming. In addition, we coded infant prelinguistic vocalizations, excluding cries, grunts, burps, and other organic sounds. For all dimensions coded, coders (blind to specific hypotheses) annotated the videos frame-by-frame at 10 Hz. The set of behaviors coded and reported here is detailed in Table 1.1. Figure 1.2 shows a sample time series of data from one session.

For reliability purposes, a second coder independently annotated infant gaze in 37% of the entire sample, mother gaze in 22%, infant hands in 33%, and mother hands in 37% of the sample (reliability samples were quasi-randomly chosen and age-stratified). Reliability was calculated separately for each variable. Intercoder agreement was 83% for infant gaze, 85% for mother gaze, 89% for infant hands, and 95% for mother hands. Cohen's kappa (Cohen, 1968) was .79 for infant gaze, .78 for mother gaze, .85 for infant hands, and .92 for mother hands. A second coder also independently transcribed maternal speech and infant vocalizations for 30 sessions. Agreement for the timing of vocalization onsets was calculated as the average proportion of all vocal events with onsets matching within 200 ms. Agreement averaged 90% for mother utterances and 68% for infant vocalizations.

Statistical analysis

Our analytic approach was as follows: We first report overall age-related trends in behavior. Next, we describe contingencies between infant actions and mother speech. Finally, we present the detailed time course of dyadic gaze and hand activity preceding and following naming utterances.

Repeated measures ANOVAs (rmANOVA) were used to test for age-related trends in the overall rates of all maternal utterances, maternal naming utterances, infants' and mothers' gaze to objects, and infants' and mothers' handling of objects. Conditional probability models were then used to examine rates of maternal utterances—all utterances, and specifically naming utterances—as a function of infants' gaze, manual behaviors, and vocalizations. Separate probability models were computed for all utterances and naming-utterances, relative to each type of infant behavior. Infant behaviors included the infants' gaze target, and each possible shift in gaze targets (seven types: object to object, object to face, face to object, object to other, other to object, face to other, or other to face); infants' object handling (i.e., number of objects handled: zero, one, or multiple), and each possible shift in object-handling (five types: zero to one, one to zero, one to multiple, multiple to one, and one to one). To model contingencies to gaze and object handling shifts, we compared the rate of utterances within 2 seconds after the shift to the rate over all other times[1]. Contingencies to infant gaze targets and object handling states were modeled directly using the rates of utterances at times infants were in each state.

Models were estimated using mixed-effects Poisson regression. This procedure tests for differences in the rates of a random point event (e.g., utterance onsets) across defined periods of time (infants' behavior states). It is a suitable analytic approach because other measures (e.g., proportion of actions that elicit a response) are sensitive to differences in base rate, and because the relative rate statistic can be interpreted as the magnitude of the signal available to the infant (Coxe, West, & Aiken, 2009). Models included age as a within-subjects continuous predictor, subject identity as a random effect, and infant behavior states (e.g., gaze at mother's

[1] The 2-second window was chosen based on previous reported results showing that a 3-second window is optimal for detecting contingencies between infant actions and caregiver responses (Van Egeren, Barratt, & Roach, 2001); however, because we are considering more fine-grained and frequent behaviors (e.g., gaze shifts), we adopted a shorter window to minimize spurious contingencies. Nevertheless, to ensure that the results do not depend narrowly on our use of the 2-second criterion, we repeated all analyses with 1.5-and 2.5-second windows. The results using those windows were all qualitatively similar to those reported in the text. Contingencies to infant gaze targets and object handling states were modeled directly using the rates of utterances at times infants were in each state.

face vs. objects vs. other) as a within-subjects factor. Age and subject identity were included in the models to control for individual and age differences in the base rate of maternal speech; however, for conciseness, only the effect of the infant behavior variable is reported. Whenever this effect was significant, we also report a model that includes the infant behavior X age interaction term, to show whether the contingency differed in strength with age.

Because we found a significant, novel effect in which infants' object-handling shifts from one to multiple objects (or vice versa) predicted mothers' naming utterances, we further investigated those shifts using rmANOVA to test for age-related changes in frequency, as well as session-wise correlations between the frequency of such shifts and mothers' naming rates.

We then characterized the time course of infants' and mothers' allocation of visual and manual attention to named objects before and after naming utterances. At each age, we computed the probabilities that each of these four modalities was engaged with the named object at different times relative to naming utterances. For each modality, all naming events were aligned (i.e., utterance onset time = 0) and, for every 0.1 second time step (between −10.0 and +10.0 sec), the proportion of events during which that modality (i.e., gaze or hands of mother or infant) was focused on the named object was calculated. Each modality is thereby represented as a time series of proportions (i.e., that the modality was focused on the named object at given time step over a window of ± 10 sec from naming utterance onset). At each time step, the probability of modality-engagement was compared to a baseline defined as the proportion of the entire session during which that modality was engaged with the object. The p-values of t-tests were thresholded at p = .05.

Results

Descriptive statistics

The average rates per minute of all maternal utterances, and of naming utterances specifically, at each age, are presented in Table 1.2. rmANOVAs with age as a within-subjects factor revealed significant effects of age on rate of total utterances, $F(2,67) = 3.59$, $p < .05$, and on rate of naming utterances, $F(2,67) = 11.31$, $p < .001$. Post hoc tests comparing the 4 and 6 month sessions and the 6 and 9 month sessions were computed, using critical $p = .025$ to correct for multiple comparisons. These showed that rate of total utterances decreased from 6 to 9 months, $t(35) = -2.94$, $p < .01$, but the rate of naming utterances increased marginally from 4 to 6 months, $t(32) = 2.13$, $p < .05$, and significantly from 6 to 9 months, $t(35) = 2.82$, $p < .01$.

Table 1.3 shows the mean proportions of time that the mother or infant, respectively, either looked at or handled at least one object, at each age. These analyses replicate our previous findings (de Barbaro et al., 2016) in the current larger dataset. Specifically, rmANOVAs with age as a within-subjects factor revealed a significant effect of age on prevalence of infant gaze, infant handling, and mother handling (ps < .001), but not on mother gaze, $p = .39$. Post hoc tests showed that infant gaze to objects decreased from 4 to 6 months, $t(40) = -3.11$, $p < .005$, and from 6 to 9 months, $t(40) = -4.82$, p < .001. Infant object handling increased from 4 to 6 months, $t(40) = 7.21$, $p < .001$, and showed an increasing trend from 6 to 9 months, $t(41) = 1.84$, $p = .07$. Mothers' object handling decreased from 4 to 6 months, $t(38) = -6.20$, $p < .001$, and from 6 to 9 months, $t(41) = -2.87$, $p < .01$.

Verbal responses to infants' gaze

We investigated maternal responsiveness to infant gaze patterns by testing whether mothers' utterance rate or naming rate differed as a function of either infants' current gaze target or recent changes in infants' gaze target. Infants' gaze target was classified as one of three types: Face, Object, or Other. Rates of maternal speech contingent on infants' gaze target are shown in Table 1.4. Mothers produced fewer total utterances when the infant looked at objects ($p < .001$) and more utterances when the infant looked at her face ($p < .001$), than at other locations. In contrast, mothers produced more naming utterances when the infant looked either at objects ($p < .001$) or at her face ($p < .01$). Thus, when infants looked at their mother's face, mothers talked more overall, but when infants looked at objects, mothers named objects more and talked less overall.

Next we examined whether maternal utterance rates were contingently related to any of seven types of infant gaze shifts: object-to-other, other-to-object, object-to-face, face-to-object, face-to-other, other-to-face, and object-to-object. Applying a Bonferroni correction to each set of seven tests, the significance level was set at $p = .007$.

The results, summarized in Table 1.5, are generally consistent with the analyses of gaze targets. Mothers' total utterances were significantly more frequent after infants' gaze shifts from objects to face, other to face, and face to other (*p*s < .001). Naming utterances were significantly more frequent after infants' shifts from face to objects ($p < .005$) and from other to objects (p < .005).

To test whether contingencies between infant gaze and maternal speech changed as a function of infant age, we repeated the regressions that showed significant effects, with the addition of age X infant gaze interaction terms. The interaction was not significant in any case except for a positive interaction between age and gaze to objects on naming utterances, $\beta = .147$, $p < .005$. That is, as infants got older, mothers responded more contingently to infants' object-gaze by naming objects.

Verbal responses to infants' hand activity

We investigated maternal responsiveness to infants' hand activity by testing whether mothers' utterance rate or naming rate differed as a function of either infants' number of objects handled or recent changes in object handling. Infants' object handling was divided into times when infants handled no object, one object, or multiple objects. No significant relations were found between number of objects handled, and either naming or total utterances (Table 1.6).

Next we examined whether maternal utterance rates were contingently related to any of five types of shifts in infants' object handling: from no object to one object, one object to no object, one object to multiple objects, multiple objects to one object, or one object to another object. Applying a Bonferroni correction to each set of 5 tests, the critical significance level was set at $p = .01$. Results are summarized in Table 1.5. Mothers' total utterances were not significantly contingent on object-handling shifts. However, naming utterances were significantly contingent on two shift types: from one object to multiple objects and multiple objects to one object (ps < .01).

To test whether maternal contingent vocal responsiveness to infant hand actions changed as a function of the infant's age, we repeated the regressions that showed significant effects, adding age X infant behavior interaction terms. None of the interaction terms reached significance (ps > .2). Therefore, maternal vocal responsiveness to infants' manual shifts did not change with age, although the low frequency of shifts involving multiple objects at 4 months limited our ability to detect an age interaction.

Verbal responses to infant vocalizations

We investigated maternal responsiveness to infants' prelinguistic vocalizations by testing whether mothers' utterance rate or naming rate were predicted by infants' vocalizations (Table 1.5). The 2-second window defining contingent responses started at the offset of infant vocalizations rather than the onset because vocal turn-taking tends to avoid overlap (Sacks, Schegloff, & Jefferson, 1974). As expected, mothers' total utterances were significantly contingent on infant vocalizations ($p < .01$). However, naming utterances were not contingently related to infant vocalizations ($p = .58$).

Relations between multiple-object handling and naming utterances

The foregoing analyses confirmed our prediction that mothers' speech is contingent on their infant's manual activity. Specifically, naming utterances were predicted by infants' shifts from handling one object to multiple objects or vice versa. In addition, naming utterances increased in overall frequency as infants aged. Previous work in our lab indicates that infants increasingly manipulate multiple objects simultaneously from 4 to 6 months and from 6 to 9 months (de Barbaro et al., 2016). This suggests a potential developmental pathway whereby changes in the language infants hear—specifically naming— are mediated by developmental changes in their manual activity. To evaluate whether that pathway is consistent with our current results, we first investigated whether infants' multiple-object handling shifts indeed increased with age. Second, we investigated whether, at each age, those shifts correlate with the overall rate of naming utterances.

The rmANOVAs with age as a within-subjects factor revealed that infants' multiple-object handling shifts increased with age, both as a rate per minute, $F(2,80) = 11.9$, $p < .001$, and as a proportion of handling shifts, $F(2, 76) = 27.1$, $p < .001$ (Figure 1.3). Post hoc tests showed that the rate of multiple-object handling shifts increased significantly from 4 to 6 months, $t(40) = 3.22$, $p < .005$, and marginally from 6 to 9 months, $t(41) = 2.21$, $p < .05$. Similarly, the proportion of handling shifts that involved multiple objects increased from 4 to 6 months, $t(39) = 4.62$, $p < .001$, and from 6 to 9 months, $t(40) = 3.23$, $p < .005$. The partial correlation between infants' multiple-object handling shifts and mothers' rate of total and naming utterances, controlling for the infant's age in days, was calculated at each month (Table 1.7). Applying a Bonferroni correction to each group of 3 tests (across ages), the significance level was set at $p = .017$. As expected, infants' rates of multiple-object handling shifts were not significantly related to mothers' total utterances (ps > .06). However, multiple-object shifts were significantly positively correlated with mothers' rate of naming utterances at 4 months ($p = .01$) and at 6 months ($p = .01$), though not at 9 months ($p > .3$). This pattern is consistent with the hypothesis that developmental changes in infants' manual coordination influence not only the timing but also the content of mothers' speech: not only did infants' multiple-object handling shifts increase with age in parallel with mothers' increasing rate of naming, but in addition, at 4 and 6 months individual differences in multiple-object handling predicted mothers' rate of object naming.

Time course of activity before and after naming utterances

We next described the time course of infants' and mothers' allocation of visual and manual attention to named objects before and after naming utterances, computing, for each of four modalities (infant gaze, infant hands, mother gaze, mother hands), time series of proportions representing the probability that the modality was engaged with the named object at each time relative to the onset of a naming utterance. Infant gaze and hands time-series are shown in Figure 1.4, and mother gaze and hands time-series are shown in Figure 1.5. At each time step, we tested whether each modality was engaged with the named object significantly more than a baseline defined as the proportion of the entire session during which that modality was engaged with the object.

At 4 months, infants frequently directed gaze to the named object briefly before naming, and for a more prolonged period after naming, whereas their hands were not differentially directed to the named object around naming. At 6 months, both gaze and hands were significantly more often directed to the object for the entire window. However, the shape of the temporal profile changed from 4 to 6 months in that at 6 months, infant gaze peaked around the time of naming, and handling was more frequent after naming than before. At 9 months, the shapes of the temporal profiles were similar to those at 6 months, but infants' gaze and hands were significantly directed to the named object during a more precise window of time relative to naming.

Mothers' gaze and object handling around naming utterances also changed with infant age. Notably, at 6 months, associations between naming and both maternal modalities also robustly exceeded chance across the entire 20 s window, whereas the associations were more temporally precise at 9 months. Because mothers' visuomotor skills presumably did not change, this suggests that mothers adapted their sensorimotor patterns to infants' increasingly fluid exploration. However, the temporal profile of mothers' handling of named objects differed from that of infants. Specifically, at all ages, mothers' handling peaked in synchrony with naming onset, whereas infants, with age, increasingly handled the named object *after* naming onset.

Discussion

Although recent research demonstrates the importance of contingency detection in infant learning, we are only beginning to understand how structured social contingencies change during the first year and contribute to infants' experience with language, people, and objects.

We examined relations between occurrences of mothers' speech acts, and object-naming utterances in particular, and infants' and mothers' looking and object-handling actions. The results reveal regularities in the patterning of speech and exploratory actions that could support infants' word learning across the first year. Mothers' speech was contingent on infants' gaze, manual actions, and vocalizations. These contingencies are potential cues that could help infants learn not only specific object names, but also how their own actions influence caregivers' speech. In addition, contingencies were specific to speech content and changed with age as infants produced different sets of actions.

Some contingencies involve simply infant's visual attention: mothers were more likely to speak when their infant looked at them, and less likely to speak when their infant looked at objects. This pattern is consistent with Lloyd and Masur's (2014) report that mothers responded less to 13-month-olds' object initiatives than social initiatives. Notably, there is evidence that contingencies between infant gaze and maternal speech emerge quite early: Lavelli and Fogel (2005) found that mothers spoke more to infants as young as 1 month when their infant was looking at them, although the effect became stronger by 3 months, when infants' social gaze is more expressive and differentiated (see also Henning, Striano, & Lieven, 2005). The current data did not reveal change in mothers' contingent responses to their infant looking at them, suggesting that the contingency is well established by 4 months.

The results also indicate that the content of maternal speech is related to the infant's gaze. In particular, mothers' naming utterances were contingent on infants looking at the mother's face, or shifting gaze to objects. As infants got older, the timing of mothers' naming utterances became more tightly contingent on infants' gaze to objects. These results are consistent with previous reports: Penman, Cross, Milgrom-Friedman, and Meares (1983) found higher proportions of maternal speech about external referents when infants looked at objects at 3 and 6 months, and at 4 months, joint attention predicts lexical content in maternal vocalizations (Brousseau, Malcuit, Pomerleau, & Feider, 1996). In a cross-cultural study of

American and Japanese 3-month olds, the distribution of speech act types and referents differed based on infants' gaze targets, although the specific contingencies differed between cultures (Morikawa, Shand, & Kosawa, 1988). The current study thus confirms that mothers adapt their distribution of functional utterance types in response to infants' gaze and extends that evidence to object-naming utterances.

Maternal speech was also contingent on infants' manual activity. Although total speech rate did not differ in response to infants' object handling, mothers' production of object-naming utterances was contingent on infants' shifts from handling one object to multiple objects, and from multiple objects to one object. Because we were interested in the effect of developmental changes in infants' manual activity on their contingent caregiver speech, we further investigated age-related changes in infants' shifts in handling multiple objects. From 4 to 9 months, infants increased their object-handling shifts involving multiple objects, both per minute and as a proportion of total object-handling shifts. 4- and 6-month-olds who produced more multiple-object shifts heard more object-naming utterances. By 9 months, when most infants can effectively manipulate multiple objects, there were no significant individual differences in mothers' object naming, although the moment-to-moment contingencies between infants' multiple-object shifts and maternal speech remained robust. Thus, changes in infants' motor coordination are related to increased exposure to naming utterances.

These results build on those of de Barbaro et al. (2016), who described nonverbal dynamics of object play in a randomly selected subset of the sessions reported in the current study. Those results suggest that development of infants' ability to distribute their attention during play, as reflected by infants' "decoupling" of gaze and hands (i.e., looking at one object while handling another, or handling multiple objects simultaneously), has implications for social interactions. Notably, decoupling is linked to emerging social behaviors such as turn-taking and imitation (de Barbaro et al., 2013). The current results suggest that increased object naming represents another aspect of social routine maturation accompanying infants' advances in object exploration and attention. One possible interpretation is that changes in maternal speech occur because infants produce more of the actions that tend to elicit object naming. However, it is also possible that infants' advances in object handling influence maternal speech by changing mothers' perceptions of their infants. Mothers' perceptions of their 4- and 8-month-old infants' intentionality correlate positively with mothers' sensitive interaction style (Feldman & Reznick, 1996), and between 10 and 13 months, parents' perception of their infant as an individual positively predicted receptive vocabulary (Walle, 2016). Future research could therefore investigate the relationship between infants' object-handling development and caregivers' perceptions of their cognitive maturity.

Why might mothers disproportionately name objects when infants pick up or put down a second object? One possibility is that at these moments mothers simply perceive infants to be more attentive to the objects. However, it is also possible that these handling-switches are adaptive times for object naming because they highlight contrasts between objects. At these times infants are likely focusing attention on one object, but the other is still available and represented in working memory. Object names in utterances that occur at those times can be associated with features that distinguish the new focal object from the previous one. To clarify whether mothers used object-handling shifts specifically to highlight the new object, we computed the proportion of naming utterances contingent on one-to-multiple shifts that named each of the two objects. Out of these, 58% named the newly handled object, whereas 32% named the previously handled object. Highlighting new objects around shifts is consistent with evidence that comparisons facilitate children's learning of words and categories (Gentner, Loewenstein, & Hung, 2007; Gentner & Namy, 1999). By producing naming utterances when infants start or stop handling multiple objects, rather than during prolonged episodes of handling one or more objects, mothers increase the probability that the object label will be associated with contrastive features of one object versus the other, rather than irrelevant properties such as an object's location or motion. If caregivers regularly distribute naming utterances in such informative ways, it might not only help infants build associations between object labels and referents, but also guide their inferences about which object features to assign to nouns, which are otherwise highly indeterminate (Quine, 1960).

Mothers' utterances were also more frequent following infants' vocalizations. There is ample evidence that mothers spontaneously impose turn-taking rhythm in vocal response to infants' preverbal vocalizations (e.g., Papoušek & Papoušek, 1989). Such adult-imposed contingent input ostensibly socializes infants for discourse conventions that show culturally predictable temporal parameters (Stivers et al., 2009). Accordingly, by 4 months infants actively participate in vocal turn-taking (Stevenson, Ver Hoeve, Roach, & Leavitt, 1986). Notably, however, we found that infant vocalizations predicted mothers' utterances in general, but not naming utterances in particular. This suggests that contingencies between infant behaviors and maternal speech have differentiated functions: some might highlight the responsive nature of verbal interactions in general, whereas others might help infants associate maternal speech types or specific words with external referents.

We also observed different time courses of infants' and mothers' gaze and hand engagement with named objects, relative to the onset of naming utterances. At all ages, mothers' gaze and hand engagement both tended to peak around the onset of naming. The time course of infants' gaze and hand engagement relative to naming, however, showed a more complex developmental trajectory. Infants tended to look at the named object at all ages, but as they got older their looks peaked closer to the onset of naming. Infants handled named objects more by 6 months, and their handling increased after the onset of naming utterances. This increase suggests either that mothers name objects in anticipation of infants' activity, or that infants use mothers' naming utterances as a cue to sustain attention to objects. The latter possibility would create a positive feedback loop that might help infants maintain joint attention with caregivers; however, controlled experiments are necessary to determine whether caregiver speech affects infants' subsequent attention in naturalistic contexts. Similar to the present results, Yu and Smith (2012) found that older (18-month-old) infants held named objects more than did parents after naming events; however, in that study, infants' object holding peaked at the time of naming, while in the present study infants' object holding peaked later. Nonetheless, the time courses of modalities in both studies indicate that object naming does not simply overlap with infants' object gaze and handling, but is embedded in temporally structured sequences of co-exploration of potential referents. Often, for example, infants watched as their mother held, looked at, and named an object, and then retrieved the object themselves. Consistency in these sequences might help infants associate naming utterances with patterns of sensorimotor experience in order to ground the possible meanings of object names within those utterances.

Underlying the process of learning words, infants' multimodal experiences may contribute to the formation of neural networks that process language jointly with ongoing manual action. In adults, inferior frontal cortex is activated in language production as well as action production and recognition tasks (Hamzei et al., 2003). Language processing networks also integrate speech with manual activity in the form of co-speech gesture. In fMRI experiments, inferior frontal gyrus (Broca's area) showed greater metabolic response when co-speech gesture conveyed additional information than when it reiterated information present in speech (Dick, Mok, Beharelle, Goldin-Meadow, & Small, 2014). Co-speech gesture also elicited a stronger response than speech without gesture in cortical regions associated with language comprehension, both in adults (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009) and in children aged 8–11 years (Dick, Goldin-Meadow, Solodkin, & Small, 2012).

Relatedly, a form of multimodal experience with language and manual activity also seems to influence mothers' integration of speech and manual sensorimotor activity. Mothers showed differential N1 and P2 ERP components (relating to selective attention and discrimination) following mismatches between tactile cues and tactile-related words, whereas nonmothers did not show such responses, presumably because tactile-lexical associations were more salient to mothers, who spend more time explicating such associations with infants and toddlers (Tanaka, Fukushima, Okanoya, & Myowa-Yamakoshi, 2014). Thus, mothers' cortical networks might develop speech-action integration in an activity-dependent manner. We speculate that a similar process may occur in infants as a result of experience with associations between speech and motor and/or haptic experience.

Although it is difficult to determine whether changes in integration of sensorimotor and speech processing in infants are a result of experience or maturation, at least one study (Imada et al., 2006) shows that such integration does develop during the first year. Neonates showed MEG responses to speech sounds only in temporal auditory areas, but at 6 and 12 months activation was observed in both temporal areas and inferior frontal gyrus (Imada et al., 2006). However, it is not known whether such early-developing cross-modal processing is limited to direct, temporally precise links such as those between mouth movements and speech sounds, or whether it also encompasses less deterministic associations such as the social contingencies we observed.

Word-referent associative learning is unlikely to account for the entirety of early lexical development (Waxman & Gelman, 2009). Nevertheless, contingencies between infant actions and maternal speech suggest that associative learning plays a broader role in language acquisition. One way action contingencies could facilitate learning is if infants associate their own actions with expectations of informative input from caregivers. For example, Rochat, Querido, and Striano (1999) found that from 2 to 6 months infants learn to expect a turn-taking action pattern during "peek-a-boo" games–including the expectation that the infant's own action will elicit a particular kind of response, within a certain interval, from the adult. This illustrates that infants can learn to anticipate that their own actions will elicit specific communicative acts from adults. These expectations can then cue infants' attention in the service of word learning (Smith, Colunga, & Yoshida, 2010). Indeed, by 9 months, infants attend more to novel visual stimuli in the presence of novel words (Balaban & Waxman, 1997). Contingent responses to infants' self-generated actions are optimal for developing such expectations, both because infants seem predisposed to detect the causal force of their own actions (Bahrick & Watson, 1985; Bigelow, 1999; Watson, 1972), and because infants can generate the eliciting signal at times and contexts when they are receptive to input.

In addition to effects on word learning, social contingencies could potentially facilitate infants' understanding of others' attention or other mental states. Caregivers' responsiveness depends on their being attentive to the infant. Therefore, infants might become sensitized to caregivers' attention once they have detected the social contingencies that joint attention affords. If so, then participation in increasingly sophisticated social contingencies may form a foundation for infants to imbue their caregivers' attention and actions with meaning (Baldwin, 1991; Carpendale & Lewis, 2004, 2010; Rączaszek-Leonardi, Nomikou, & Rohlfing, 2013; Reddy, 2001).

Taken together, our results underscore that no unitary construct such as "responsiveness" can precisely capture the various ways caregivers act contingently on infant behavior. Instead, dyadic multimodal speech-and-sensorimotor contingencies provide a rich source of event-sequential information for infants. That information is available to young infants, but it changes as infants mature and acquire new behavioral capacities. At each point in development, an infant's social environment emerges differently from the set of contingent responses that are active at that time.

The current data set has several limitations. The infant behaviors that appeared to generate contingent responses might be correlated with unobserved behaviors that mediate the mother's response. It is also likely that combinations of infant behaviors, such as object handling while vocalizing or looking at the caregiver's face, interact to elicit contingent responses that cannot be captured by independent contingencies to individual infant behaviors. However, the current dataset did not have enough statistical power to test for all possible interactive effects of multiple infant behaviors on maternal speech.

The current study involved observations of play in dyads in an urban, primarily English-speaking community in the United States. Practices of playing and speaking with infants vary widely across cultures and across contexts within a culture (e.g., Altınkamış, Kern, & Sofu, 2014; Bornstein, Toda, Azuma, Tamis-LeMonda, & Ogino, 1990). Therefore, the observed patterns cannot be assumed to generalize to other populations. However, in studies of early mother-infant interaction across several cultures, caregivers have been observed to respond contingently to infants' vocalizations, even in cultures with low overall levels of infant-directed speech and toy play (Bornstein, Putnick, Cote, Haynes, & Suwalsky, 2015; Kärtner et al., 2008). Thus, infants' self-generated activity might play a role in many cultures in driving the micro-behavioral structure of their language environment by shaping caregivers' infant-directed speech across the first year. However, it is likely that the specific behaviors that contribute to this structure differ somewhat from culture to culture; for example, Fogel, Toda, and Kawai (1988) found that whereas American mothers verbalized in response to their 3-month-olds' vocalizations and gaze, Japanese mothers responded with facial expressions rather than speech, and did not respond contingently to their infants' vocalizations.

The analyses presented in this paper are exploratory in nature and would benefit from confirmatory replication. Nevertheless, these results show that the occurrence and object-naming content of maternal speech are contingent on infants' gaze and hand actions, and that the precise pattern of contingencies, and of infants' object exploration, evolves from 4 to 9 months. Future research should investigate contingent responsiveness at a similar level of granularity in additional contexts and cultures, and measure both the theoretical learnability and infants' actual learning of the information made available in these social interactions.

Funding

This material is based upon work supported by the National Science Foundation (NSF) Graduate Research Fellowship under Grant No. DGE-1144086, NSF grant SES-0527756 to G. Deák and J. Triesch, and by the Temporal Dynamics of Learning Center (NSF SMA-1005256).

Chapter 1, in full, is a reprint of the material as it appears in: Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology.* The dissertation author was the primary investigator and author of this paper.

Figure 1.1. Example of frames from three video angles taken from home session at six months. Objects can be seen on tray and in wells on sides of tray. Because observations were recorded in participants' homes, sessions vary in background visual scene characteristics.

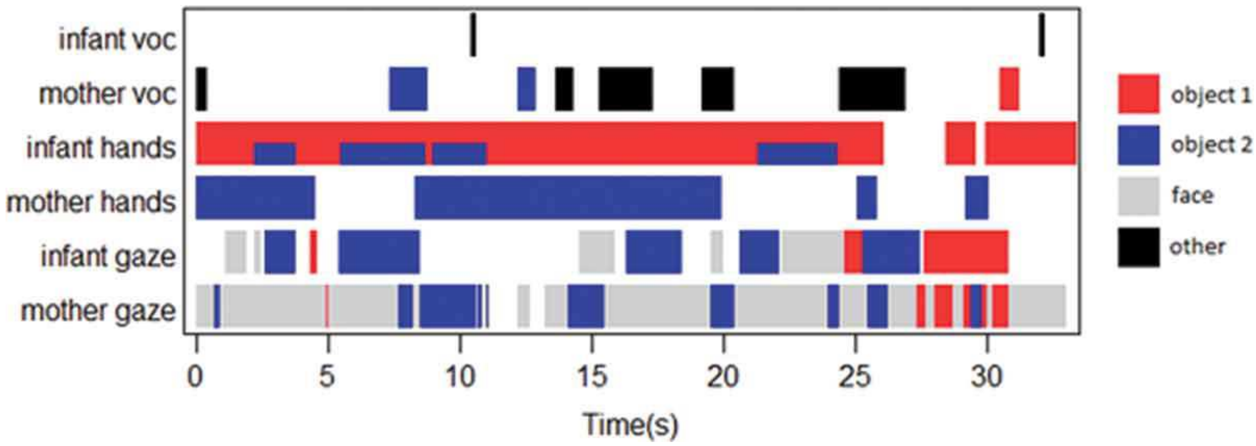

Figure 1.2. Example time series of data from part of one session. First row: Bars represent times when the infant vocalized. Second row: Bars represent times when the mother was speaking, and colors indicate naming utterances for the respective objects. Remaining rows: Colored bars represent times when the modality was directed to a specific target (objects or partner's face). Portions of bars split into two colors represent periods of simultaneous contact with multiple objects.

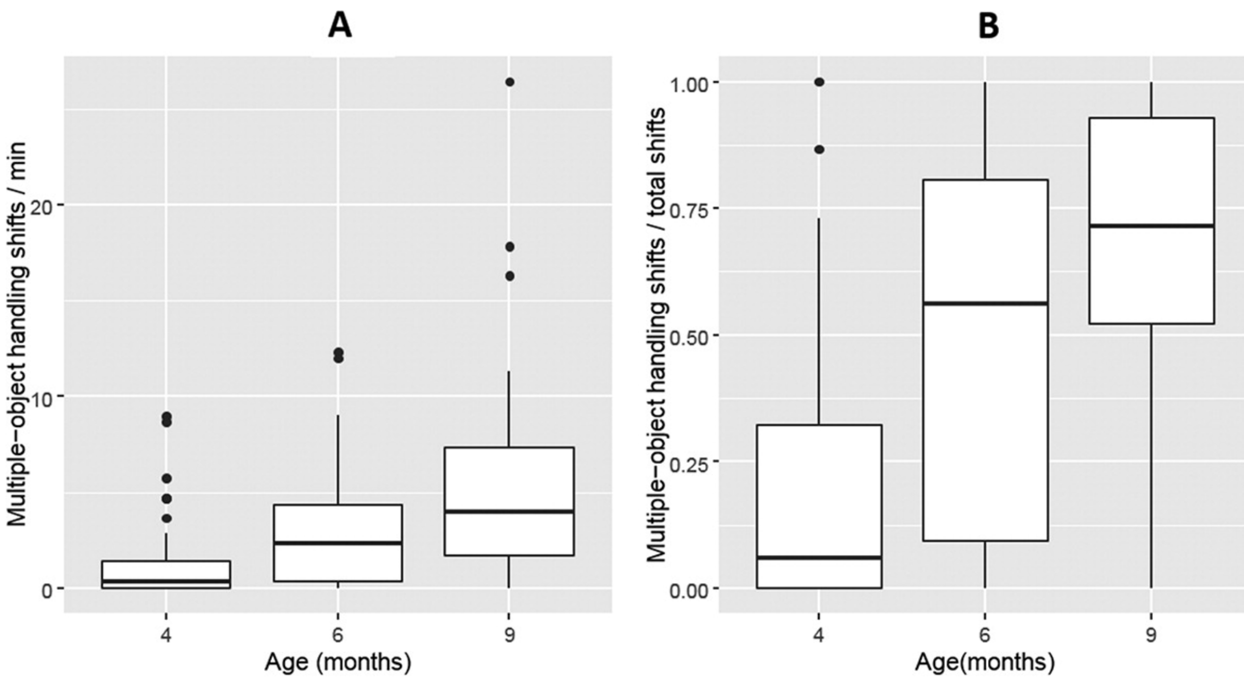

Figure 1.3. Increase in multiple-object handling shifts with age. A: multiple-object handling shifts per minute at each age. B: multiple-object handling shifts as a proportion of total shifts at each age.

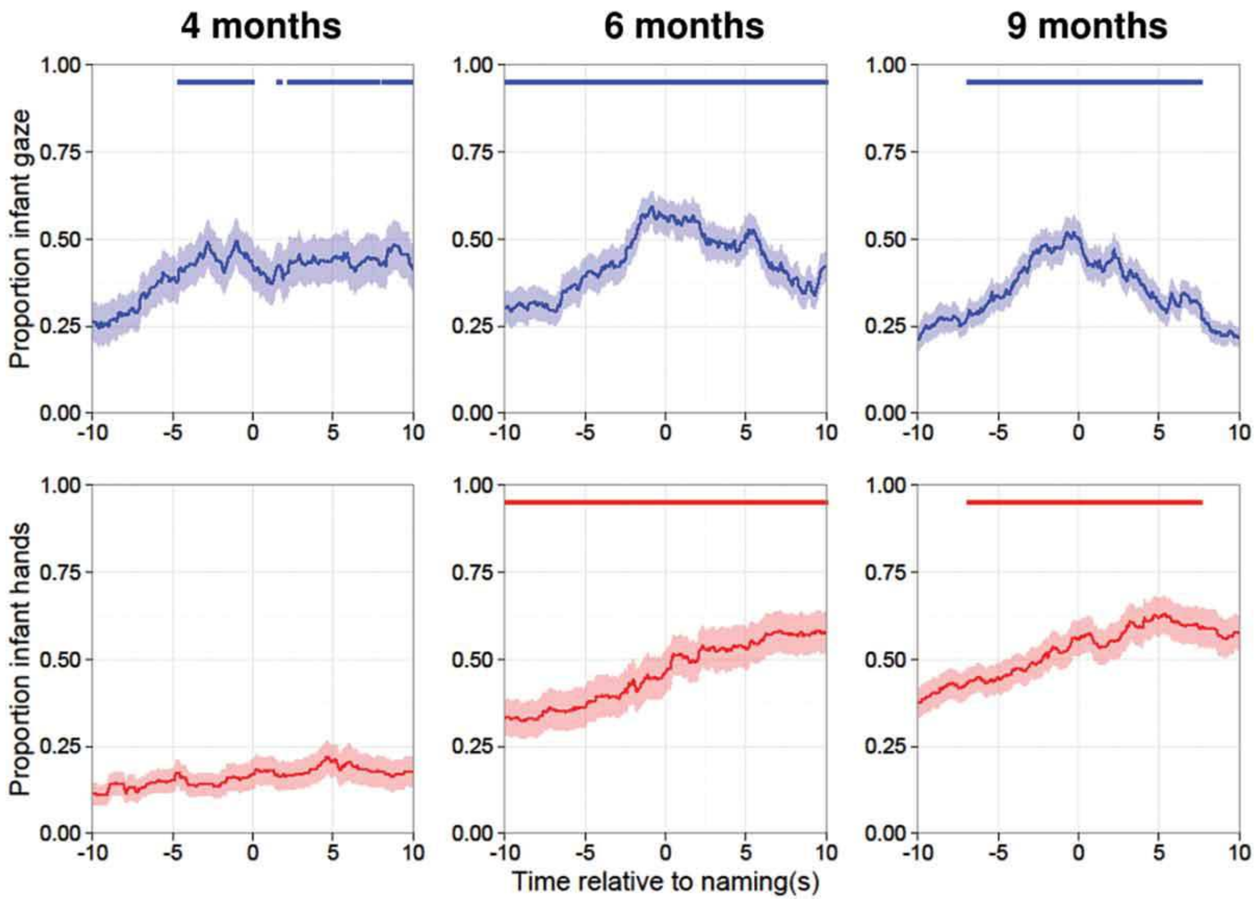

Figure 1.4. Infant behavior around naming utterances at each month. Lines represent average proportions of instances in which, at that time, infants' hands or gaze were focused on the target object, time-locked to onsets of naming utterances. Shaded regions represent the standard error of the mean. Bars at the top of graphs represent times when the naming-association index was greater than chance at p < .05.

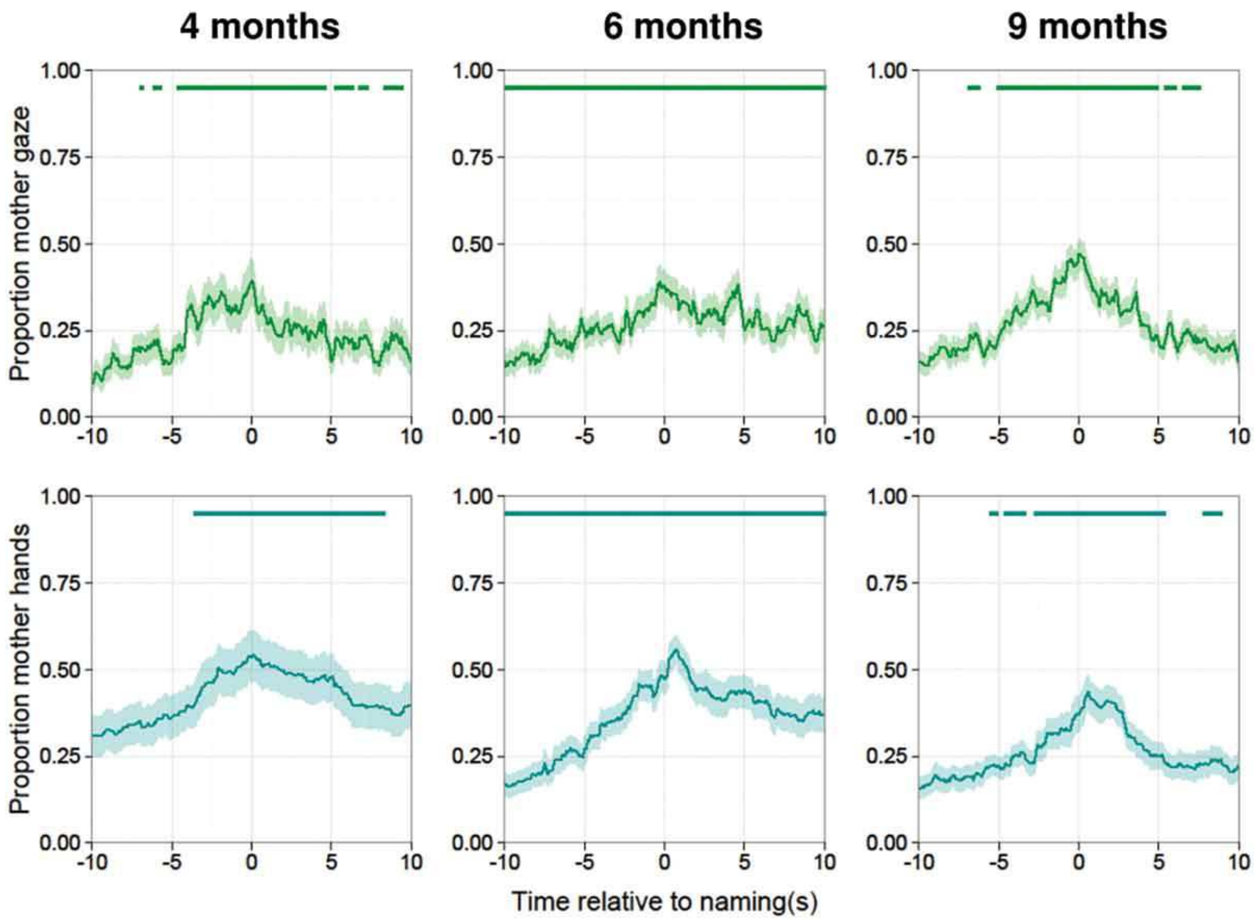

Figure 1.5. Mother behavior around naming utterances at each month. Lines represent average proportions of instances in which, at that time, mothers' hands or gaze were focused on the target object, time-locked to onsets of her naming utterances. Shaded regions represent the standard error of the mean. Bars at the top of graphs represent times when the naming-association index was greater than chance at $p < .05$.

Table 1.1. Behavioral coding scheme

Behavior	Definition
Infant Gaze	Target: one or more of the toys, or mother's face, or other (any location except one of the toys or the mother's face, e.g., the tray, walls, furniture, or extraneous objects)
Mother Gaze	Target: one or more of the toys, or infant's face, or other (any location except one of the toys or the infant's face, e.g., the tray, walls, furniture, or extraneous objects)
Infant Hands	Empty, or contacting one or more of the toys (ignoring gaps in contact of < 2 s). Mouthing the toys was also included, but was almost exclusively accompanied by handling.
Mother Hands	Empty, or contacting one or more of the toys (ignoring gaps in contact of < 2 s)
Infant Vocalization	Any vocal sound, excluding cries, grunts, burps, and other organic sounds
Mother Speech	All verbalizations, excluding sound effects and other non-linguistic vocalizations. Utterances are defined as bouts of speech separated by ≥ 200 ms of silence. Each utterance was further categorized as naming (e.g., *"You got the ball"*) or nonnaming (e.g. *"What is that?"*).

Table 1.2. Average number of naming and total utterances per minute at each month. SD in parentheses.

	Total utterances/min	Naming utterances/min
4 months	20.5 (4.6)	1.5 (1.5)
6 months	21.5 (5.3)	2.3 (1.4)
9 months	19.1 (4.9) *	3.1 (1.9) *

Note. * denotes significant change from previous month, $p < .025$.

Table 1.3. Mean proportion of time infants or mothers were looking at, or handling, at least one object.

	Infant gaze	Infant hands	Mother gaze	Mother hands
4 months	.73 (.13)	.43 (.29)	.36 (.12)	.68 (.19)
6 months	.65 (.12) *	.74 (.25) *	.39 (.11)	.47 (.18) *
9 months	.54 (.15) *	.83 (.20)	.39 (.12)	.39 (.18) *

Note. SD in parentheses. * denotes significant change from previous month, $p < .025$.

Table 1.4. Poisson regressions for the contingencies between infant gaze targets and mothers' utterances.

	Total Utterances			Naming Utterances		
Gaze Target	Relative Rate	β	p	Relative Rate	β	p
Object vs. None	0.87 *	−.144	< .001	1.91 *	.646	< .001
Face vs. None	1.18 *	.164	< .001	1.46 *	.380	< .010

Note. Exponentiated coefficients, which are interpreted as the ratio of the rate of responses between the two infant gaze targets, are reported as Relative Rate. * $p < 0.5$.

Table 1.5. Poisson regressions for the contingencies between infant actions and mothers' utterances.

	Total Utterances			Naming Utterances		
Infant Behavior	Relative Rate	β	p	Relative Rate	β	p
Gaze: Object → Other	1.01	.008	.800	1.23	.208	< .05
Gaze: Other → Object	1.00	.000	.990	1.29 *	.260	< .005
Gaze: Object → Face	1.19 *	.176	<.001	1.21	.194	.200
Gaze: Face → Object	1.05	.047	.370	1.49 *	.399	< .005
Gaze: Face → Other	1.14 *	.131	< .001	0.50 *	−.681	< .001
Gaze: Other → Face	1.34 *	.294	< .001	1.03	.034	.810
Gaze: Object → Object	0.99	−.014	.780	1.14	.132	.340
Hand: None → One	0.90	−.104	.110	1.23	.211	.230
Hand: One → None	1.07	.070	.230	0.91	−.084	.670
Hand: One → Multiple	0.94	−.061	.560	1.43 *	.361	< .010
Hand: Multiple → One	0.99	.006	.930	1.47 *	.385	< .010
Hand: Object → Object	0.97	−.032	.740	1.34	.294	.190
Infant Vocalization	1.18 *	.163	< .010	1.10	.174	.580

Note. For each infant behavior, we compared maternal verbalizations in the 2 s after each infant behavior with other periods. The regression models test differences in rates of maternal response between these two types of intervals. Exponentiated coefficients, which are interpreted as the ratio of the rate of responses between the two types of periods, are reported as Relative Rate. * adjusted $p < 0.05$.

Table 1.6. Poisson regressions for the contingencies between infant hand states and mothers' utterances

	Total Utterances			Naming Utterances		
Objects held	Relative Rate	β	*p*	Relative Rate	β	*p*
One vs. None	1.02	.022	.46	1.01	.009	.92
Multiple vs. None	1.00	−.004	.93	1.04	.044	.76

Note. Exponentiated coefficients, which are interpreted as the ratio of the rate of responses between the two types of periods, are reported as Relative Rate. $^*p < 0.05$.

Table 1.7. Partial correlations between rate of infants' multiple-object handling shifts and maternal speech rates, controlling for infant's age in days

Age	Utterances/min	Labels/min
4 months	−.03	.42*
6 months	−.01	.40*
9 months	−.30	−.15

Note. $^*p < .017$.

References

Altınkamış, N. F., Kern, S., & Sofu, H. (2014). When context matters more than language: Verb or noun in French and Turkish caregiver speech. *First Language, 34*(6), 537-550.

Bahrick, L. E., & Watson, J. S. (1985). Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy. *Developmental Psychology, 21*(6), 963-973.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology, 64*, 3-26.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*(5), 874-890.

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.

Bigelow, A. E. (1999). Infants' sensitivity to imperfect contingency in social interaction. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 137-154). Mahwah, NJ: Lawrence Erlbaum Associates.

Bornstein, M. H., Putnick, D. L., Cote, L. R., Haynes, O. M., & Suwalsky, J. T. D. (2015). Mother-Infant Contingent Vocalizations in 11 Countries. *Psychological Science, 26*(8), 1272-1284.

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C.-S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology, 44*(3), 867-874.

Bornstein, M. H., Toda, S., Azuma, H., Tamis-LeMonda, C., & Ogino, M. (1990). Mother and infant activity and interaction in Japan and in the United States: II. A comparative microanalysis of naturalistic exchanges focused on the organisation of infant attention. *International Journal of Behavioral Development, 13*(3), 289-308.

Brousseau, L., Malcuit, G., Pomerleau, A., & Feider, H. (1996). Relations between lexical-temporal features in mothers' speech and infants' interactive behaviours. *First Language, 16*(46), 41-59.

Carpendale, J. I., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 27*(1), 79-96.

Carpendale, J. I., & Lewis, C. (2010). The development of social understanding. In L. Liben, U. Müller, & R. M Lerner (Eds.), *Handbook of child psychology and developmental science: Cognitive processes* (vol. 2, 7th ed., pp. 381–424). Hoboken, NJ: John Wiley & Sons, Inc.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213-220.

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment, 91*(2), 121-136.

de Barbaro, K., Johnson, C. M., & Deák, G. O. (2013). Twelve-month “social revolution” emerges from mother-infant sensorimotor coordination: A longitudinal investigation. *Human Development, 56*, 223-248.

de Barbaro, K., Johnson, C. M., Forster, D., & Deák, G. O. (2016). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother-infant-object interactions. *Child Development, 87*(2), 494-512.

Deák, G. O. (2015). When and where do infants follow gaze? *Proceedings of International Conferences on Development and Learning and on Epigenetic Robotics*, Providence, RI, USA.

Deák, 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*(2), 270-281.

Deák, G. O., Triesch, J., Krasno, A., de Barbaro, K., & Robledo, M. (2013). Learning to share: The emergence of joint attention in human infancy. In B. Kar (Ed.), *Cognition and brain development: Converging evidence from various methodologies* (pp. 173-210). Washington, DC: American Psychological Association.

Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., & Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping, 30*(11), 3509-3526.

Dick, A. S., Goldin-Meadow, S., Solodkin, A., & Small, S. L. (2012). Gesture in the developing brain. *Developmental Science, 15*(2), 165-180.

Dick, A. S., Mok, E. H., Beharelle, A. R., Goldin-Meadow, S., & Small, S. L. (2014). Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Human Brain Mapping, 35*(3), 900-917.

Dimitrova, N., Özçalışkan, Ş., & Adamson, L. B. (2016). Parents’ translations of child gesture facilitate word learning in children with autism, Down syndrome and typical development. *Journal of Autism and Developmental Disorders, 46*(1), 221-231.

Ellis, E. M., Gonzalez, M. R., & Deák, G. O. (2014). Visual prediction in infancy: What is the association with later vocabulary? *Language Learning and Development, 10*(1), 36-50.

Fagard, J., & Lockman, J. J. (2010). Change in imitation for object manipulation between 10 and 12 months of age. *Developmental Psychobiology, 52*(1), 90-99.

Feldman, R., & Reznick, J. S. (1996). Maternal perception of infant intentionality at 4 and 8 months. *Infant Behavior and Development, 19*(4), 483-496.

Fogel, A., Toda, S., & Kawai, M. (1988). Mother-infant face-to-face interaction in Japan and the United States: A laboratory comparison using 3-month-old infants. *Developmental Psychology, 24*(3), 398-406.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Erlbaum.

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development, 8*(3), 285-307.

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development, 14*(4), 487-513.

Gogate, L., Maganti, M., & Bahrick, L. E. (2015). Cross-cultural evidence for multimodal motherese: Asian Indian mothers' adaptive use of synchronous words and gestures. *Journal of Experimental Child Psychology, 129*, 110-126.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development, 71*(4), 878-894.

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6-to 8-monthold infants and learning of word-object relations. *Infancy: The Official Journal of the International Society on Infant Studies, 9*(3), 259-288.

Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science, 10*(6), 778-785.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19*(5), 515-523.

Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic Object-Directed vocalizations indicate a readiness to learn. *Infancy: The Official Journal of the International Society on Infant Studies, 15*(4), 362-391.

Gros-Louis, J., West, M. J., & King, A. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy: The Official Journal of the International Society on Infant Studies, 19*(4), 385-408.

Hamzei, F., Rijntjes, M., Dettmers, C., Glauche, V., Weiller, C., & Büchel, C. (2003). The human action recognition system and its relationship to Broca's area: An fMRI study. *Neuroimage, 19*(3), 637-644.

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

Henning, A., Striano, T., & Lieven, E. V. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development, 28*(4), 519-536.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*(5), 1368-1378.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *Neuroreport, 17*(10), 957-962.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language, 37*(2), 229-261.

Iverson, J. M., & Fagan, M. K. (2004). Infant vocal-motor coordination: Precursor to the gesture-speech system? *Child Development, 75*(4), 1053-1066.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental Science, 17*(3), 388-395.

Kärtner, J., Keller, H., Lamm, B., Abels, M., Yovsi, R. D., Chaudhary, N., & Su, Y. (2008). Similarities and differences in contingency experiences of 3-month-olds across sociocultural contexts. *Infant Behavior and Development, 31*(3), 488-500.

Kaye, K., & Fogel, A. (1980). The temporal structure of face-to-face communication between mothers and infants. *Developmental Psychology, 16*(5), 454-464.

Kaye, K., & Wells, A. J. (1980). Mothers' jiggling and the burst-pause pattern in neonatal feeding. *Infant Behavior and Development, 3*, 29-46.

Kochanska, G., & Aksan, N. (2004). Development of mutual responsiveness between parents and their young children. *Child Development, 75*(6), 1657-1676.

Lavelli, M., & Fogel, A. (2005). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: The 2-month transition. *Developmental Psychology, 41*(1), 265–280.

Lloyd, C. A., & Masur, E. F. (2014). Infant behaviors influence mothers' provision of responsive and directive behaviors. *Infant Behavior and Development, 37*(3), 276-285.

Masur, E. F. (1982). Mothers' responses to infants' object-related gestures: Influences on lexical development. *Journal of Child Language, 9*(1), 23-30.

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word–object relations. *Infancy: The Official Journal of the International Society on Infant Studies, 13*(2), 172-184.

McGillion, M. L., Herbert, J. S., Pine, J. M., Keren-Portnoy, T., Vihman, M. M., & Matthews, D. E. (2013). Supporting early vocabulary development: What Sort of Responsiveness Matters? *IEEE Transactions on Autonomous Mental Development, 5*(3), 240-248.

Meyer, M., Hard, B., Brand, R. J., McGarvey, M., & Baldwin, D. A. (2011). Acoustic packaging: Maternal speech and action synchrony. *IEEE Transactions on Autonomous Mental Development, 3*(2), 154-162.

Morikawa, H., Shand, N., & Kosawa, Y. (1988). Maternal speech to prelingual infants in Japan and the United States: Relationships among functions, forms and referents. *Journal of Child Language, 15*(2), 237-256.

Olson, J., & Masur, E. F. (2011). Infants' gestures influence mothers' provision of object, action and internal state labels. *Journal of Child Language, 38*(5), 1028-1054.

Olson, J., & Masur, E. F. (2013). Mothers respond differently to infants' gestural versus nongestural communicative bids. *First Language, 33*(4), 372-387.

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language, 9*(6), 137-157.

Penman, R., Cross, T., Milgrom-Friedman, J., & Meares, R. (1983). Mothers' speech to prelingual infants: A pragmatic analysis. *Journal of Child Language, 10*(1), 17-34.

Quine, W. V. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Rączaszek-Leonardi, J., Nomikou, I., & Rohlfing, K. J. (2013). Young children's dialogical actions: The beginnings of purposeful intersubjectivity. *IEEE Transactions Autonomic Mental Developments, 5*, 210-221.

Reddy, V. (2001). Mind knowledge in infancy: Understanding attention and intention in the first year. In G. J. Bremner, & A. Fogel (Eds.), *Blackwell handbook of infancy research* (pp. 241-264). Oxford, UK: Blackwell.

Rochat, P. (1989). Object manipulation and exploration in 2-to 5-month-old infants. *Developmental Psychology, 25*(6), 871-884.

Rochat, P., Querido, J. G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology, 35*(4), 950-957.

Rollins, P. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics, 24*, 221-234.

Ruff, H. A. (1984). Infants' manipulative exploration of objects: Effects of age and object characteristics. *Developmental Psychology, 20*(1), 9-20.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696-735.

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science, 34*, 1287-1314.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language, 4*(1), 1-22.

Stevenson, M. B., Ver Hoeve, J. N., Roach, M. A., & Leavitt, L. A. (1986). The beginning of conversation: Early patterns of mother-infant vocal responsiveness. *Infant Behavior and Development, 9*(4), 423–440.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106*(26), 10587-10592.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*(3), 748-767.

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

Tamis-LeMonda, C. S., Kuchirko, Y., & Tafuro, L. (2013). From action to interaction: Infant object exploration and Mothers' Contingent Responsiveness. *IEEE Transactions on Autonomous Mental Development, 5*(3), 202-209.

Tanaka, Y., Fukushima, H., Okanoya, K., & Myowa-Yamakoshi, M. (2014). Mothers' multimodal information processing is modulated by multimodal interactions with their infants. *Scientific Reports, 4*, 1-7.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology, 32*(3), 492-504.

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

Van Egeren, L. A., Barratt, M. S., & Roach, M. A. (2001). Mother-infant responsiveness: Timing, mutual regulation, and interactional context. *Developmental Psychology, 37*(5), 684–697.

von Hofsten, C. (1991). Structuring of early reaching movements: A longitudinal study. *Journal of Motor Behavior, 23*(4), 280-292.

Walle, E. A. (2016). Infant social development across the transition from crawling to walking. *Frontiers in Psychology, 7*, 1-10.

Watson, J. S. (1972). Smiling, cooing, and "the game." *Merrill-Palmer Quarterly of Behavior and Development, 18*(4), 323-339.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences, 13*(6), 258-263.

Wu, Z., & Gros-Louis, J. (2014a). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *Journal of Child Language, 41*(4), 1-24.

Wu, Z., & Gros-Louis, J. (2014b). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language, 34*(1), 72-90.

Yoshida, H., & Burling, J. M. (2013). An embodied perspective of early language exposure. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Berlin, Germany (pp. 1635–1640).

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

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *Plos One, 8*(11), e79659.

Zumbahlen, M. R. (1997). The role of infant locomotor onset in shaping mother-infant communication (Doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois.

CHAPTER 2: Maternal discourse continuity and infants' actions organize 12-month-olds' language exposure during object play

Abstract

Infant language learning depends on the distribution of co-occurrences within language–between words and other words–and between language content and events in the world. Yet infant-directed speech is not limited to words that refer to perceivable objects and actions. Rather, caregivers' utterances contain a range of syntactic forms and expressions with diverse attentional, regulatory, social, and referential functions. We conducted a distributional analysis of linguistic content types at the utterance level, and demonstrated that a wide range of content types in maternal speech can be distinguished by their distribution in sequences of utterances and by their patterns of co-occurrence with infants' actions. We observed free-play sessions of 38 12-month-old infants and their mothers, annotated maternal utterances for 10 content types, and coded infants' gaze target and object handling. Results show that all content types tended to repeat in consecutive utterances, whereas preferred transitions between different content types reflected sequences from attention-capturing to directing and then descriptive utterances. Specific content types were associated with infants' engagement with objects (declaratives, descriptions, object names), with disengagement from objects (talk about attention, infant's name), and with infants' gaze at the mother (affirmations). We discuss how structured discourse might facilitate language acquisition by making speech input more predictable and/or by providing clues about high-level form-function mappings.

Introduction

Infant-directed discourse

From the earliest ages, infant-directed speech (IDS) constitutes a rich discourse including interconnected utterances of varied syntactic and semantic content. These utterances include but are not limited to object-naming, descriptive comments, attention-directing, imperatives, questions, and stereotyped routines such as greetings and games. Caregivers' varied utterance types have interactive functions including regulating infants' affective state and attention (e.g., Fernald, 1993; Ninio & Snow, 1996; Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990). Distinct functions can be at least partially discriminated using low-level acoustic features (Fernald, 1989) and by linguistically naive observers (Bryant & Barrett, 2007). Yet, most research has focused on how infants' environment supports learning object words (e.g., Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Suanda, Smith, & Yu, 2016; Trueswell, Lin, Armstrong, Cartmill, Goldin-Meadow, & Gleitman, 2016). The diversity of utterance types in natural discourse, however, exacerbates the inherent difficulty of mapping words to meanings: for example, there are frequent opportunities for spurious associations not only between words and objects, but between words or phrases and varied social purposes ranging from action-imperatives such as “don't touch that!” to non-referential utterances such as “Hi, baby!”

Despite this potential difficulty, the amount of complex and diverse speech heard by infants predicts their vocabulary growth (Akhtar, Dunham, & Dunham, 1991; Bornstein, Haynes, & Painter, 1998; Hart & Risley, 1995; Hoff, 2003; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Pan, Rowe, Singer, & Snow, 2005). This appears to be true even though a richer discourse environment would potentially complicate the task of a learner who relies on co-occurrences of words and referents to construct a lexicon (Yu & Ballard, 2007; Yu & Smith, 2012). Perhaps, however, infants also start to detect and discriminate patterns of occurrence of different discourse content types, even as they are learning their first object words. Different functions of IDS utterances such as attention-getting, prohibition, encouragement, and comforting might be indicated not only by prosodic differences (e.g., Fernald, 1989; Katz, Cohn, & Moore, 1996), but also by the frequency of utterance contexts or frames that associate words with phrase structures and syntactic roles (Mintz, 2003). Such findings underscore the question: how does syntactically and pragmatically diverse discourse shape not only infants' linguistic environment, but also their language acquisition? One possibility is that discourse diversity is merely incidental: parents simply produce complex language out of habit, but the diversity of words and discourse-functions does not directly affect early language acquisition. A second possibility is that caregivers use IDS to regulate infants' attention, to make word-learning opportunities more predictable and therefore more effective (Benitez & Smith, 2012). A third possibility, not mutually exclusive with the second, is that complex discourse is itself an input for infant learning–not, perhaps, of word-object mappings, but of the different types of utterances that can accompany various social-interactive events. That is, perhaps a burgeoning awareness of discourse types, working in tandem with infants' growing social-event knowledge (Cannon & Woodward, 2012; Tomasello, 1992) allows infants to classify or even predict adults' communicative actions during social interactions. However, the relevant statistical patterns characterizing infant-directed discourse–its structure over time and correspondences with other observable events–are poorly characterized. Thus, it remains unclear what role the diversity of varied discourse content plays in language acquisition.

Discourse continuity

Conversational analysis studies show that successive speakers' turns show well-structured, predictable sequences of functional content types, such as question-answer pairs (Ninio, Snow, Pan, & Rollins, 1994; Schegloff, 2007). Conversations between adults and young children show developmental trajectories toward a variety of such functional conversational sequences, including question-answering (Gallagher, 1981) and repair (Tomasello, Conti-Ramsden, & Ewert, 1990). However, relatively few studies have focused on discourse structure in parents' speech to prelinguistic infants, even though infant-directed discourse is presumably the developmental context for the emergence of early conversational patterning. These studies are reviewed below.

One proposed mechanism by which infant-directed discourse could facilitate learning is continuity of reference (Messer, 1980). Frank, Tenenbaum, and Fernald (2013) identified object references in a corpus of IDS and found that mothers' consecutive utterances were more likely to refer to the same object than to other objects that were present. Moreover, a computational model was able to use this continuity of reference to more accurately infer which object was the topic of each utterance. However, sequential continuity might also characterize discourse functions other than object reference. For instance, if utterances that either name objects, describe objects, or refer to actions each tend to follow previous utterances of the same type, this continuity might be useful to a learner tasked with learning a lexicon of nouns, adjectives, and verbs, respectively. That is, sequences of utterances of the same type might provide not only repetition of topical content words and continuity of reference, but also correlated repetition of syntactic forms. Repetition of these parallel levels of content, function, and context variables might facilitate semantic and pragmatic mapping.

Another form of discourse continuity might be available to infants: short, formulaic sequences of different utterances that tend to occur in predictable order. For example, utterances that capture the infant's attention might be followed by utterances that direct their attention to an object, and then by utterances that name the object, and finally by elaborating utterances that describe or evaluate it. By this hypothesis, IDS would be organized around stereotyped sequences (or "paths") of content-types ("states"), with different probabilities associated with all of the possible transitions from one content type to the next. For example, Rohde and Frank (2014) found that mothers produced more utterance-final object names early within runs of same-topic utterances; conversely, they produced more pronouns later in runs. However, the aforementioned studies analyzed only object-naming and/or object-referring utterances, and therefore did not address whether continuity over time or predictable transitions characterize other types of infant-directed utterances. Therefore, in the current study we recorded transition probabilities among different types of infant-directed utterances.

Contextual correlates of discourse types

In addition to sequential continuity from one utterance to the next, the discourse structure of IDS could facilitate language learning if different types of utterances appear preferentially in specific contexts. Relatively few studies have examined the contextual correlates of utterances other than object-referring utterances. Moreover, studies on the nonlinguistic contexts of object-referring utterances often measure only the degree to which the infant and/or parent's gaze or manual actions are directed to the particular object being spoken about (e.g. Frank et al., 2013; Yu & Smith, 2012). However, independent of object-specific associations, object-naming as a discourse function occurs coupled with infants' general object-directed gaze and developmentally advanced manual object exploration (Chang, de Barbaro, & Deák, 2016), and analogous coupling might exist between other discourse functions and other non-verbal actions or states within social interactions. If infants can learn to associate utterance-types with the contexts they most often occur in, this could form the basis for infants to develop expectations that are well coordinated with caregiver actions (Rochat, Querido, & Striano, 1999). These contexts can be studied at a macro level (e.g. the activity being engaged in and the participants present), or at a micro level: that is, the moment-to-moment actions and attention-states of participants within an activity. Several studies have investigated the frequency of various types of linguistic content in IDS across different activities: book-reading and object play (Choi, 2000; Gros-Louis, West, & King, 2016; Tamis-LeMonda, Song, Leavell, Kahana-Kalman, & Yoshikawa, 2012; Yont, Snow, & Vernon-Feagans, 2003), as well as eating and dressing (Hoff-Ginsberg, 1991; Pan, Perlmann, & Snow, 2000). These efforts have revealed significant variability in the proportions of types of maternal utterances, such as imitations, directives, questions, and naming, produced in these varied social contexts.

Other studies have investigated caregivers' use of different types of utterance content in response to infants' moment-to-moment actions within an activity context. One series of studies found that mothers follow infants' social initiatives with responsive utterances, whereas they tend to ignore or redirect infants' object initiatives, and to redirect infants when they disengage from toys (Masur, Flynn, & Lloyd, 2013; Lloyd & Masur, 2014). Similarly, Tamis-LeMonda, Kuchirko, and Tafuro (2013) found that mothers contingently respond to 14-month-old infants' object exploration actions by handling the objects themselves and by producing more referential language. Another study found that mothers respond contingently to infants' object exploration and play with utterances of various types including questions, descriptions, affirmations, imitations, and prompts (Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008). Beyond social and object play, infants' walking and crawling influences parents' use of prohibitions and action directives (Karasik, Tamis-LeMonda, & Adolph, 2014; Zumbahlen, 1997), and infants' pointing encourages parents to ask questions (Wu & Gros-Louis, 2014). Collectively, these studies show that caregivers' utterance content types depend on infants' actions; however, each has limitations. In several studies, infants' activity was judged subjectively by different criteria, making comparisons across studies difficult. In others, parental utterances were only counted if they were judged by human observers to be appropriate responses to the infants' actions, or utterance types were defined based on how they related to infant actions. These designs do not allow for comparisons between observed utterance-action co-occurrence rates and chance, and are difficult to interpret in terms of informativeness to naive learners. Therefore, in the current study we comprehensively documented both infants' action primitives, specifically gaze target and object handling, and mothers' speech, which was then classified for different types of discourse content.

Utterance types

Several utterance content types have been studied in IDS and infant language, and the evidence points to predictable developmental trajectories. At the lexical level, researchers have been interested in infants' first-acquired words. These include many "routines" (e.g. thank you, all gone, hello), and an increasing proportion of object nouns in the first months of production (Caselli et al., 1995). As early as 6 months of age, infants show evidence of having formed associations with common nouns (Bergelson & Swingley, 2012) and their own name (Mandel, Jusczyk, & Pisoni, 1995).

Other researchers have examined which types of discourse content in speech input correlate with developmental outcomes. A consistent finding has been that a higher proportion of responsive utterances –typically defined as utterances that refer to the infants' focus of attention, or non-referential affirmations of infants' communicative actions –correlate with positive outcomes in language development (Baumwell, Tamis-LeMonda, & Bornstein, 1997; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Masur, Flynn, & Eichorst, 2005; Rollins, 2003; Tamis-LeMonda, Bornstein, & Baumwell, 2001). In contrast, a higher proportion of directive utterances is unrelated to (Baumwell et al., 1997; Carpenter et al., 1998) or negatively correlated with vocabulary growth (Hughes, Dote-Kwan, & Dolendo, 1999; Nelson, 1973; Tomasello & Todd, 1983), perhaps because directives disrupt infants' attention (Akhtar et al., 1991; Lloyd & Masur, 2014; Masur et al., 2005).

We therefore sought to code maternal utterances in a way that captured the salient lexical, syntactic, and topical content types reviewed above. Standardized taxonomies of communicative acts have been proposed (Ninio et al., 1994; Searle, 1976); however, these schemes either did not cover the desired range of levels of description or did not match well with the set of content types that were developmentally and contextually appropriate to the current study. Therefore, we chose to create a coding system suitable for the population and context represented in the current study. We chose to tag infant-directed utterances at the syntactic level as declarative, imperative, or questions; and at the topical level as about attention, infant actions, or object-properties (i.e., descriptions). We also tagged utterances for several other lexical or functional content types with specific relevance to infant language: infant's name, object names, social routines, and affirmations. Similar approaches to coding IDS have been used in studies of infant-parent interaction (Bornstein et al., 2008; Gros-Louis, West, Goldstein, & King, 2006; Toda, Fogel, & Kawai, 1990).

Infant activity

While previous studies have extensively documented parents' tendency to refer to specific objects of infant visual attention (e.g. Yu & Smith, 2012), we chose to analyze more general categories of gaze: at toys, at the mother's face, or other. We expected that mothers would tailor their speech content to the infant's focus of attention (Frank et al., 2013), preferentially using speech with referential, interpersonal, or attention-directing content, depending on infants' gaze state. We also expected that simple versus complex object handling would elicit different speech patterns. Chang et al. (2016) and de Barbaro, Johnson, Forster, & Deák (2016) showed that infants' manipulation of two objects is related to their cognitive and motor maturity, and that bimanual manipulation elicited more object naming by mothers of infants between 4 and 9 months. Handling of multiple objects might also be related to subjective measures of infant manual behavior such as the distinction between object exploration and play (Bornstein et al., 1998). Object handling was therefore classified as involving no objects, one object, or two objects.

Current study

In the current study, we observed mother-infant dyads in a free-play in-home interaction with toys. We transcribed the mothers' speech and tagged utterances as containing declaratives, imperatives, questions, talk about action, talk about attention, object descriptions, affirmations, social routines, object names, and the infant's name. We also annotated infants' gaze at toys and at the mother's face, and infants' handling of the toys. The first set of analyses describes the structure of mothers' discourse over time. We measured the repetitiveness of all content types, and report transitions between the types that occur at above-chance rates. The second set of analyses describes contingencies between infants' actions and mothers' discourse. We first investigated whether changes in infants' gaze target or object handling affect mothers' probability of repeating each content type in successive utterances. We also investigated whether each content type varies in frequency as a function of infants' gaze target and/or object handling. Finally, we quantified the degree of predictability of each content type as a function of previous utterance content and of infant activity state. We expected that most content types would tend to follow utterances of the same type at above-chance levels, and that infant gaze or hand shifts would decrease this tendency. We also expected that mothers' speech would include both significant sequential transitions between different utterance content types, and correlations between utterance content and infants' gaze and object handling. Furthermore, the specific patterns should not be arbitrary, but should reflect the social, attentional, and referential functions of the utterances.

Methods

Participants

The participants were 42 mother–infant dyads (20 female) from a longitudinal study of infant social development (Chang et al., 2016; Deák, Triesch, Krasno, de Barbaro, & Robledo, 2013). Participants were recruited as a sample of convenience from the greater San Diego area. Mothers' mean age upon recruitment was 32.1 years (range = 21–42) and they had completed a mean of 16.1 years of formal education (range = 12–21). Twenty-nine infants were Caucasian, two were Asian, four were Hispanic, five were "other" or multiple races, and two parents did not report ethnicity. No infants had any neurological, cognitive, or sensory deficits, according to parental report.

An experimenter visited the participants' home each month between 3 and 9 months, and again at 12 months, and participants also visited the laboratory to complete various tasks every month. All participants gave informed consent before participating in each session. For the current study, only the 12-month home session was used. Six additional participants were recruited but dropped out of the study before the 12-month session, and four participants were tested but excluded from analyses due to equipment failure or use of non-English language, resulting in a final sample of 38 dyads. Infants' mean age was 371 days (range: 356–450; SD = 14.5 days); all but one infant was tested before 13 months of age, but excluding this participant did not affect the qualitative results.

Procedure

An experimenter visited the participants' home. Infants and mothers participated in a free-play task in which they were seated on the floor and experimenters provided three different sets of toys (switched out by the experimenter every ~3 min; Figure 2.1). Similar tasks have been widely used to study infant language learning (Song, Spier, & Tamis-Lemonda, 2014; Sosa, 2016; Suanda et al., 2016; Tamis-LeMonda et al., 2001; Yu & Smith, 2012). Object play tasks elicit rich parental speech with many opportunities for interaction, and are representative of "peaks" in infants' daily language exposure rather than average periods (Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017). Multiple sets of toys reduced the possibility of spurious results due to specific toys. The first set of toys included a set of three colorful blocks, two wooden ladybugs (red and green), and a chain of plastic rings. The second set included a set of stacking cups, a rubber duck, and a ball. The third set included a turtle, a boat, a parrot, and two pirate dolls. Free play duration averaged 9.8 min (range: 2.6–13.8; SD = 2.6). Mothers were instructed to play as they normally would, but to remain in view of the two stationary Canon mini-DV video cameras used to record the sessions. During the session, the experimenter waited quietly in a nearby room, and only entered to switch out the toy sets as quickly and unobtrusively as possible. All procedures were approved by the university's Institutional Review Board.

Coding

We coded the videos for infant gaze, infant object handling, and maternal speech. In each frame we identified the infant's gaze target; the possible targets included TOY, the mother's FACE and OTHER (e.g. furniture, the floor, or otherwise indeterminate). For each frame, we also identified whether the infant handled one or multiple toys, and the identity of the toy(s). Finally, we also transcribed all maternal speech, divided into utterances separated by gaps of at least 200 ms, with the start and stop time of each utterance identified. After transcription, utterances were tagged for 10 content types: DECLARATIVE, IMPERATIVE, QUESTION, DESCRIPTION, ATTENTION, ACTION, OBJECT NAME, INFANT'S NAME, AFFIRMATION, and SOCIAL ROUTINE. Because the content types spanned multiple levels of description, they were not mutually exclusive: utterances frequently were tagged with multiple content types, accurately reflecting the fact that many utterances contain different types of information at different levels. Example tagged utterances are shown in Table 2.1, and overlap rates are given in Table A.1. A second coder coded 10 randomly selected sessions (26% of sample). Average Cohen's kappa for utterance tags ranged from 0.76 (DESCRIPTION) to 0.96 (QUESTION). Cohen's kappa averaged 0.76 for infant gaze and 0.81 for infant object handling (Table A.2). A representative sequence is shown in Table A.3.

Analysis

We analyzed mothers' utterance content types with mixed-effects logistic regression models, using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015) to model the probability of occurrence of each utterance type. These models fit the log-odds of occurrence of a content type as a linear function of the predictor variables, with random intercepts fit for each participant dyad. *p*-values for each coefficient are calculated using likelihood ratio tests between the full model and the model with that predictor removed.

The first set of analyses examined sequences of mothers' utterance content types, testing for which content types predicted the type(s) of the following utterance. We first selected all consecutive utterance pairs, so long as the following utterance occurred within 5 s of the end of the previous one (including all utterance pairs regardless of the delay yielded qualitatively similar results). We then fit separate mixed-effects logistic regression models for each content type as a function of all the types in the previous utterance. Because most content types were most strongly predicted by previous occurrence of the same type, we conducted an additional analysis of temporal and interactive factors affecting the probability of repetition. For each content type, we selected utterances that followed an utterance containing that content type. We then fit mixed-effects logistic regression models predicting the probability of repeating the content type as a function of three factors: time elapsed between utterances, infants' shifts in gaze target, and infants' shifts in object handling.

The next set of analyses examined whether the proportion of utterances with each content type differed as a function of infants' gaze target and infants' object handling. For each content type, we fit mixed-effects logistic regression models with infants' gaze target (FACE, TOY, or OTHER) and previous utterance content types as predictors, and then with object handling (none, one, or two objects) and previous utterance type as predictors. These models were compared against the baseline models using previous utterance type alone (including these baseline predictors was necessary to remove serial correlation between consecutive utterances).

The final set of analyses examined how both factors–the previous utterance content, and the infant's gaze and object handling state–predicted the content of the next utterance. We fit mixed-effects logistic regression models using previous utterance content, the infants' gaze state, and the infant's object handling state. We then quantified how much each subset of factors (previous utterance, gaze and hands, both) predicts each content type in the next utterance. Performance was evaluated using subsets of fixed effects estimated from the joint model to predict the content types for participants held out under 10-fold cross validation (Zhang, 1993). Results were quantified using the area under the ROC curve.

To correct for the total number of comparisons, we adopted a stringent $\alpha = 0.01$ criterion for statistical significance throughout the results (note that a correction for multiple comparisons such as Bonferroni would be inappropriate because correlations among tested variables are predicted by previous results; e.g., Chang et al., 2016; Frank et al., 2013; Suanda et al., 2016). Effects with $0.01 < p < 0.05$ are indicated in figures and tables but not discussed in detail. The data and code used to generate the results may be found at: https://github.com/cogdevlabucsd/discourse-continuity-dev-sci.

Results

Descriptive statistics

The frequencies of mother's production of utterances with each content type are shown in Table 2.2, and the proportion of time that infants spent in each gaze and object-handling state are shown in Table 2.3.

Continuity and sequential structure of discourse

We analyzed the sequential structure of infant-directed discourse by identifying the predictive relations between utterance content types in consecutive utterances (Figure 2.2). The data included 7,432 utterances, of which 6,535 were preceded by an utterance within 5 s. All utterance types showed significant repetitiveness: values greater than 0 on the main diagonal of Figure 2.2 indicate that all content types were more likely to occur when the previous utterance contained the same content type (*ps* < 0.01). Pairs of successive utterance content types that were significantly predictive are highlighted in Figure 2.2 and listed in Table 2.4. The presence of significant transitions between different content types indicates that sequential contingencies in content types are a source of regularity in mothers' utterances, over and above content repetition (Figure 2.3); however, the largest magnitude effects are observed for repetitiveness. Although exact repetitions were sometimes observed, the repetition coefficients remained positive when exact repetitions were removed from the dataset for all content types except AFFIRMATION, and remained significant for all content types except AFFIRMATION, IMPERATIVE, and INFANT'S NAME (Table A.4). The proportion of exact repeats ranged from 0.08 (DECLARATIVE) to 0.41 (AFFIRMATION). Partial repeats (i.e. at least one overlapping word, but not exact repeats) constituted the majority of repeats for all content types except AFFIRMATION and SOCIAL ROUTINE (Table A.5).

Relations between maternal discourse and infants' gaze and object handling

To further characterize the repetitiveness observed in the previous section, we investigated whether infants' activity plays a role in the repetitiveness of maternal discourse. Specifically, we tested whether infants' gaze shifts or object handling shifts affect the probability of repetition of each content type. Gaze shifts were defined as any difference in the gaze target at the start times of the two utterances, and handling shifts were defined as any difference in the object(s) being handled at the start times of the two utterances. For each content type, we first identified all utterances that followed an utterance with that same type. We then modeled the probability of repetition for each type using mixed-effects linear models with time (i.e., seconds between end of utterance and start of next utterance), gaze shift, and hand shift as fixed effects, and with dyad as a random effect. Results showed that probability of repetition decreased over time for nine of ten content types, reliably so for six types (Table 2.5). Although infants' gaze and hand shifts did not generally affect the probability of repetition, gaze shifts reliably decreased the probability of repeating INFANT'S NAME ($\beta = -1.21$, $p < 0.01$). Marginal effects suggest that gaze shifts might decrease the probability of repeating SOCIAL ROUTINE ($\beta = -0.44$, $p = 0.04$), and hand shifts might decrease the probability of repeating ACTION ($\beta = -0.30$, $p = 0.02$).

Next, for each utterance content type, we used mixed-effects logistic models to test whether the proportion of utterances containing that type differed as a function of infants' gaze target (FACE, TOY, or OTHER), and/or object handling (zero, one, or two objects), controlling for effects of previous utterance content. Infant gaze and hands were codable for 6763 utterances. Significant effects ($ps < 0.01$) of gaze target were found for the content types IMPERATIVE,

DESCRIPTION, ATTENTION, OBJECT NAME, INFANT'S NAME, and AFFIRMATION (see Figure 2.4). *Post hoc* tests showed that DESCRIPTION and OBJECT NAME were more frequent when infants looked at TOY than other, whereas IMPERATIVE and ATTENTION utterances were more frequent when infants looked at OTHER than TOY. Also, INFANT'S NAME was more frequent when infants looked at OTHER than TOY or FACE, and AFFIRMATION was more frequent when infants looked at FACE than TOY.

Significant effects ($ps < 0.01$) of object handling were found for IMPERATIVE, DESCRIPTION, and INFANT'S NAME (Figure 2.5). *Post hoc* tests showed that IMPERATIVE was more frequent when infants handled zero objects than two, DESCRIPTION was more frequent when infants handled one or two objects than zero, and INFANT'S NAME was more frequent when infants handled zero objects than one.

Predictability of utterance content

We next investigated how much infants' concurrent gaze and object handling and the previous utterance content contributed to the predictability of utterance content type. Using mixed-effects logistic regression, we fit models predicting the probability of each utterance content type based on the full set of predictors (gaze, handling and previous utterance content). We then used 10-fold cross-validation (assigning each dyad to one fold) to quantify how much fixed effects for gaze/handling, previous utterance content, or both improved predictability over chance levels in sessions not used for model fitting. We evaluated the predictability using the area under the ROC curve (AUC; Table 2.6). This measure is computed by plotting the true positive rate against the false positive rate as the decision boundary (i.e., the predicted probability of the content type necessary for a positive prediction) is varied. An AUC value of 0.5 indicates chance prediction (true positives do not outnumber false positives for any decision boundary), and 1.0 indicates perfect accuracy (i.e., some decision boundary detects all true positives and no false positives).

Consistent with the previous analyses, prediction was better than chance for most gaze/handling models (AUC from 0.49 to 0.60), for all discourse models (AUC from 0.56 to 0.69), and for all full models (AUC from 0.58 to 0.70). The previous utterance content was generally a stronger predictor than gaze and handling information. However, gaze/handling state was a stronger predictor than previous utterance content for INFANT'S NAME. The two factors performed approximately equally well for predicting AFFIRMATION (Figure 2.6). For most content types, the model using only previous content performed as well as (i.e., AUC within 0.01 of) the full model, indicating that gaze/handling information was redundant with previous content; however, for utterances with INFANT'S NAME and OBJECT NAME, both sources of information made independent contributions.

Discussion

The current results confirm that a combination of sequential discourse structure and concurrent infant gaze and manual activity predict a wide range of content types in naturalistic maternal speech to 12-month-old, English-learning infants. Unlike previous studies on the linguistic and action context of IDS (e.g., Frank et al., 2013; Trueswell et al., 2016; Yu & Smith, 2012), these effects were not measured in terms of relations between specific words and their referents. Nor were they limited to concrete utterances referring to observable objects or actions. Instead, sequential and cross-modal relations were found for a wide range of abstract content types that are prevalent in IDS and relevant to language acquisition.

Maternal speech exhibited pervasive repetitiveness, in that utterances with a given type of content were likely to be followed by one or more successive utterances with the same type of content. Notably, this pattern held across all the 10 syntactic, thematic, and lexical content types tested. For most content types, most repetitions showed only partial (not complete) overlap with the previous utterance. In addition, discourse type repetition was more likely when utterances were closely spaced in time.

Repetitiveness in IDS has been documented previously: it is known that caregivers frequently produce both exact and partial repetitions (Hoff-Ginsberg, 1990; Kaye, 1980; Snow, 1972). Infants, notably, prefer the expected level of repetitiveness of IDS (McRoberts, McDonough, & Lakusta, 2009). This repetition seems to have specific pragmatic functions: for example, parents' consecutive utterances tend to refer to the same object even after taking into account infants' and parents' ongoing and coordinated gaze and object handling (Frank et al., 2013). Moreover, this structured repetition impacts language learning. Continuity in discourse has been found experimentally to facilitate word learning: 2-year-olds learn new object words more effectively when they are repeated in consecutive utterances (Schwab & Lew-Williams, 2016), and toddlers can use continuity of reference to infer the referent of otherwise ambiguous utterances (Horowitz & Frank, 2015). Effects of repetition on learning might be more general: for example, repetitive exposure to the same individuals in time has been proposed to facilitate learning in other domains such as face processing (Jayaraman & Smith, 2018).

Our findings demonstrate that content repetition in IDS is more general than previously recognized, in that discourse continuity is not specific to one or a few types of linguistic content. If infants use the continuity of adjacent utterances to reduce the ambiguity of object-referential utterances, they could potentially use the same mechanism to more accurately identify the communicative function of a wide variety of infant-directed utterances.

In addition to repetition of content types, we also found significant sequential patterning of mothers' use of different content types across utterances, demonstrating that continuity of infant-directed discourse also includes sequential transitions between content types. We identified sequences or paths of frequent transitions, such as INFANT'S NAME → ATTENTION → DECLARATIVE and INFANT'S NAME → ATTENTION → DESCRIPTION. These sequences might help to focus infants' attention on the most informative utterances by first capturing the infant's attention by saying their name, then orienting the infant to an object, then providing an informative comment. A prototypical example is the following sequence:

what's under here [name]? good job! look what it is. it's the blue one and a purple.

Infants might initially experience such sequences while spontaneously exploring their environment or following overt cues. However, as they learn to predict and comply with caregivers' attentional bids, infants might also increasingly choose actions that facilitate the emergence of more complex and reliable sequences.

Consistent with this idea, our results show systematic patterning of mothers' discourse content depending on infants' state of engagement with the physical and social environment. Although we did not find general effects of infant attention shifts on content type repetition, there was at least one specific effect: a lower rate of repetition of the infant's name after infant gaze shifts, suggesting that infants' reorientations in relation to maternal utterances might facilitate sequential transitions in discourse. Given that only two infant behavioral states were coded, a more thorough characterization of infant states might reveal more specific relations to the caregiver's speech repetition. Additionally, several content types differed in frequency based on infants' attention states. When infants gaze at toys, they hear more descriptive language and more object names; when they gaze at faces, they hear more affirmations, and when they gaze away from faces and toys, they hear more imperatives, their own name, and attention-related language. Similarly, when infants handle one or more toys, they hear more descriptive language; whereas when they do not handle objects, they hear more imperatives and their own name. These results complement and expand a sizable literature (reviewed above) on maternal speech about objects, and maternal responsiveness in dyadic and triadic interactions.

Some of these differences might be due to mothers responding to infants' attentional engagement. Nevertheless, such associations could still help establish the pragmatic force of some utterances as regulatory and others as informative. Furthermore, some utterance types are observed to differ in frequency across "on-task" states. Notably, affirmations are associated with gaze at the mother's face rather than at objects, reflecting their interpersonal rather than referential role. Also, whereas descriptions are associated with both gaze to and handling of toys, object names were associated only with gaze.

We also compared how much the infant's gaze and object handling and/or the previous utterance content predicted occurrences of each utterance content type. One or both types of contextual factors predicted every content type above chance, although substantial uncertainty remained for all content types, suggesting that other predictive contextual factors have not yet been identified. Previous utterance content was generally a stronger predictor than concurrent gaze and object handling, but the relative importance of the cues differed substantially across content types. Thus, some discourse types were typically produced within a longer discourse sequence, whereas other functions were more immediately responsive to the infant's actions. At one extreme, object names and descriptions were highly dependent on the previous utterance, but occurred frequently in all configurations of the infant's gaze and object handling; at the other, the infant's name occurred in response to infants' gaze away from toys and the mother's face but depended only weakly on previous utterances. These differences can reflect the interacting roles of different discourse types: for instance, mothers typically did not talk about attention as a direct consequence of an infant's disengagement with the toys, but rather oriented the infant to an interesting object only after first capturing their attention by calling their name.

Previous research suggests how correspondences between infants' actions and parents' speech might facilitate language learning. In the well-studied case of object-name learning, 17- to 18-month-old infants are more likely to learn words that are uttered during episodes of joint attention (Baldwin, 1991; Tomasello & Farrar, 1986), and 18-month-old infants learn more object labels that are presented while the object dominates their visual field (Yu & Smith, 2012). The effectiveness of object naming, therefore, depends not only on the correspondence between object and label, but also on a supportive context in which the utterance is expected and appropriate. The current results suggest that nonverbal context could similarly support the usage of a range of referential (descriptions, object names, talk about action) and regulatory (attention, imperative, infant's name, affirmations) discourse types. Our approach complements other empirical studies of structured relations between maternal speech and concurrent activities that are predictable within a variety of contexts (e.g., diaper-changing; Nomikou & Rohlfing, 2011). Characterizing language input at the level of associations between maternal discourse content and infants' embodied behavioral states is also consistent with the proposal that infants represent linguistic meanings not as specific objects or events, but as frames of routinized goal-directed activity (Rohlfing, Wrede, Vollmer, & Oudeyer, 2016). As infants learn their role in interactive sequences, they might further improve their ability to participate in and learn from routines, eventually leading to the emergence of conversation. The current results characterize speech-activity relations experienced by infants at a particularly important developmental stage, as predictable, structured activities play an increasingly complex and integral role in 1-year-old' social interactions (De Barbaro, Johnson, & Deák, 2013; Nelson, 1998), despite their still immature and incomplete linguistic knowledge and symbolic skills.

Discourse continuity in IDS might facilitate language acquisition in several ways. Partial repetitions of parental utterances correlate positively with children's syntactic development (Baker & Nelson, 1984; Hoff-Ginsberg, 1985, 1986). Discourse continuity has been previously argued to facilitate syntax acquisition because partial overlap in adjacent utterances constrains word and phrase boundaries (Onnis, Waterfall, & Edelman, 2008), and facilitates lexical acquisition by clustering object-labeling events (Schwab & Lew-Williams, 2016; Vlach & Johnson, 2013). Structured sequences in discourse and interaction might cue infants to expect utterances with novel information such as object names or descriptions. In fact, these were the two most predictable content types in our data. Our finding of significant transitions from attention-regulating to declarative and descriptive utterances suggests that attention-directing talk might be partially responsible for this predictability. This raises the possibility that findings of negative correlations between directive maternal talk and child language outcomes (Hughes et al., 1999; Nelson, 1973; Tomasello & Todd, 1983) are related to the disruption of these sequences, perhaps due to infants' failure to respond to bids for attention, rather than differences in speech style per se.

Sequential continuity and repetitiveness might further facilitate learning of syntactic and/or pragmatic categories if continuity effects are stronger for valid categories than for arbitrary ones. Repetition in discourse might highlight similarities among exemplars of a content type, whereas sequences emphasize contrastive features (Carvalho & Goldstone, 2017). Moreover, infants might better recognize repeated word tokens if prosody or other utterance-level features are consistent across utterances (Bortfeld & Morgan, 2010). Consistent sequences in discourse might also make it possible to infer similar meanings between utterances that occur in similar discourse contexts, analogous to inferring word meaning by familiar sentence context (Miller & Charles, 1991). Once children learn to detect discourse continuity, contextual utterances could constrain the interpretation of utterances that would be ambiguous in isolation. This effect has been called discourse bootstrapping. For example, by 2 years of age children can use discourse continuity to select a referent for an unfamiliar word suggested by a speaker's previously expressed preferences (Sullivan & Barner, 2016), and parents use multiple cues such as syntactic frames, utterance-final position, and prosodic emphasis to introduce unfamiliar words to 2- to 5-year-old children (Clark, 2010). Although we cannot yet determine how much discourse bootstrapping facilitates meaning-inference relative to syntactic and semantic bootstrapping, the current results indicate that information supporting the former is available by 12 months.

Finally, discourse continuity might facilitate language acquisition via general information processing factors such as increased processing speed (see Conway, Bauernschmidt, Huang, & Pisoni, 2010) and reduced working memory load. These mechanisms might bridge correlational findings indicating that parental speech style influences language acquisition (Barnes, Gutfreund, Satterly, & Wells, 1983; Hart & Risley, 1995; Tamis-LeMonda et al., 2001) with mechanistic accounts of how infants find structure in linguistic input (Hoff & Naigles, 2002; see also Tamis-LeMonda, Kuchirko, & Song, 2014).

The current study has several limitations. We observed mother-infant dyads interacting in a restricted context that captured only a limited portion of their daily activities. Also, the dyads were a sample of convenience of mostly middle- to upper-class families that did not faithfully represent local population demographics with respect to education and ethnicity. This limits the generality of the findings–specifically, the patterns of experience found in this sample might not hold in different contexts or populations. However, context-specific structure in infants' language exposure might facilitate language learning over and above regularities that are evident when aggregating over all contexts (Beals, 1997; Roy, Frank, DeCamp, Miller, & Roy, 2015; Roy, Frank, & Roy, 2012). Another limitation of the current study is the necessarily somewhat arbitrary classification of verbal content types. Further research could attempt to derive a "bottom-up" taxonomy based on raw verbal data, and investigate how much the sequential and cross-modal structure of maternal speech affords unsupervised discovery of content types.

The current study simplified interactions that unfold over time by studying state transitions and concurrent co-occurrences between utterances and other events. Although a lagged analysis of relations between parental speech and infants' past attention states is outside the scope of the study, concurrent relations between infants' attention state and mothers' speech can also be interpreted as contingent responses to infant state changes. However, mothers' speech might also depend on sequences and durations of infant activities. Infants' object handling, for example, tends to occur later in episodes of shared attention to objects, compared to object gaze (Chang et al., 2016; Rohde & Frank, 2014). Therefore, another layer of contingency remains to be revealed. Future work also could explore how parents' discourse patterns are related to infants' comprehension of specific words or constructions, as the current study and related studies have not disaggregated contingencies based on the familiarity or novelty of linguistic elements to the infant. In addition, an extension of the current study could examine discourse contingencies at multiple ages because parents integrate speech into play differently depending on the infant age or developmental status (Bornstein et al., 2008; Murphy, 1978). Finally, future research might improve models by adding individual differences in caregiver speech (e.g., level of discourse repetitiveness) as a predictor of infants' emerging speech processing fluency or early linguistic knowledge.

Conclusion

The current results demonstrate that mothers' speech to 12-month-old infants is repetitive at multiple levels of analysis, is sequentially structured, and is contingently related to infants' exploratory activity. All of the syntactic, thematic, and lexical content types that we examined tended to occur preferentially in specific positions in the discourse. Also, most types showed significant associations with infants' ongoing gaze and/or manual activity, even when ignoring specific word-object pairings. Yet the specific predictors of each utterance content type varied: some types were more repetitive, others more sequential, others more responsive to infant actions. This richly structured input thus provides opportunities for infants to distinguish between pragmatic functions of different content types in caregiver speech. The patterns we identified might also reflect mothers' efforts to ensure that informative utterances occurred at times when their infant was attentive to the referent object, and therefore might expect informative input to quickly follow. If infants are sensitive to these regularities, then the diversity of structured discourse contexts might constitute a source of information that facilitates, rather than complicates, learning. Future research on language acquisition should account for such information within connected, multimodal, interactive discourse, thereby more fully explicating the mechanisms by which social interaction endows infants' linguistic experience with an organizing structure.

Acknowledgements

We would like to thank Jordan Danly for recording the sessions, Shirlene Wade for supervising the speech transcription, and Scott Ragland, Kimberli Ng, Tiffany Ziebol, Vivian Zhang, Maria Kremliovsky, Christen Buetz, and Erin Maxwell for assisting with coding. Kaya de Barbaro and Christine Johnson contributed to developing behavioral codes used in the study, and Sean Trott provided comments on a draft. This research was supported and by the National Science Foundation (NSF) Graduate Research Fellowship under grant no. DGE-1144086, NSF grant SES-0527756 to G. Deák and J. Triesch, and by the Temporal Dynamics of Learning Center (NSF SMA-1005256).

Chapter 2, in full, is a reprint of the material as it appears in: Chang, L., & Deák, G. (2019). Maternal discourse continuity and infants' actions organize 12-month-olds' language exposure during object play. *Developmental science,* 22(3), e12770. The dissertation author was the primary investigator and author of this paper.

Figure 2.1. Sample screen-shots of mother-infant dyads in free play with standardized toy sets.

Next Prev.	Dec	Imp	Que	Dsc	Att	Act	ONm	INm	Aff	Soc
Dec	**.63****	-.13	-.03	-.09	.01	-.18	-.26	-.56	.06	.01
Imp	.07	**.60****	-.14	-.28	-.09	-.02	-.18	.12	-.41	.11
Que	.03	-.26	**.83****	.08	-.18	-.13	-.17	-.20	-.07	-.19
Dsc	**.31****	-.42	-.30	**1.89****	-.27	-.24	-.24	-.16	-.21	-.22
Att	**.35***	-.02	-.27	**.51***	**.94****	-.27	.14	-.11	-.25	-.23
Act	-.33	**.57****	-.25	-.47	.05	**1.27****	-.22	-.17	**.31†**	.07
ONm	.06	-.19	-.02	-.07	-.10	-.26	**1.66****	-.11	-.20	-.18
INm	-.58	**.64****	.01	-.77	**.64****	.08	.16	**.90****	-.72	-.04
Aff	.03	.04	.12	-.00	-.14	.10	.07	-.34	**.41***	.09
Soc	-.07	-.03	.05	-.13	-.12	.17	-.15	-.14	-.14	**1.07****

Figure 2.2. Predictive relations between content types in consecutive utterances. Row indicates content types present in the preceding utterance. Column indicates content type being predicted. Values shown are coefficients representing the difference in log-odds when the previous content type is present. †$p < 0.05$, *$p < 0.01$, and **$p < 0.001$. Significance is only shown for positive predictors. Dec: DECLARATIVE; Imp: IMPERATIVE; Que: QUESTION; Dsc: DESCRIPTION; Att: ATTENTION; Act: ACTION; ONm: OBJECT NAME; INm: INFANT'S NAME; Aff: AFFIRMATION; Soc: SOCIAL ROUTINE

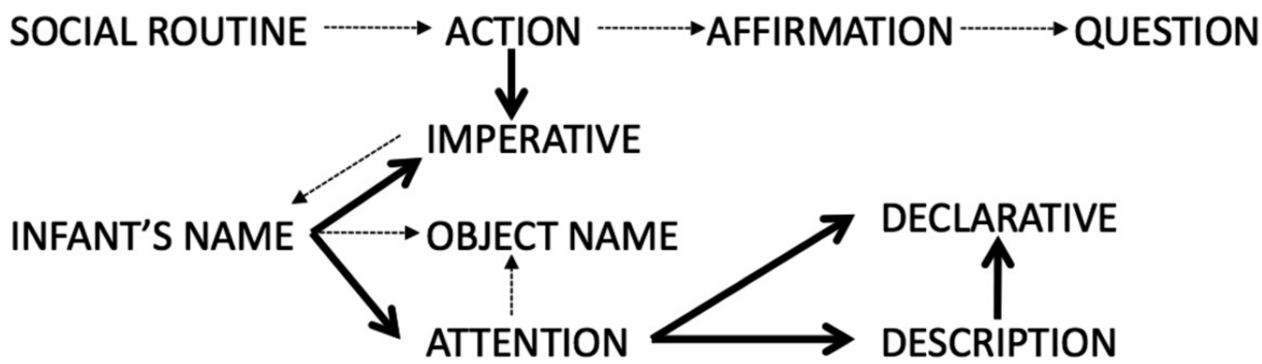

Figure 2.3. Graph of above-chance transitions in mothers' discourse. Significant predictive relations are shown as bold arrows. Then, smaller dashed arrows were added for the highest remaining coefficients until the graph was fully connected.

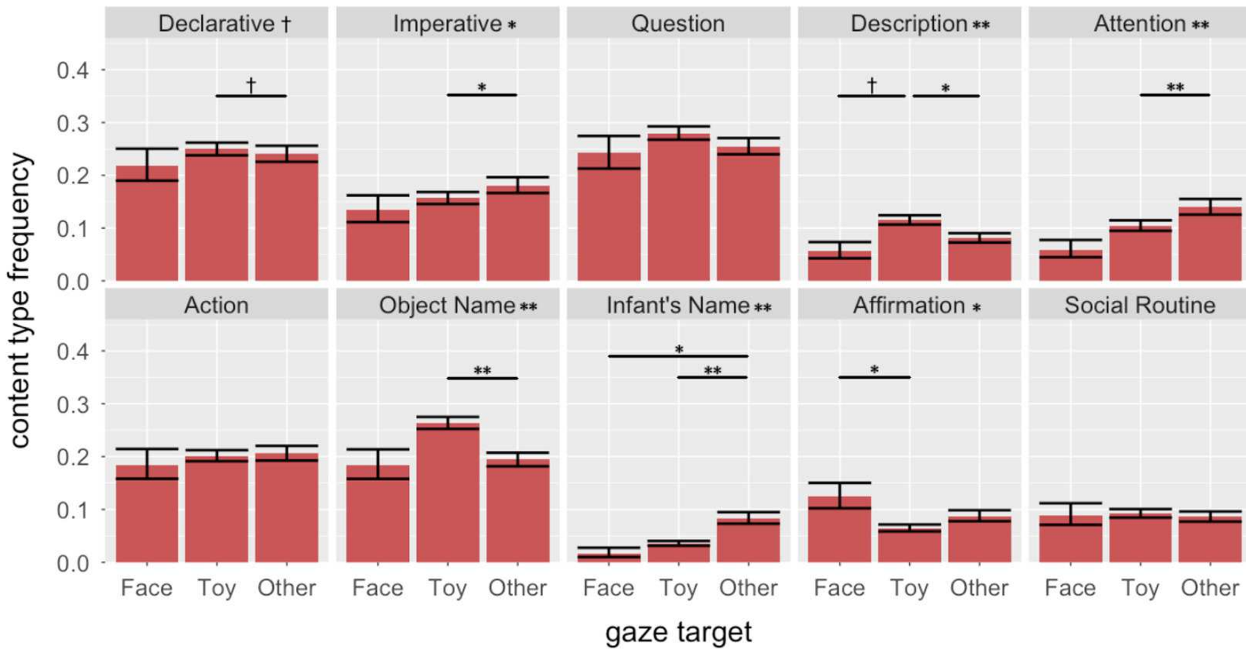

Figure 2.4. Estimated frequency of each content type as a function of infant gaze target type, evaluated at mean values of previous-utterance predictors. Error bars represent the standard error of the fitted value. Significance symbols in the panel titles represent likelihood ratio tests for the effect of infant gaze target in the linear mixed effects model. Significant symbols within the panels represent pairwise *post hoc* tests. †$p < 0.05$, *$p < 0.01$, and **$p < 0.001$.

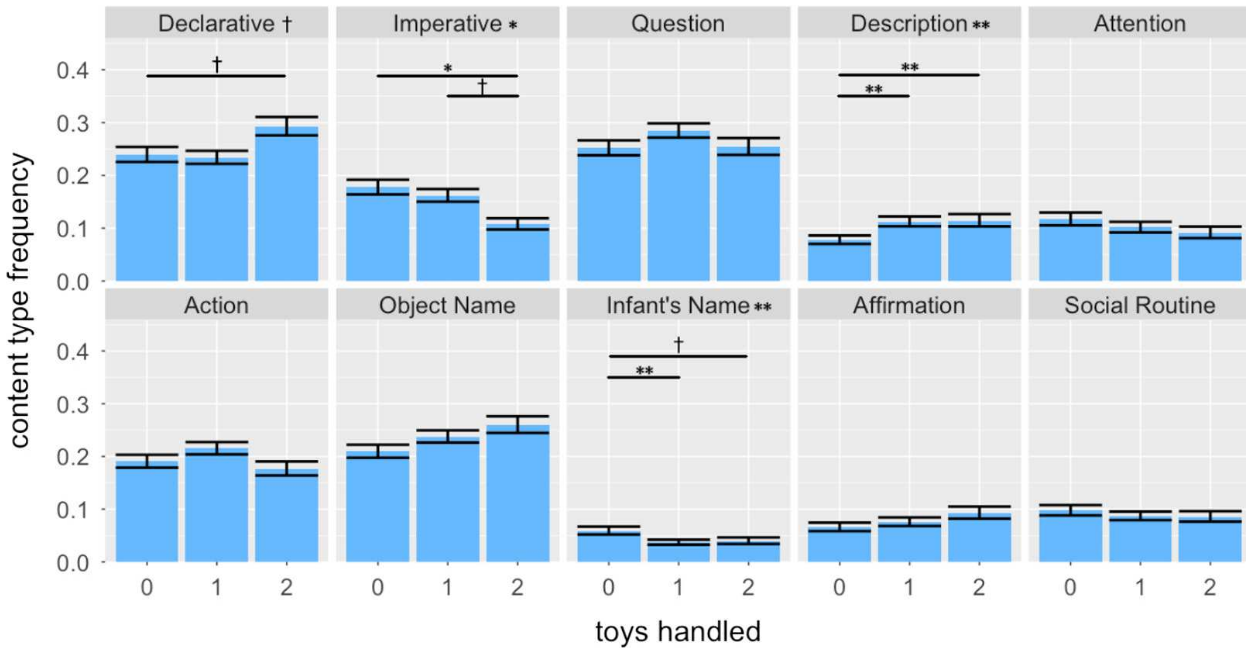

Figure 2.5. Estimated frequency of each content type as a function of infant object handling, evaluated at mean values of previous-utterance predictors. Error bars represent the standard error of the fitted values. Significance symbols in the panel titles represent likelihood ratio tests for the effect of infant object handling in the linear mixed effects model. Significance symbols within the panels represent pairwise *post hoc* tests. †$p < 0.05$, *$p < 0.01$, and **$p < 0.001$.

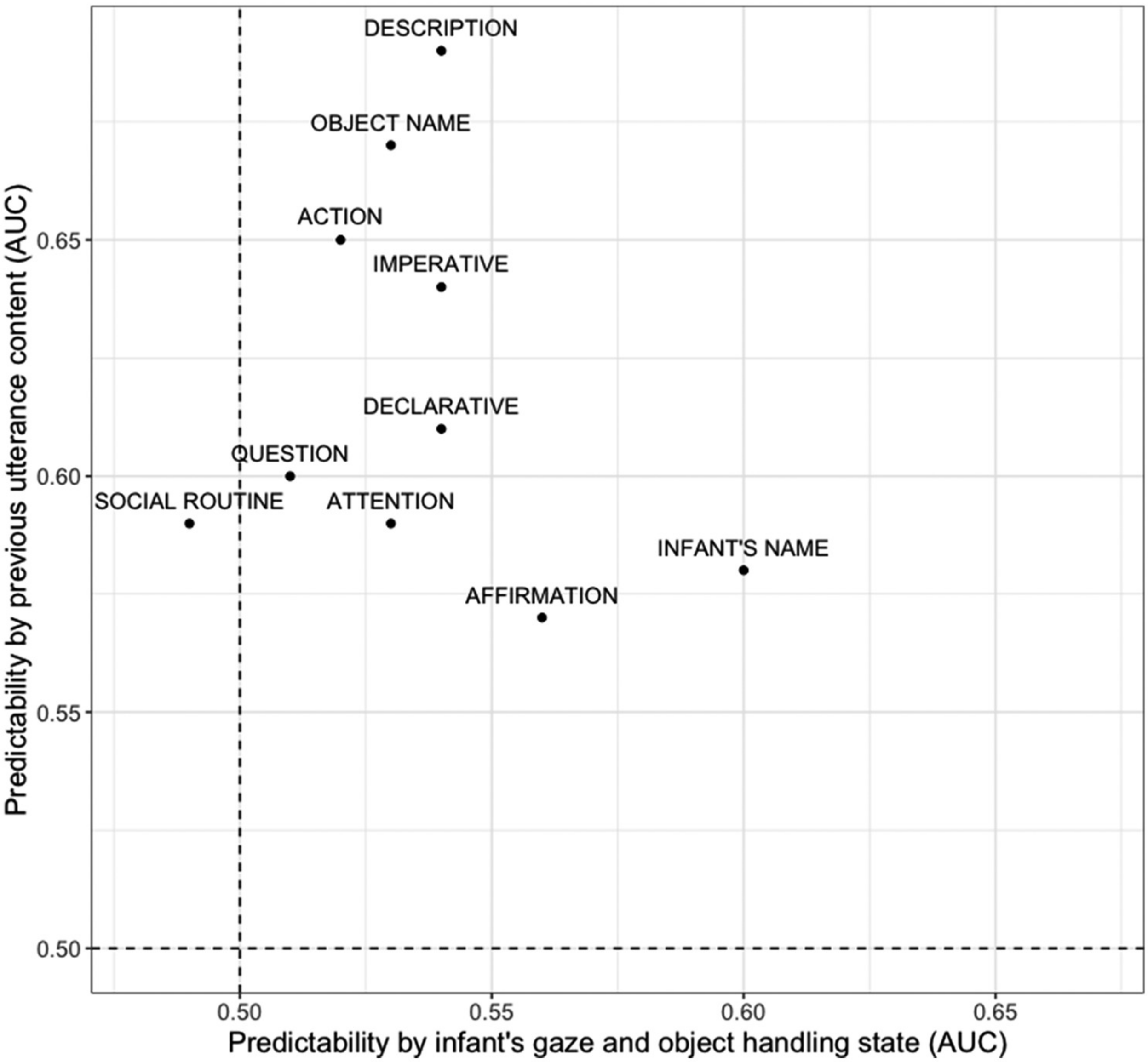

Figure 2.6. Content types plotted by degree of predictability (AUC) when predicting using the infant's gaze and object handling state or the previous utterance content. Dashed lines indicate chance prediction.

Table 2.1. Discourse content types, definitions of types, and a representative utterance. Because content types are not mutually exclusive, the right-most column indicates additional content types that were tagged in the example utterance, if any.

Content type	Definition	Example	Example also contains
DECLARATIVE	Declarative statements, excluding one-word utterances	*That's pretty neat*	–
IMPERATIVE	Imperative syntax, including *Let's*	*Go get it*	ACTION
QUESTION	Question syntax, or utterance-final pitch rise in appropriate context	*What's over here?*	–
DESCRIPTION	Adjectives or predicates referring to toys	*It's the blue one*	DECLARATIVE
ATTENTION	Directing or commenting on infant's focus of attention	*Can you see mommy?*	QUESTION
ACTION	Directing or commenting on infant's non-perceptual actions	*You wanna try and squeeze it?*	QUESTION
OBJECT NAME	Contains name for physically present toys	*We got a little rubber ducky*	DECLARATIVE
INFANT'S NAME	Contains infant's name or other term of address	*Hi [name]*	SOCIAL ROUTINE
AFFIRMATION	Contains a form of *Yes* or acknowledges a conversational turn without further content	*Yeah*	–
SOCIAL ROUTINE	Fixed expressions such as greetings or exclamations	*Thank you*	–

Table 2.2. Frequency of utterance content types

Variable	Frequency/minute: mean (range), *SD*
Total utterances	18.1 (8.7–26.4), 4.0
DECLARATIVE	4.9 (1.9–8.8), 1.8
IMPERATIVE	3.5 (0.9–7.7), 1.9
QUESTION	4.9 (2.1–7.9), 1.6
DESCRIPTION	2.2 (0.2–8.2), 1.5
ATTENTION	2.3 (0.4–6.0), 1.4
ACTION	4.0 (1.0–7.5), 1.5
OBJECT NAME	4.5 (1.2–8.1), 1.6
INFANT'S NAME	1.0 (0.2–4.1), 0.8
AFFIRMATION	1.6 (0.3–3.6), 0.9
SOCIAL ROUTINE	2.0 (0.5–5.1), 1.1

Table 2.3. Proportion of time spent in gaze/object handling states

Variable	Proportion: mean (range), *SD*
Gaze::toy	0.74 (0.27–0.92), 0.15
Gaze::face	0.03 (0–0.13), 0.03
Gaze::other	0.23 (0.04–0.68), 0.15
Handling::0 objects	0.31 (0.01–0.63), 0.13
Handling::1 objects	0.48 (0.21–0.70), 0.11
Handling::2 objects	0.21 (0.03–0.43), 0.11

Table 2.4. Significant sequential structure in mothers' discourse. Effects are listed if their coefficients were positive with $p < 0.01$.

Content type	Predicts
DECLARATIVE	DECLARATIVE
IMPERATIVE	IMPERATIVE
QUESTION	QUESTION
DESCRIPTION	DESCRIPTION, DECLARATIVE
ATTENTION	DESCRIPTION, DECLARATIVE, ATTENTION
ACTION	ACTION, IMPERATIVE
OBJECT NAME	OBJECT NAME
INFANT'S NAME	INFANT'S NAME, IMPERATIVE, ATTENTION
AFFIRMATION	AFFIRMATION
SOCIAL ROUTINE	SOCIAL ROUTINE

Table 2.5. Effects of time between utterances and infant attention shifts on probability of content type repeats.

Content type	Time (s)	Gaze shift	Hand shift
DECLARATIVE	−0.36**	−0.05	0.04
IMPERATIVE	−0.20**	−0.20	0.08
QUESTION	−0.15*	0.15	−0.10
DESCRIPTION	−0.54**	−0.08	−0.10
ATTENTION	−0.17†	−0.04	0.07
ACTION	−0.19**	−0.17	−0.30†
OBJECT NAME	−0.05	0.00	0.19
INFANT'S NAME	−0.06	−1.21*	0.30
AFFIRMATION	0.02	0.01	−0.45
SOCIAL ROUTINE	−0.26*	−0.44†	0.09

† $p < 0.05$, * $p < 0.01$, and ** $p < 0.001$.

Table 2.6. Area under curve scores for logistic regression models predicting each utterance content type. 0.50 represents chance performance, and 1.0 represents perfect accuracy.

	Gaze + handling	Previous utterance	All (full model)
DECLARATIVE	0.54	0.61	0.61
IMPERATIVE	0.54	0.64	0.64
QUESTION	0.51	0.60	0.60
DESCRIPTION	0.54	0.69	0.70
ATTENTION	0.53	0.59	0.59
ACTION	0.52	0.65	0.65
OBJECT NAME	0.53	0.67	0.69
INFANT'S NAME	0.60	0.58	0.62
AFFIRMATION	0.56	0.57	0.58
SOCIAL ROUTINE	0.49	0.59	0.59

References

Akhtar, N., Dunham, F., & Dunham, P.J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language, 18*(1), 41-49.

Baker, N.D., & Nelson, K.E. (1984). Recasting and related conversational techniques for triggering syntactic advances by young children. *First Language, 5*(13), 3-21.

Baldwin, D.A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*(5), 874-890.

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Baumwell, L., Tamis-LeMonda, C.S., & Bornstein, M.H. (1997). Maternal verbal sensitivity and child language comprehension. *Infant Behavior and Development, 20*(2), 247-258.

Beals, D.E. (1997). Sources of support for learning words in conversation: Evidence from mealtimes. *Journal of Child Language, 24*(3), 673-694.

Benitez, V.L., & Smith, L.B. (2012). Predictable locations aid early object name learning. *Cognition, 125*(3), 339-352.

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 of the United States of America, 109*(9), 3253-3258.

Bornstein, M.H., Haynes, M.O., & Painter, K.M. (1998). Sources of child vocabulary competence: A multivariate model. *Journal of Child Language, 25*(2), 367-393.

Bornstein, M.H., Tamis-LeMonda, C.S., Hahn, C.S., & Haynes, O.M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology, 44*(3), 867.

Bortfeld, H., & Morgan, J.L. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive Psychology, 60*(4), 241-266.

Bryant, G.A., & Barrett, H.C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science, 18*(8), 746-751.

Cannon, E.N., & Woodward, A.L. (2012). Infants generate goal-based action predictions. *Developmental Science, 15*(2), 292-298.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i-174.

Carvalho, P.F., & Goldstone, R.L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(11), 1699.

Caselli, M.C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10*(2), 159-199.

Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology, 41*(5-8), 342-361.

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language, 27*(1), 69-96.

Clark, E.V. (2010). Adult offer, word-class, and child uptake in early lexical acquisition. *First Language, 30*(3-4), 250-269.

Conway, C.M., Bauernschmidt, A., Huang, S.S., & Pisoni, D.B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114*(3), 356-371.

Deák, G.O., Triesch, J., Krasno, A., de Barbaro, K., & Robledo, M. (2013). Learning to share: The emergence of joint attention in human infancy. In B. Kar (Ed.), *Cognition and brain development: Converging evidence from various methodologies* (pp. 173-210). Washington, DC: APA.

de Barbaro, K., Johnson, C.M., & Deák, G.O. (2013). Twelve-month "Social revolution" emerges from mother-infant sensorimotor coordination: A longitudinal investigation. *Human Development, 56*(4), 223-248.

de Barbaro, K., Johnson, C.M., Forster, D., & Deák, G.O. (2016). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother-infant-object interactions. *Child development, 87*(2), 494-512.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60*(6), 1497-1510.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*(3), 657-674.

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.

Gallagher, T.M. (1981). Contingent query sequences within adult–child discourse. *Journal of Child Language, 8*(1), 51-62.

Gros-Louis, J., West, M.J., Goldstein, M.H., & King, A.P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development, 30*(6), 509-516.

Gros-Louis, J., West, M.J., & King, A.P. (2016). The influence of interactive context on prelinguistic vocalizations and maternal responses. *Language Learning and Development, 12*(3), 280-294.

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

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*(5), 1368-1378.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*(2), 418-433.

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

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

Hoff-Ginsberg, E. (1990). Maternal speech and the child's development of syntax: A further look. *Journal of Child Language, 17*(1), 85-99.

Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child Development, 62*(4), 782-796.

Horowitz, A.C., & Frank, M.C. (2015). Young children's developing sensitivity to discourse continuity as a cue for inferring reference. *Journal of Experimental Child Psychology, 129*, 84-97.

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

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 236.

Jayaraman, S., & Smith, L.B. (2018). Faces in early visual environments are persistent not just frequent. *Vision Research, 157*, 213-221.

Karasik, L.B., Tamis-LeMonda, C.S., & Adolph, K.E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental Science, 17*(3), 388-395.

Katz, G.S., Cohn, J.F., & Moore, C.A. (1996). A combination of vocal f_0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development, 67*(1), 205-217.

Kaye, K. (1980). Why we don't talk ‘baby talk’ to babies. *Journal of Child Language, 7*(3), 489-507.

Lloyd, C.A., & Masur, E.F. (2014). Infant behaviors influence mothers’ provision of responsive and directive behaviors. *Infant Behavior and Development, 37*(3), 276-285.

Mandel, D.R., Jusczyk, P.W., & Pisoni, D.B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science, 6*(5), 314-317.

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

Masur, E.F., Flynn, V., & Lloyd, C.A. (2013). Infants' behaviors as antecedents and consequents of mothers' responsive and directive utterances. *First Language, 33*(4), 354-371.

McRoberts, G.W., McDonough, C., & Lakusta, L. (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy, 14*(2), 162-194.

Messer, D.J. (1980). The episodic structure of maternal speech to young children. *Journal of Child Language, 7*(1), 29-40.

Miller, G.A., & Charles, W.G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes, 6*(1), 1-28.

Mintz, T.H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91-117.

Murphy, C.M. (1978). Pointing in the context of a shared activity. *Child Development, 49*(2), 371-380.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*(1/2), 1-135.

Nelson, K. (1998). *Language in cognitive development: The emergence of the mediated mind*. Cambridge, UK: Cambridge University Press.

Ninio, A., & Snow, C.E. (1996). *Pragmatic development*. Boulder, CO: Westview Press.

Ninio, A., Snow, C.E., Pan, B.A., & Rollins, P.R. (1994). Classifying communicative acts in children's interactions. *Journal of Communication Disorders, 27*(2), 157-187.

Nomikou, I., & Rohlfing, K.J. (2011). Language does something: Body action and language in maternal input to three-month-olds. *IEEE Transactions on Autonomous Mental Development, 3*(2), 113-128.

Onnis, L., Waterfall, H.R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition, 109*(3), 423-430.

Pan, B.A., Perlmann, R.Y., & Snow, C.E. (2000). Food for thought: Dinner table as a context for observing parent-child discourse. In L. Menn, & N. B. Ratner (Eds.), *Methods for studying language production* (pp. 205-224). Mahwah, NJ: Erlbaum

Pan, B.A., Rowe, M.L., Singer, J.D., & Snow, C.E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*(4), 763-782.

Papoušek, M., Bornstein, M.H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development, 13*(4), 539-545.

Pruden, S.M., Hirsh-Pasek, K., Golinkoff, R.M., & Hennon, E.A. (2006). The birth of words: ten-month-olds learn words through perceptual salience. *Child Development, 77*(2), 266-280.

Rochat, P., Querido, J.G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology, 35*(4), 950.

Rohde, H., & Frank, M.C. (2014). Markers of topical discourse in child-directed speech. *Cognitive Science, 38*(8), 1634-1661.

Rohlfing, K.J., Wrede, B., Vollmer, A.-L., & Oudeyer, P.-Y. (2016). An alternative to mapping a word onto a concept in language acquisition: Pragmatic frames. *Frontiers in Psychology, 7*, 470.

Rollins, P.R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics, 24*(2), 221-234.

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, 112*(41), 12663-12668.

Roy, B., Frank, M., & Roy, D. (2012). Relating activity contexts to early word learning in dense longitudinal data. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 935-940). Austin, TX: Cognitive Science Society.

Schegloff, E.A. (2007). *Sequence organization in interaction: Volume 1: A primer in conversation analysis* (Vol. 1). Cambridge, UK: Cambridge University Press.

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

Searle, J.R. (1976). A classification of illocutionary acts. *Language in Society, 5*(1), 1-23.

Snow, C.E. (1972). Mothers' speech to children learning language. *Child Development, 42*(2), 549-565.

Song, L., Spier, E.T., & Tamis-Lemonda, C.S. (2014). Reciprocal influences between maternal language and children's language and cognitive development in low-income families. *Journal of Child Language, 41*(2), 305-326.

Sosa, A.V. (2016). Association of the type of toy used during play with the quantity and quality of parent-infant communication. *JAMA Pediatrics, 170*(2), 132-137.

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.

Sullivan, J., & Barner, D. (2016). Discourse bootstrapping: Preschoolers use linguistic discourse to learn new words. *Developmental Science, 19*(1), 63-75.

Tamis-LeMonda, C.S., Bornstein, M.H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*(3), 748-767.

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.

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

Tamis-LeMonda, C.S., Kuchirko, Y., & Tafuro, L. (2013). From action to interaction: Infant object exploration and mothers' contingent responsiveness. *IEEE Transactions on Autonomous Mental Development, 5*(3), 202-209.

Tamis-LeMonda, C.S., Song, L., Leavell, A.S., Kahana-Kalman, R., & Yoshikawa, H. (2012). Ethnic differences in mother–infant language and gestural communications are associated with specific skills in infants. *Developmental Science, 15*(3), 384-397.

Toda, S., Fogel, A., & Kawai, M. (1990). Maternal speech to three-month-old infants in the United States and Japan. *Journal of Child Language, 17*(2), 279-294.

Tomasello, M. (1992). The social bases of language acquisition. *Social Development, 1*(1), 67-87.

Tomasello, M., Conti-Ramsden, G., & Ewert, B. (1990). Young children's conversations with their mothers and fathers: Differences in breakdown and repair. *Journal of Child Language, 17*(1), 115-130.

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

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language, 4*(12), 197-211.

Trueswell, J.C., Lin, Y., Armstrong, B., Cartmill, E.A., Goldin-Meadow, S., & Gleitman, L.R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition, 148*, 117-135.

Vlach, H.A., & Johnson, S.P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition, 127*(3), 375-382.

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language, 34*(1), 72-90.

Yont, K.M., Snow, C.E., & Vernon-Feagans, L. (2003). The role of context in mother–child interactions: An analysis of communicative intents expressed during toy play and book reading with 12-month-olds. *Journal of Pragmatics, 35*(3), 435-454.

Yu, C., & Ballard, D.H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing, 70*(13), 2149-2165.

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

Zhang, P. (1993). Model selection via multifold cross validation. *The Annals of Statistics, 21*(1), 299-313.

Zumbahlen, M.R. (1997). The role of infant locomotor onset in shaping mother-infant communication. Doctoral dissertation. University of Illinois at Urbana-Champaign.

CHAPTER 3: Adjacent and Non-Adjacent Word Contexts Both Predict Age of Acquisition of English Words: A Distributional Corpus Analysis of Child-directed Speech

Abstract

Children show a remarkable degree of consistency in learning some words earlier than others. What patterns of word usage predict variations among words in age of acquisition? We use distributional analysis of a naturalistic corpus of child-directed speech to create quantitative features representing natural variability in word contexts. We evaluate two sets of features: one set is generated from the distribution of words into frames defined by the two adjacent words. These features primarily encode syntactic aspects of word usage. The other set is generated from non-adjacent co-occurrences between words. These features encode complementary thematic aspects of word usage. Regression models using these distributional features to predict age of acquisition of 656 English words indicate that both types of features improve predictions over simpler models based on frequency and appearance in salient or simple utterance contexts. Syntactic features were stronger predictors of children's production than comprehension, whereas thematic features were stronger predictors of comprehension. Overall, earlier acquisition was predicted by features representing frames that select for nouns and verbs, and by thematic content related to food and face-to-face play topics; later acquisition was predicted by features representing frames that select for pronouns and question words, and by content related to narratives and object play.

Introduction

Infants' linguistic, social, and physical environment provides a wide array of features that might support early word learning. How do the myriad cues and correlations in the world combine to determine which words are learned, and when? Laboratory studies have demonstrated, at various ages and situations, influences of factors including dominance of the word's referent in the child's visual field (Yu & Smith, 2012), speech-motion synchrony (Gogate & Bahrick, 1998), children's biases to attend preferentially to relevant cues such as object shape

(Landau, Smith & Jones, 1988), discourse context (Horowitz & Frank, 2015; Sullivan & Barner, 2016), and the distribution of exposures over time (Childers & Tomasello, 2002; Vlach & Johnson, 2013). However, laboratory studies typically involve mapping novel nouns to novel toys during a single experimental session, which is not representative of normal word acquisition. During typical language acquisition, early-learned words represent diverse syntactic and semantic types and are experienced repeatedly over weeks or months. To understand the role of these experiences in word learning, researchers have related the normative age of acquisition (AoA) of words to measures of their frequency and usage patterns in infant-directed speech, as reviewed below.

Predicting Age of Acquisition

Recently, large databases of infants' vocabularies collected using the MacArthur-Bates Communicative Development Inventory (CDI; Fenson, Marchman, Thal, Dale, & Reznick, 2007) have made it easier to study the normative AoA of individual words (Frank, Braginsky, Yurovsky, & Marchman, 2017). Infant-directed speech corpora such as those in CHILDES (MacWhinney, 2014) make it possible to derive predictors from real-world usage patterns. Using these data, Braginsky et al. found that normative ages of comprehension and production in seven languages are predicted by high word frequency, low mean length of utterance (MLU), frequent appearance in isolated or utterance-final positions, high concreteness ratings, and high "babiness" ratings by adults–i.e., relevance of a word to babies (Braginsky, Yurovsky, Marchman, & Frank, 2016, 2017). Machine-extracted prosodic features also predict age of comprehension (Frermann & Frank, 2017). Swingley and Humphrey (2018) extended these analyses to predict individual differences. By pairing individual children's CDIs with samples of their parents' speech, they found that word frequency, frequency of occurrence in isolation, shorter utterance length, and longer spoken duration predicted both comprehension and production, and these predictors were stronger for matched mother-infant pairs than for randomized pairings of mothers and infants. Another study used dense data collected for a single child, finding that word production was predicted by frequency of a word in the input and by distinctiveness of a word's spatial, temporal, and topic distribution (Roy, Frank, DeCamp, Miller, & Roy, 2015).

The above factors illustrate that words can be easier or harder to learn in a few distinct ways: some predictors relate to the sheer number of opportunities to learn the word, some relate to the conceptual accessibility of the word meanings for young children, and some relate to the position of words within a linguistic context. Within the last category, only a few simple predictors have been evaluated. These show that simplicity of the sentential context, distinctiveness of contexts, and placement at utterance boundaries, all positively predict word learning. However, data-driven distributional representations of word usage patterns have not been evaluated as predictors of AoA, even though distributional models have been proposed as mechanisms for infants to learn about word class (Monaghan, Chater, & Christiansen, 2005; Clair, Monaghan, & Christiansen, 2010) and verb semantics (Laakso & Smith, 2007), and can successfully represent semantic similarity and categorical structure when trained on infant-directed speech (Huebner & Willits, 2018).

The goal of the current study is to investigate whether, and how, distributional properties of words contribute to predicting age of acquisition. Distributional features are derived from co-occurrences among words or word sequences, but might reflect, to varying degrees, different types of information including syntactic, thematic, and taxonomic relations among words (Huebner & Willits, 2018). Therefore, a second goal of the study is to design and extract a set of distributional features from real-world language use that segregates information types, so that the contributions of these types to lexical development can be evaluated separately. One simple way to segregate distributional features into two complementary streams is by tracking distributional statistics at different time scales. That is, a word's distribution can be described both in terms of *adjacency relations* such as transition probabilities or simple constructions across successive words, and in terms of *non-adjacent co-occurrence* with other words in a wider scope. We expect this simple distinction to correlate with some higher-level conceptual boundaries: specifically, we expect that adjacency relations will primarily capture word class and syntactic information, whereas non-adjacent co-occurrence will primarily reflect thematic information. In the next section, we review the strengths and limitations of existing methods of generating distributional semantic representations of words.

Distributional Word Representations

General-purpose distributional lexical representations have a history of effective use in natural language processing, and therefore serve as a starting point for the design of developmentally relevant distributional representations. Three main approaches can be distinguished. One family of models represents words by their distribution across large-scale contextual units. These contexts can be defined by document boundaries in the training corpus, as in latent semantic analysis (Dumais, 2004), or they may represent topics inferred by a generative model, as in latent Dirichlet allocation (Blei, Ng, & Jordan, 2003). A second group of models represents words by their co-occurrence rates with other words. This approach was introduced as the Hyperspace Analogue to Language (Lund & Burgess, 1996), and statistical refinements such as COALS (Rohde, Gonnerman, & Plaut, 2004). Recently, new levels of performance on large datasets have been achieved by recasting this approach as a neural network, as in Skip-gram (Mikolov, Chen, Corrado & Dean, 2013; Mikolov, Sutskever, Chen, Corrado & Dean, 2013). Finally, a third group of models learns to predict sequences of words, such that its internal states after training can be interpreted as word representations. This group includes recurrent neural networks (Elman, 1990) and their refinements such as LSTM (Hochreiter & Schmidhuber, 1997), Gated Recurrent Unit (Chung, Gulcehre, Cho, & Bengio, 2014), and Delta-RNN (Ororbia, Mikolov, & Reitter, 2017). Despite their successes, existing distributional models suffer from a few drawbacks. Models that represent words as points in a vector space have trouble accounting for the fact that pairwise word similarity is not a valid metric: word A may be similar to word B, and word B may be similar to word C in an unrelated way, without implying that words A and C are at all similar (e.g., Barclay, Bransford, Franks, McCarreli & Nitsch, 1974; Griffiths, Steyvers, & Tenenbaum, 2007; Medin & Shoben, 1988). Another potential problem is that computational models often achieve good performance by processing vast amounts of data at arbitrary timescales and memory loads, which are unrealistic models of human infants' learning, and might be counterproductive for certain kinds of learning (Elman, 1993; Newport, 1990; Phillips & Pearl, 2015).

Corpus linguistic studies from a developmental perspective have demonstrated that adjacent distributional regularities encode rich information about syntax. For instance, many *frequent frames,* consisting of pairs of flanking words for a target word, are strongly selective for nouns and verbs, and have been proposed to form the basis for early syntactic learning (Mintz, 2003). Frames are an attractive measure for predicting AoA of specific words, for several reasons. They can be detected in corpora of child-directed speech, and they encode information about word class and usage without requiring advanced syntactic knowledge. Frames might support word learning in several ways. Frequent frames likely facilitate word segmentation by demarcating the boundaries of the framed word, increasing the probability that it will be further processed. Another possibility is that frames help children infer word meaning by embedding words in familiar and meaningful constructions (see Goldberg, 2003; Gleitman, 1990). A third possibility is that frequent frames might facilitate production if infants acquire productive language in a construction-specific manner (Tomasello, 2000; Cameron-Faulkner, Lieven, & Tomasello, 2003): if producing a word depends not only on inferring its meaning, but also on having a construction in which it is known to occur, then the words that appear most often in the earliest-acquired productive constructions are likely to be produced first. Finally, appearing in one or a few specific frames might be a proxy for a word's syntactic class. In sum, features derived from word-frame co-occurrence statistics merit closer study because frames are thought to be relevant for infant learning, they address open questions about interactions between syntactic and lexical learning, and they can be derived exclusively from the words immediately adjacent to each target word.

In contrast to adjacency relations such as frames, it has been noted that models based on word-co-occurrences, such as Skip-gram, tend to be biased toward thematic rather than syntactic or taxonomic information (Huebner & Willits, 2018). To further emphasize thematic information at the expense of syntax, a model based on word co-occurrence can be trained using a sliding window that counts co-occurrences between nearby words, but *not* adjacent ones. The representation modified in this way is a simple, associative model of coarse-grained distributional information whose input is complementary to that of frames.

Because co-occurrences are derived from bottom-up distributional statistics in infant-directed speech, they reflect the structure of naturalistic language environments more directly than theory-based measures of grammatical class and word meaning that have been used to predict acquisition (e.g. Braginsky et al., 2016; 2017). Thus, taken together we expect frames and non-adjacent co-occurrences to reflect primarily syntactic and thematic aspects of word meaning, respectively, and to generate a rich set of quantitative features that jointly predict AoA better than either level of information alone.

Current Study

The current study addresses how distributional properties of words predict AoA of individual words. We derive quantitative features from a large corpus using two types of information: *syntactic features* are derived from distribution of words in frames, and *thematic features* are derived from non-adjacent co-occurrences between words. We first demonstrate that these two types of features succeed in capturing different levels of structure in the lexicon, both qualitatively and by quantifying the extent to which they distinguish between syntactic and thematic word classes as defined by the groupings on the CDI. Then, we evaluate how well normative AoA of words is predicted by the two types of distributional features, using CDI production norms as the primary measure of AoA. We also used the features to predict CDI comprehension norms to increase generality and comparability with previous studies. We fit models predicting AoA using a set of simple word-use metrics that have been previously described (i.e., frequency, MLU, final frequency, solo frequency; Braginsky et al., 2016). We then investigate how these models are improved by including the syntactic and/or thematic distributional features.

Methods

Corpus

The corpus was constructed by downloading all American English transcripts from CHILDES (MacWhinney, 2014). These were narrowed down to those containing speech by primary caregivers directed to typically-developing children aged 48 months or younger. Special codes indicating non-words were removed. All punctuation was removed except for utterance boundary tokens. Dialectal, spelling, word segmentation, contractions, and transcription-style variants were standardized. Whenever possible, we attempted to standardize variants so that each CDI item corresponded to a single word type in the corpus. Finally, we removed common inflections and contractions (*-ing, -ed, -s, 'll, 've, 'm, 're, -n't*), leaving word lemmas. The resulting corpus contained 952,564 utterances and 5,083,634 tokens from 3377 transcribed sessions (range: 2 - 11,722 tokens per recording session) of 2093 children (range: 2 - 462,442 tokens per child).

2.2. Age of Production

CDI data were obtained from Wordbank (Frank et al., 2017). American English Words and Gestures (age 8-18) and Words and Sentences (age 16-30) were both used. For each age, the proportion of children producing the word was calculated. We then fit a logistic curve to the proportion, constraining the proportion to approach 1 eventually. The age at which this curve crossed .5 was considered the age of production for the word. For model fitting, values were normalized to have zero mean and unit variance. Three items were dropped because they consisted of family-specific names; nine were dropped because they were phrases that could not be segmented as distinct units in the corpus; and two were dropped because they are sex-specific body parts and were very infrequent in the corpus. This left scores for 656 words.

Age of Comprehension

Comprehension scores were calculated the same way as production scores, except that only the shorter Words and Gestures form includes a comprehension checklist (see Fenson et al. 2007). Thus, data were available for a smaller set of words and ages. In addition to the items that were dropped from the production dataset, logistic curves could not be fit to four words (brother, sister, mommy, daddy) because no increasing trend was present over the range 8-18 months. This resulted in scores for 383 words.

Syntactic Features

Frequent frames were defined as pairs of words (including the utterance boundary marker, so that the framed word could be in utterance-initial or -final position) that co-occurred, separated by one word, at least 1000 times in the corpus. This cutoff was chosen to eliminate most frames involving content words specific to a theme or activity, and those that occurred too infrequently to be familiar to most children. This yielded 403 frequent frames. All other frames were collapsed into a single "other" frame, representing rare or unknown frames. For each target word, we then constructed a vector of the number of times the word occurred in each frame. Thus, 656 words' occurrences in 404 frames were counted, yielding a 656 x 404 matrix. Values were first normalized so that each word's features (i.e., frame-occurrences) summed to 1, and then these features were scaled to have a mean of zero and variance of 1. Each row of the resulting matrix thus consisted of a 404-dimensional representation of a word, where each dimension represents its tendency to occur in a specific frame. To generate a smaller set of features, we applied principal components analysis (PCA). This operation produces a set of abstract features representing combinations of frames that best explain the variability in the data, subject to the constraint that they are uncorrelated with each other. For instance, all frames that allow mainly nouns are likely to contribute to a single principal component (PC) that represents, roughly, the affinity of a word for “noun” contexts. The resulting principal components (PCs) are ordered by amount of variance explained, and by selecting the first few PCs, we obtain a low-dimensional set of syntactic features that capture the most divergent and systematic differences in the typical frame-contexts of different words. Finally, to improve robustness of regressions, PCs were winsorized, clipping values more extreme than 10 times the mean absolute deviation from the median. We focus on the first 10 PCs for analysis.

Thematic Features

Thematic (i.e., non-adjacent) features were generated using a version of the COALS model, based on correlations between words (Rohde et al., 2004). This model was chosen for its simplicity and lack of domain- or task-specific assumptions, and because it represents co-occurrences between words as statistical associations, in accordance with statistical learning theories of language acquisition (Romberg & Saffran, 2010; Smith, Suanda, & Yu, 2014; Erickson & Thiessen, 2015). Context words were defined as any word that occurred at least 1000 times in the corpus (this cutoff eliminated rare context words that were unlikely to be familiar to children). 1376 distinct context words were thus identified. For each target word, we counted the number of times it co-occurred with each context word, using a sliding window to count co-occurrences that were within five flanking words but not adjacent. This resulted in a 656 x 1376 matrix. Co-occurrence frequencies were then normalized to represent correlations, and negative values were set to zero (because the model is only interested in finding regular associations between words). Finally, values were square-root transformed (this transformation is mathematically arbitrary, but as Rohde et al. (2004) note, it increases the weight of weak associations between words relative to strong ones, thereby increasing sensitivity to patterns in a limited dataset, and it puts the values on a more interpretable scale). Each row of the matrix thus consists of a 1376-dimensional representation of a word, where each dimension represents its tendency to co-occur non-adjacently with a specific context word. As above, we applied PCA to produce abstract features representing combinations of context words that explain the most variability among target words, and PCs were winsorized in the same way. This produces a set of 10 thematic features that are comparable in structure, but complementary in content, to the syntactic features.

Other Metrics

Following Braginsky et al. (2016), we computed frequency, mean length of utterance (MLU), solo frequency, and final frequency. Frequency was calculated as the logarithm of the number of times each word appeared in the corpus. Stemming and standardization described above ensured that this frequency represents that of the word overall and not just the specific form listed in the CDI. MLU was calculated as the mean number of words in utterances containing each word, where utterance boundaries were defined by the original transcription, and did not always correspond to grammatical sentences. Solo frequency and final frequency were calculated by taking the logarithm of the number of times each word appeared in a one-word utterance or as the last word of an utterance respectively, and finally taking the residual with respect to log frequency.

Evaluation

We first evaluate the extent to which syntactic and thematic features encode distinct information. We investigate this qualitatively by reporting the frames/context words that contribute most to the top PCs, and the words with the highest and lowest values on each PC. We then quantitatively evaluate the extent to which each PC distinguishes among either syntactic or thematic word categories, as defined by the groupings on the CDI, and confirm that the syntactic and thematic distributional features differentiate the syntactic and thematic word groups, as hypothesized.

We then fit multiple linear regression models predicting both age of production and age of comprehension for the target words. First, a baseline model included frequency, MLU, solo frequency, and final frequency. Next, a full model including all features as predictors. Finally, we fit models with the baseline and syntactic features, the baseline and thematic features, and the syntactic and thematic features, leaving out one feature set at a time. Models are evaluated using R^2 and adjusted R^2, and significance for each predictor set was evaluated using likelihood ratio tests between the full model and the model without the predictors. Finally, predictive robustness was evaluated using 10-fold cross-validation to estimate the root mean squared error for predicted AoA of words not used to fit the model parameters.

We additionally repeated the analysis with different values (increasing or decreasing by a factor of 2) for minimum feature frequency, number of PCs, and winsorization levels to ensure that the overall model fits are not strongly dependent on the specific values chosen.

Results

We examined the top principal components of the syntactic (adjacent) and thematic (non-adjacent) distributional features. First, we examined scree plots to validate the number of PCs selected for each feature type (Figure 3.1). These plots show the proportion of the total variance in the raw features that is explained by each additional syntactic and thematic principal component. Inspection of the plots shows that the proportion of variance explained levels off around 5 to 10 PCs, slightly more sharply for the syntactic features. Therefore, we conduct the main analyses using the first 10 PCs, but also compare the results using 5 or 20 PCs for robustness.

Next, we inspected the top PCs to determine whether they qualitatively capture mainly syntactic and thematic distinctions, respectively. Table 3.1 shows the words with lowest and highest values, and the frames with lowest and highest weights, for the first ten syntactic PCs.

Examining the sets of words frames qualitatively, PC1 separates pronoun frames from noun and verb frames, and PC2 separates nouns from verbs. Thus, the first two PCs encode distinctions among three major word classes. PC3 selects for modal verbs that appear between a pronoun and a verb. PC4 distinguishes between modal verbs and question words. PC5 seems to encode the difference between *you* and other pronouns. PC6 selects for adjectives, PC8 selects for helping verbs, and PC10 seems to encode the difference between *don't* and other verbs. PC7 and PC9 are difficult to interpret.

Similarly, Table 3.2 shows for the first ten thematic PCs, the words with lowest and highest values, and the context words with lowest and highest weights. PC1 separates words that appear in food contexts from words that appear in narrative contexts (some of the extreme values occur for low-frequency words with spurious associations, but the context words are more stable). PC2 selects for the topic of animals; PC3 and PC4 collectively distinguish between object play contexts and talk about the past. PC5 separates object play from interpersonal play contexts. PC6 selects for body-related contexts; PC7 separates food from object-describing contexts; PC8 separates clothing from vehicle contexts. PC9 seems to encode additional distinctions between animal-related and narrative contexts, and PC10 selects for colors. Taken together, the syntactic features reflect salient aspects of word class, whereas the thematic features represent different topics and activities. We next verified that the syntactic and thematic PCs quantitatively discriminated between different syntactic and thematic groups of words, respectively. For this purpose, we used the word categories on the CDI. We collapsed the categories into a subset that primarily reflects syntactic distinctions, and a subset that primarily reflects thematic distinctions. The "syntactic" categories were quantifiers, locations, helping verbs, connecting words, descriptive words, action words, pronouns, question words, and the eleven noun categories (combined into a single set). Time words, sounds, and games/routines were dropped because they were not syntactically homogeneous. The "thematic" categories were the eleven noun categories taken separately: vehicles, animals, body parts, food/drink, people, outside, toys, furniture/rooms, household objects, places, and clothing. For each PC, we calculated the F-statistic reflecting the ratio of between-category and within-category variance for each category set. Figure 3.2 shows the F-statistics for the first 10 PCs in each feature type for both category sets. Frame features reflected syntactic categories significantly better than thematic categories (median = 50.9 for syntactic categories, median =

9.9 for thematic categories; Wilcoxon test, $p < .001$; rank-biserial correlation = .96), and co-occurrence features reflected thematic categories significantly better than syntactic categories (median = 8.3 for syntactic categories, median = 28.8 for thematic categories; Wilcoxon test, $p < .005$, rank-biserial correlation = .64). For each feature type, 9 out of 10 PCs were consistent with this distinction. This confirms that adjacent context better predicts syntactic class, whereas non-adjacent context better predicts thematic relations within a syntactic class (nouns).

Having established that the two feature sets differentiate distinct aspects of word semantics, we then used the distributional features to predict age of production and comprehension of words using linear regression models. For both production and comprehension, we evaluated a baseline model using frequency, MLU, solo frequency, and final frequency, a full model including all features, and models with only the baseline and syntactic features, the baseline and thematic features, and the syntactic and thematic features, leaving out one set of features at a time. Models were evaluated using R^2, adjusted R^2, and root mean squared error (RMSE) under 10-fold cross-validation (Table 3.3).

For both comprehension and production, the full model (RMSE = 2.83 months for comprehension, 2.49 months for production) performed substantially better than the baseline model (RMSE = 3.31 months for comprehension, 2.84 months for production). Moreover, for both comprehension and production, the model fit was significantly degraded when any of the three feature sets was removed (likelihood ratio tests, *p*s < .001). For both comprehension and production, removing the baseline features caused the biggest increase in prediction error relative to the full model. However, the relative importance of syntactic and thematic features depended on the AoA measure: for comprehension, prediction error was higher for the model without thematic features (RMSE = 3.01 months) than for the model without syntactic features (RMSE = 2.97 months), whereas for production, prediction error was higher for the model without *syntactic* features (RMSE = 2.68 months overall, 2.71 for the subset of words that are on the comprehension form) than for the model without thematic features (RMSE = 2.57 months overall, 2.51 months for the comprehension-form subset). This pattern suggests that syntactic features are relatively more important for production, whereas thematic features are more important for comprehension. Finally, error was higher for comprehension (RMSE = 2.83 months) than production (RMSE = 2.49 months overall, 2.46 on the comprehension-form subset), though it is unclear whether this is due to weaker predictive relations or to noisier data due to the difficulty of assessing comprehension in infants. Predictions of the full model of age of production and comprehension of individual words were plotted against actual AoA values for the production corpus and the comprehension corpus (Figure 3.3).

To further characterize the contributions of individual syntactic and thematic features to the AoA values, we also report the estimated weights for each feature in the full models in Tables 3.1 and 3.2. These are visualized in Figure 3.4 for production (top) and comprehension (bottom) data. Accuracy of estimation of the contribution of each individual feature is decreased when predictors are correlated; therefore, we calculated the variance inflation factor for each feature in the full production and comprehension models (Tables B.1, B.2). VIFs ranged from 1.07 to 7.34 in the production model, and from 1.19 to 5.98 in the comprehension model, indicating a moderate amount of multicollinearity.

Finally, to test whether our results were dependent on the specific choices for numerical parameters, we repeated all analyses with different values (increasing or decreasing by a factor of 2) for minimum feature frequency, number of PCs, and winsorization levels. The overall pattern of model fits was similar in all cases (Tables B.3, B.4, B.5). Performance was slightly degraded by increasing the minimum feature frequency, suggesting that less-common items add noise to the PCs. Performance was also slightly degraded by increasing the winsorization threshold, suggesting that a few extreme PC values can obscure the information encoded in smaller differences among other words. Changing the number of PCs had different effects depending on the model: Decreasing the number of PCs from 10 to 5 resulted in worse performance of all models, although the model with baseline and syntactic features was relatively less affected, which is consistent with the smaller coefficients on syntactic features 6-10. Increasing the number of PCS from 10 to 20 improved predictions slightly for production, but degraded predictions for comprehension. However, increasing the number of PCs did not affect the pattern of relative performance among models. Thus, models using 10 PCs avoid overfitting while encoding most of the relevant information in the word usage distributions.

Discussion

In the current paper, we introduce a novel distributional representation of word usage in child-directed speech. By deriving separate representations of words' distribution over frames and (non-adjacent) co-occurrences between words, we produce two sets of features that capture word usage patterns at different timescales. These features primarily encode syntactic and thematic information, respectively, which we confirm both by qualitative inspection of the feature content and by comparing feature values across syntactic and thematic word groups. Furthermore, the features are consistent with a simple model in which learners track the co-occurrences of frequent items in naturalistic input.

We then examine the degree to which each feature type predicted children's normative age of acquisition (AoA) of English words, compared to a baseline model that includes previously identified word usage features including mean length of utterance and three frequency-based measures (Braginsky et al., 2016). The baseline features, especially word frequency and solo frequency, were consistently among the strongest predictors in all models, suggesting that a word's distribution over usage contexts complements but does not supersede overall frequency and frequency in salient contexts. Nonetheless, for both production and comprehension, both distributional feature types predicted AoA over and above the baseline model. The syntactic and thematic feature types were complementary, in that a model that includes both feature types predicted AoA over and above either feature type alone. Together, syntactic and thematic distributional features explained 15% more variance in production AoA than the baseline model, and they explained 27% more variance in comprehension. These improvements were robust both to corrections for the number of parameters, and to cross-validation.

Although comprehension and production showed similar patterns of incremental predictive utility for distributional features over and above frequency-based (baseline) features, age of comprehension was predicted less accurately overall than age of production. It is likely that comprehension data are inherently noisier due to limitations in parent report accuracy (Feldman et al., 2000; Oliver et al., 2003; Eriksson, Westerlund, & Berglund, 2002; Tomasello & Mervis, 1994). Syntactic features were a stronger predictor for production, whereas thematic features were a stronger predictor for comprehension. This suggests that availability of syntactic constructions (and related cognitive resources such as verbal working memory) might be a limiting factor in children's acquisition of words for production (Arnon & Clark, 2011). The relative influence of individual features was nevertheless mostly consistent between comprehension and production (see Figure 3.4).

Different types of distributional features contributed to AoA in different ways. Frequency, solo frequency, and final frequency were consistently strong predictors, which suggest that a word's inherent salience in the input may contribute to learning rate, possibly in several ways -- for example, solo frequency eliminates the problem of segmentation and suggests that the word can be meaningful without requiring the learner to represent complex relations among multiple entities (Brent & Siskind, 2001). Consistently with this, concrete nouns and interactive social words (e.g. "hi") are prevalent among children's first learned words, as are frozen multi-word phrases that express a discrete meaning and segment as a unit (Lieven, Pine & Barnes, 1992; Bates et al., 1994). In contrast, final frequency suggests a role of perceptual salience and/or working memory limitations in learning (Fernald & Mazzie, 1991). Among the syntactic features, the strongest effects indicated that content words (nouns and verbs) were learned earlier than pronouns, and that question words were produced late, consistent with previous studies of vocabulary composition by word class (e.g. Caselli et al., 1995; Bates et al., 1994). Among the thematic features, the strongest effects indicated that words used in food contexts were learned earlier than words used in narrative or descriptive contexts, and that words used in face-to-face interpersonal play were learned earlier than words used in object play. These patterns are consistent with the limited sentence complexity of early production, as well as the well-established developmental progression from dyadic to triadic and finally displaced reference (Adamson & Bakeman, 1991; de Barbaro, Johnson, Forster, & Deák, 2016; Sachs, 1983; Morford & Goldin-Meadow, 1997).

The predictive utility of these different types of distributional features derived from naturalistic caregiver speech, in addition to frequency-based features, suggests that distributional contextual information might play a role in infants' word learning. This identifies a gap in much of the research on word learning, which has focused more prominently on the referential transparency of object-naming events (Trueswell et al., 2016; Medina, Snedeker, Trueswell, & Gleitman, 2011; Yu & Smith, 2012), social cues in naming (Frank, Tenenbaum, & Fernald, 2013), and children's biases about word (mostly noun) meanings (Markman, 1990; Deák, 2000). Thus far, these explanations of word learning have been complementary, in that the relevant factors are typically manipulated experimentally but are not assessed or estimated from naturalistic data. Conversely, frequency and distributional data can be computed from corpus data but not assessed experimentally in real-time learning. Furthermore, because experimental and distributional factors have typically been evaluated independently, it has been difficult to meaningfully compare the relative importance of each type of factor, alone or jointly. Thus far, this methodological divide has prevented us from understanding interactions among factors in word learning. However, new experimental tests of distributional factors identified in the current study and related work, as well as dense recording of social cues and visual-speech co-occurrences enabled by new technology (Smith, Yu, Yoshida, & Fausey, 2015) should make it possible to estimate word learning as a function of multiple relevant factors in the local embodied social environment as well as the more protracted accumulation of distributed linguistic patterns.

Our bottom-up approach to deriving word features aims to be less theory-laden and assumption-dependent than methods that attempt to characterize word syntax or meaning directly, and by representing words based on their usage in actual child-directed speech we can be more confident that the features are observable by infants and children. Of course, many other types of word features vary with syntactic class–e.g., nouns tend to be more concrete than other words (Brysbaert, Warriner, & Kuperman, 2015). Words also vary in their degree of perceptual grounding. For example, early-acquired nouns tend to refer to objects that occur frequently in infants' visual environments (Clerkin, Hart, Rehg, Yu, & Smith, 2017), and object-naming events vary in their referential clarity (Cartmill et al., 2013). Social cues, such as caregivers' gaze and pointing may also vary in frequency across different words and constructions (Murphy, 1978; Chang & Deák, 2019; Mason, Kirkpatrick, Schwade, & Goldstein, 2018), and words vary in their spatial and temporal distinctiveness (Roy et al., 2015). In our view, it is unlikely that distributional features exert their influence directly and independently of other cues. Instead, they provide us with a way to identify those contexts and constructions that most reliably afford word learning. Future research might then determine how those events support learning. It is likely that no single information source dominates; rather, the current results are consistent with the possibility that multiple correlated features in different modalities and in the language input conspire to distinguish groups of words. (Yu & Ballard, 2007; Sahni, Seidenberg, & Saffran, 2010; Bhatt, Wilk, Hill, & Rovee-Collier, 2004). Indeed, it might be the quantity of correlated features, rather than their specific type, that is most important. Thus, it might be possible to achieve comparable performance in predicting age of acquisition using independent feature sets such as distributional, acoustic, visual, and other features. As more modalities are added, redundant predictors might converge to the same set of words that occur frequently in distinctive, salient, and referentially transparent contexts.

It is important to note that distributional features may be proxies for other correlated cues, so that correlations between distributional features and AoA do not necessarily imply that children learn by tracking word context distributions. We designed our features to differentiate words by syntactic class and by thematic/activity contexts. Therefore, we cannot determine how much our effects depend on the distributional features we measured as opposed to the associated syntactic classes and semantic features of words, which children might detect in other ways. However, distributional features might help to identify precisely which syntactic and semantic distinctions matter for word learning, thereby contributing to, rather than detracting from, theories of word learning based on other types of more abstract features. Future research might further distinguish the contributions of word distributions by computing distributional features for each child based on their individual language input, rather than combining data from many children. By focusing on the *differences* in distributional features between children, it would thus be possible to eliminate any confounding effect of word features, such as grammatical class or concreteness, that are shared within a language community.

By measuring individual differences in exposure to words in syntactic and thematic contexts, future work might also identify sources of individual differences in vocabulary size and composition. It might be possible to measure the degree to which a caregiver's speech style supports word learning by characterizing how often they use words in contexts that are most strongly associated with word learning. Future research could investigate whether such measures, derived from actual associations between word usage and children's learning, can supplement, or even outperform existing summary measures of caregivers' speech quantity and style in predicting children's vocabulary outcomes (Weisleder & Fernald, 2013; Tamis-LeMonda, Kuchirko, & Song, 2014).

The current study is subject to several limitations. Our measures of caregiver word use and AoA data were taken from disjoint samples of children. The caregiver corpora come from a biased sample of situations recorded by researchers, and so it is not clear how faithfully the corpora represent natural input to English-learning infants. Conversely, the AoA estimates are flawed by virtue of their sole basis in parental reports, which have documented limitations as noted above. Another limitation is that the current study treats AoA only as a population average. It is not known whether different features increase the probability of acquisition for all words across all children equally, or if there are individual differences in style or trajectory of acquisition. Yet, it is suggestive that inter-individual variability exists in the distribution of syntactic types in children's vocabulary (Kauschke & Hofmeister, 2002; Bates et al., 1994). Moreover, children's existing vocabulary predicts which new words will be acquired subsequently (Beckage & Colunga, 2013). Therefore, understanding AoA as a non-stationary function of different types of predictors might enable better description and prediction of this variability. Another limitation related to the corpora is that we did not model differences in parental word usage as a function of children's age. Developmental changes in language input might induce associations between AoA and specific constructions or contexts. In addition, the types of contexts that predict learning might change across children's age ranges. Finally, a further limitation of the generalizability of the results is that languages other than English might not have the same degree of correspondence between syntactic constructions and frames defined by adjacent words. Frames might be expected to be less informative in languages that make greater use of inflectional morphemes and non-adjacent dependencies such as agreement, and less use of word order. Thus, although semantic and frequency predictors of AoA have been found to be largely consistent in a variety of languages (Braginsky et al., 2017; Fourtassi, Bian, & Frank, 2018), it is not known how well the distinction between adjacent and non-adjacent contextual information would generalize cross-linguistically.

Age-of-acquisition norms provide a unique opportunity to investigate how children's environment supports their acquisition of knowledge and competence. The current study demonstrates strong and detailed links between words' usage patterns and AoA. A challenge for future research is to integrate these results with studies of individual differences, computational learning theory, and laboratory word learning studies. Under this vision, the goal is to integrate micro-level causal, cognitive, and neural explanations for word learning with observed developmental trajectories within natural environments. If successful, we believe this would serve as a model for how to explain ontogenesis in complex systems generally.

Acknowledgements

Chapter 3, in full, has been submitted for publication of the material as it may appear in: Chang, L. M., & Deák, G. O. (2020). Adjacent and Non-Adjacent Word Contexts Both Predict Age of Acquisition of English Words: A Distributional Corpus Analysis of Child-directed Speech. Cognitive Science. The dissertation author was the primary investigator and author of this paper.

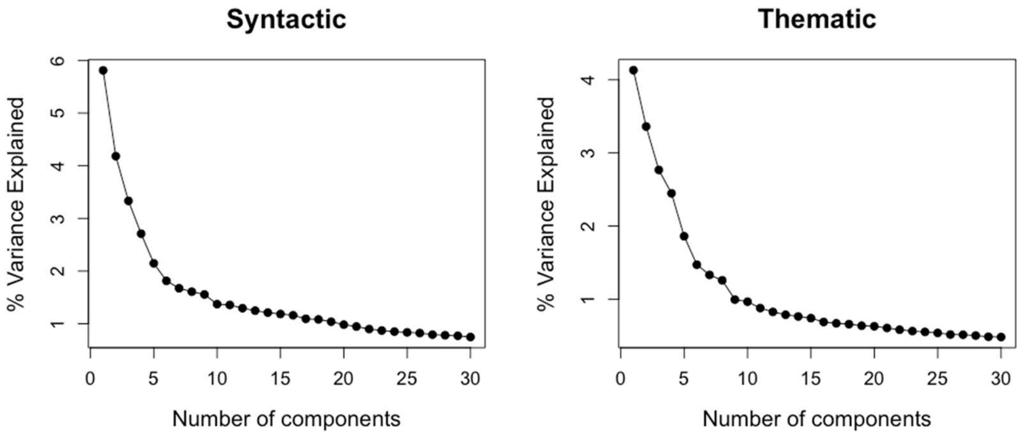

Figure 3.1. Scree plots for PCA components.

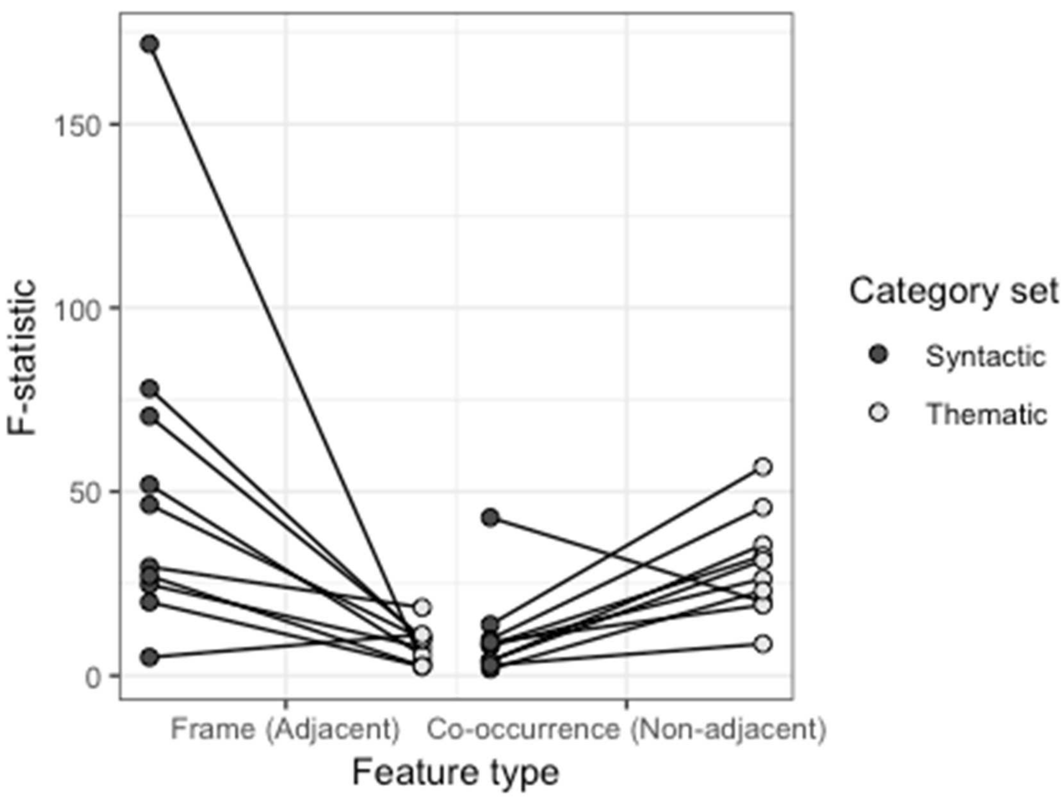

Figure 3.2. Frame versus Co-occurrence PCs segregate syntactic and thematic information. Each feature is represented as a line connecting its F-statistic with respect to the syntactic and thematic categories on the CDI. Downward sloping lines correspond to features that are more diagnostic of syntactic category, and upward sloping lines correspond to features that are more diagnostic of thematic category.

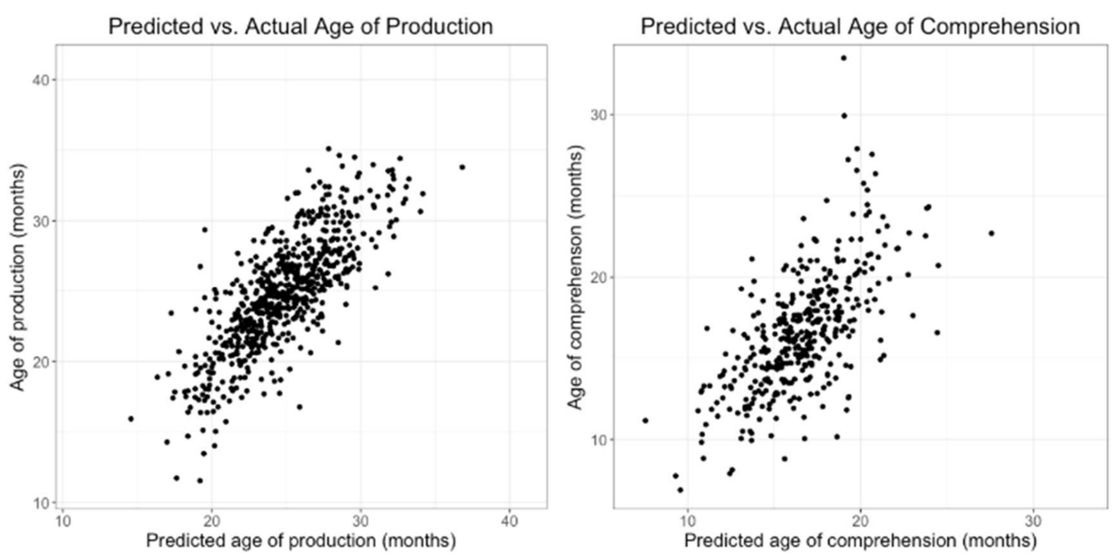

Figure 3.3. 10-fold cross-validation predicted versus actual AoA (left: production; right: comprehension) for all words.

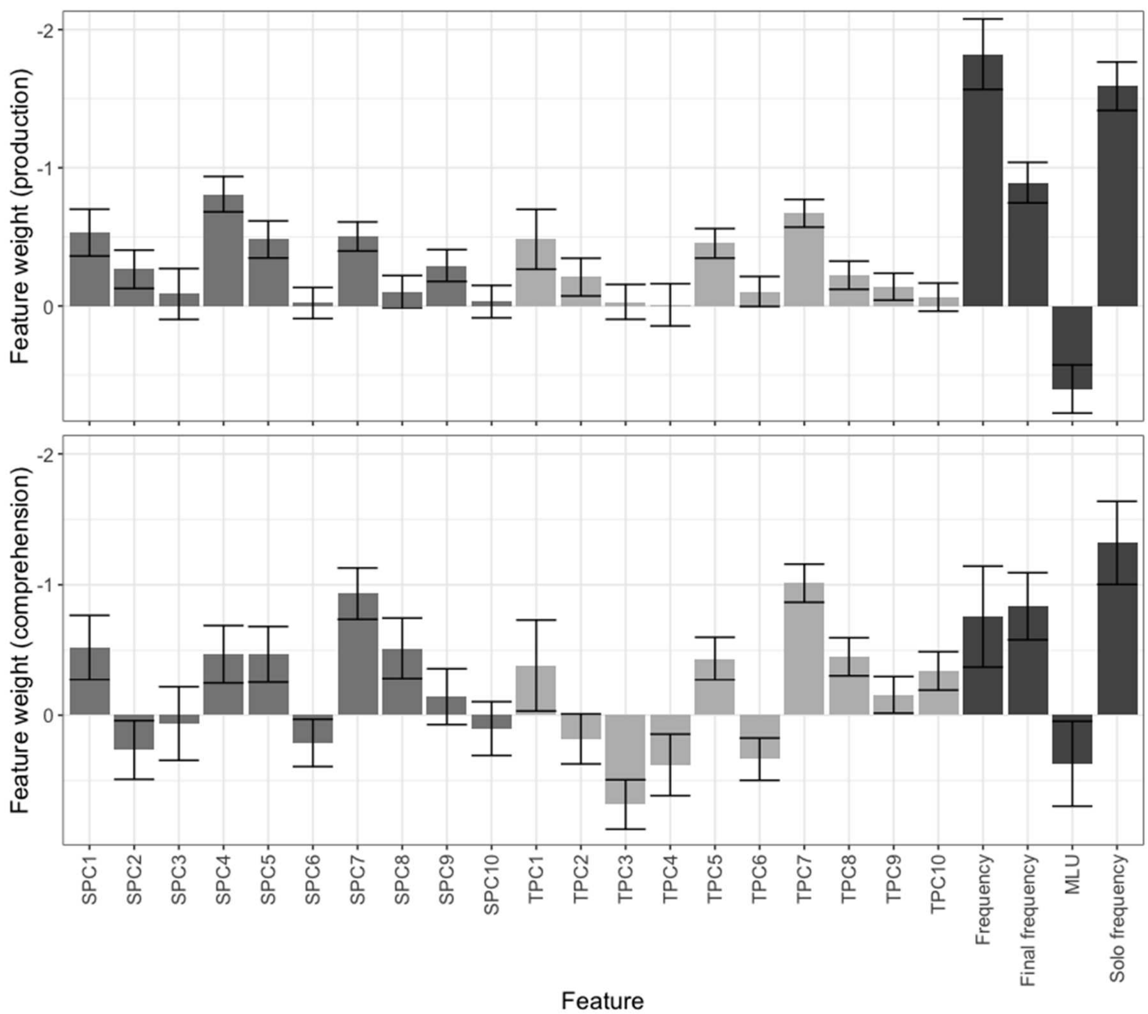

Figure 3.4. Feature weights (top: production; bottom: comprehension). Distributional feature signs are chosen so that higher feature values predict earlier production; i.e., negative weights (plotted up). Therefore, for comprehension weights, bars pointing up represent coefficients with the same sign as those for production, whereas bars pointing down represent coefficients with opposite sign. Error bars indicate standard errors of the coefficient estimates.

Table 3.1: First ten syntactic PCs. Signs are arbitrary, so “highest” is chosen for the extreme that predicts lower age of production. Coefficients are reported for the full model including baseline predictors and both sets of PCs. Significance is computed with likelihood ratio tests; **p*<.05, ***p*< .01, ****p*< .001.

PC	Highest words	Lowest words	Highest weight frames	Lowest weight frames	Production coefficient	Comprehension coefficient
1	*hear* *read* *fix* *feed*	*you* *we* *they* *i*	*you_the* *to_the* *gonna_the* *you_it*	*did_do* *so_can* *._gotta* *are_gonna*	-.53**	-.52*
2	*tights* *cheerio* *gas station* *block*	*hear* *read* *put* *fix*	*the_.* *a_.* *the_in* *your_.*	*you_the* *can_it* *wanna_the* *you_it*	-.27	.27
3	*you* *we* *they* *read*	*wanna* *gonna* *gotta* *could*	*can_put* *do_see* *do_wanna* *do_know*	*you_put* *you_get* *you_do* *you_take*	-.09	.06
4	*you* *wanna* *gonna* *gotta*	*what* *why* *where* *who*	*you_put* *you_take* *you_get* *you_do*	*._he* *._she* *._did* *._that*	-.81***	-.47*
5	*she* *he* *we* *mommy*	*you* *are* *did* *is*	*._was* *._has* *there_is* *._had*	*can_say* *do_remember* *do_think* *do_want*	-.48***	-.47*
6	*sick* *thirsty* *cute* *broken*	*what* *where* *why* *you*	*was_.* *is_.* *he_.* *they_.*	*the_.* *a_.* *you_.* *the_in*	-.02	.21
7	*we* *are* *did* *no*	*what* *you* *why* *where*	*oh_you* *._they* *what_you* *where_you*	*._do* *._is* *._did* *know_.*	-.50***	-.93***
8	*is* *does* *are* *did*	*now* *no* *thank you* *wanna*	*where_the* *what_it* *what_that* *what_he*	*._what* *._let* *._how* *._where*	-.10	-.51*
9	*is* *under* *wanna* *does*	*don’t* *why* *what* *are*	*it_the* *this_the* *go_the* *what_that*	*why_you* *you_like* *you_wanna* *i_know*	-.29*	-.14
10	*is* *build* *draw* *need*	*don’t* *show* *give* *feed*	*we_a* *you_a* *gonna_a* *to_a*	*you_wanna* *i_know* *why_you* *i_think*	-.03	.10

Table 3.2: First ten thematic PCs. Signs are arbitrary, so "highest" is chosen for the extreme that predicts lower age of production. Coefficients are reported for the full model including baseline predictors and both sets of PCs. Significance is computed with likelihood ratio tests; **p*<.05, ***p*< .01, ****p*< .001.

PC	Highest words	Lowest words	Highest weight context words	Lowest weight context words	Production coefficient	Comprehension coefficient
1	*coke* *ankle* *melon* *green bean*	*the* *he* *and* *was*	*want* *yummy* *spoon* *milk*	*and* *day* *home* *was*	-.48*	-.38
2	*cat* *quack* *woof* *who*	*you* *i* *eat* *have*	*look* *is* *rabbit* *cat*	*some* *more* *alright* *i*	-.21	.18
3	*was* *eat* *day* *were*	*put* *it* *on* *up*	*had* *um* *eat* *was*	*whoa* *whoops* *down* *oops*	-.03	.68***
4	*to* *when* *day* *time*	*that* *a* *you* *is*	*went* *when* *were* *night*	*that* *mhm* *mm* *good*	-.01	.38
5	*hi* *you* *hello* *say*	*in* *the* *put* *of*	*infant's name* *hi* *kiss* *aw*	*of* *car* *truck* *different*	-.45***	-44**
6	*he* *his* *and* *head*	*we* *you* *bye* *wanna*	*hurt* *his* *has* *leg*	*infant's name* *later* *book* *bye*	-.10	.34
7	*the* *eat* *down* *bye*	*that* *it* *i* *blue*	*milk* *high chair* *eat* *feed*	*color* *know* *pretty* *blue*	-.67***	-.1.01***
8	*put* *her* *wash* *and*	*it* *push* *car* *stop*	*pajama* *dress* *clothes* *shirt*	*crash* *stop* *noise* *it*	-.22*	-.45**
9	*did* *you* *they* *are*	*quack* *i* *moo* *say*	*yesterday* *hurt* *park* *got*	*rooster* *moo* *cow* *sheep*	-.14	-.16
10	*in* *does* *don't* *think*	*and* *up* *green* *blue*	*garage* *because* *does* *truck*	*yellow* *green* *blue* *orange*	-.06	-.34*

Table 3.3. Model evaluation for different subsets of features. RMSE = Root Mean Squared Error, in units of months (evaluated on words not used for training).

Model		R^2	Adjusted R^2	Cross-validation RMSE
Production				
	Full	.649	.636	2.49
	Baseline	.498	.495	2.84
	Syntactic+Thematic	.399	.380	3.31
	Baseline+Thematic	.571	.561	2.68
	Baseline+Syntactic	.607	.598	2.57
Comprehension				
	Full	.511	.478	2.83
	Baseline	.239	.231	3.31
	Syntactic+Thematic	.382	.347	3.16
	Baseline+Thematic	.425	.403	2.97
	Baseline+Syntactic	.391	.368	3.01

References

Adamson, L. B., & Bakeman, R. (1991). The development of shared attention during infancy. In R. Vasta (Ed.), *Annals of child development*, Vol. 8 (pp. 1-41). London: Jessica Kingsley Press.

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth–Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107-129.

Barclay, J. R., Bransford, J. D., Franks, J. J., McCarreli, N. S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior*, 13, 471-481.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., ... & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of child language*, *21*(1), 85-123.

Beckage, N., & Colunga, E. (2013). Using the words toddlers know now to predict the words they will learn next. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society.*

Bhatt, R. S., Wilk, A., Hill, D., & Rovee-Collier, C. (2004). Correlated attributes and categorization in the first half-year of life. *Developmental psychobiology*, *44*(2), 103-115.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of machine Learning research*, *3*, 993-1022.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2016). From uh-oh to tomorrow: Predicting age of acquisition for early words across languages. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Braginsky, M., Yurovsky, D., Marchman, V., & Frank, M. C. (2017). Consistency and variability in word learning across languages. Retrieved from: https://psyarxiv.com/cg6ah/

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33-B44.

Brysbaert, M., Warriner, A., & Kuperman, V. (2014b). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904-911.

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

Cartmill, E. A., Armstrong, 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*, *110*(28), 11278-11283.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*(2), 159-199.

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental psychology*, *38*(6), 967.

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

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Clair, M. C. S., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*, *116*(3), 341-360.

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Phil. Trans. R. Soc. B*, *372*(1711), 20160055.

Deák, G. O. (2000). Hunting the fox of word learning: Why “constraints” fail to capture it. *Developmental Review*, *20*(1), 29-80.

de Barbaro, K., Johnson, C. M., Forster, D., & Deák, G. O. (2016). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother–infant–object interactions. *Child development*, *87*(2), 494-512.

Dumais, S. T. (2004). Latent semantic analysis. *Annual review of information science and technology*, *38*(1), 188-230.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179-211.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71-99.

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66-108.

Eriksson, M., Westerlund, M., & Berglund, E. (2002). A screening version of the Swedish communicative development inventories designed for use with 18-month-old children. *Journal of Speech, Language, and Hearing Research*, *45*, 948-960.

Feldman, H.M., Dollaghan, C.A., Campbell, T.F., Kurs-Lasky, M., Janosky, J.E., & Paradise, J.L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child Development*, *71*, 310-322.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., & Reznick, J. S. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual*. Baltimore, MD: Brookes.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental psychology*, *27*(2), 209.

Fourtassi, A., Bian, Y., & Frank, M. C. (2018). Word Learning as Network Growth: A Cross-linguistic Analysis. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society.*

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of child language*, *44*(3), 677-694.

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.

Frermann, L., & Frank, M. C. (2017). Prosodic Features from Large Corpora of Child-Directed Speech as Predictors of the Age of Acquisition of Words. *arXiv preprint arXiv:1709.09443*.

Gleitman, L. (1990). The structural sources of verb meanings. *Language acquisition*, *1*(1), 3-55.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of experimental child psychology*, *69*(2), 133-149.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in cognitive sciences*, *7*(5), 219-224.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological review*, *114*(2), 211.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Horowitz, A. C., & Frank, M. C. (2015). Young children's developing sensitivity to discourse continuity as a cue for inferring reference. *Journal of experimental child psychology*, *129*, 84-97.

Huebner, P. A., & Willits, J. A. (2018). Structured semantic knowledge can emerge automatically from predicting word sequences in child-directed speech. *Frontiers in Psychology*, *9*, 133.

Kauschke, C., & Hofmeister, C. (2002). Early lexical development in German: A study on vocabulary growth and vocabulary composition during the second and third year of life. *Journal of child language*, *29*(4), 735-757.

Laakso, A., & Smith, L. B. (2007). Pronouns and verbs in adult speech to children: A corpus analysis. *Journal of Child Language*, *34*(4), 725-763.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive development*, *3*(3), 299-321.

Lieven, E. V., Pine, J. M., & Barnes, H. D. (1992). Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of child language*, *19*(2), 287-310.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research methods, instruments, & computers*, *28*(2), 203-208.

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57-77.

Mason, G. M., Kirkpatrick, F., Schwade, J. A., & Goldstein, M. H. (2018). The Role of Dyadic Coordination in Organizing Visual Attention in 5-Month-Old Infants. *Infancy, 24*(2), 162-186.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*(2), 158-190.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*(22), 9014-9019.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91-117.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143-182.

Morford, J. P., & Goldin-Meadow, S. (1997). From here and now to there and then: The development of displaced reference in homesign and English. *Child development*, *68*(3), 420-435.

Murphy, C. M. (1978). Pointing in the context of a shared activity. *Child development*, *49*(2), 371-380.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive science*, *14*(1), 11-28.

Oliver, B., Dale, P.S., Saudino, K.J., Petrill, S.A., Pike, A., & Plomin, R. (2003). The validity of a parent-based assessment of cognitive abilities in three-year olds. *Early Child Development and Care*, *172*, 337-348.

Ororbia, I. I., Alexander, G., Mikolov, T., & Reitter, D. (2017). Learning simpler language models with the delta recurrent neural network framework. *arXiv preprint arXiv:1703.08864*.

Phillips, L., & Pearl, L. (2015). The utility of cognitive plausibility in language acquisition modeling: Evidence from word segmentation. *Cognitive science*, *39*(8), 1824-1854.

Rohde, D. L., Gonnerman, L. M., & Plaut, D. C. (2004). An improved method for deriving word meaning from lexical co-occurrence. *Cognitive Psychology*, *7*, 573-605.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906-914.

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*, *112*(41), 12663-12668.

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. *Children's language*, *4*, 1-28.

Sahni, S. D., Seidenberg, M. S., & Saffran, J. R. (2010). Connecting cues: Overlapping regularities support cue discovery in infancy. *Child Development*, *81*(3), 727-736.

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.

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.

Sullivan, J., & Barner, D. (2016). Discourse bootstrapping: Preschoolers use linguistic discourse to learn new words. *Developmental science*, *19*(1), 63-75.

Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific English words. *Child development*, *89*(4), 1247-1267.

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

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in cognitive sciences*, *4*(4), 156-163.

Tomasello, M., & Mervis, C.B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, *59*, 174-179.

Trueswell, J. C., Lin, Y., Armstrong III, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent-child interactions. *Cognition*, *148*, 117-135.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375-382.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological science*, *24*(11), 2143-2152.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13-15), 2149-2165.

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

CHAPTER 4: Words that Share Usage Contexts in Child-Directed Speech Co-occur in Children's Vocabularies: Network Analyses of Individual Differences in Word Learning

Introduction

When children acquire a lexicon, they learn not a collection of individual associations between concepts and labels, but rather a densely interconnected set of representations in memory. Psycholinguistic methods indicate that word associations affect children's processing of related words (e.g. *apple* and *banana*) as early as 24 months (Borovsky, Ellis, Evans, & Elman, 2016). In addition to affecting language use at a particular point in time, relations among words are a significant source of complexity as well as structure in children's vocabulary acquisition over the course of development. Computational models of early word learning suggest that relations between words can facilitate word learning by constraining the search space of potential meanings (Yu, 2008) or facilitating learning of words in a category (Borovsky & Elman, 2006). Therefore, to understand how children acquire words in the real world it is necessary to consider the broader network of relations among words, and how the structure of that network emerges from and in turn influences the process of word learning.

The idea of representing related words or concepts as a network has a long history in psychology. In adults, network representations form the basis of spreading-activation theories of memory and semantic association (Collins & Loftus, 1975; Anderson, 1983) Evidence of the consequences of word relations on language processing includes phenomena such as priming (e.g., Koriat, 1981), erroneous recall (Roediger & McDermott, 1995), inhibition of learning (Tinkham, 1993; Erten & Tekin, 2008), and patterns of anomic errors in aphasia (Martin, Roach, Brecher, & Lowery, 1998) and dementia (Lerner, Ogrocki, & Thomas, 2009). It therefore seems plausible that inter-word associations and the larger network properties of these associations have pervasive effects on how we structure and utilize lexical information.

Vocabulary networks are non-randomly structured in several ways. Related words are organized in clusters around salient conceptual and thematic categories such as animals, colors, or mental verbs. Language networks typically also have a “small-world” structure (Steyvers & Tenenbaum, 2005, Milgram, 1967): that is, short paths can be drawn between any two parts of the network despite the existence of local structure. Small-world structure can arise either through sporadic connections between clusters (Watts & Strogatz, 1998) or through *hub* nodes, which have many connections and therefore enable more paths. A particular class of small-world network in which hubs play a significant role is known as *scale-free* (Barabási, 2009). This network structure, characterized by a distribution of *degree* (i.e. number of connections) following a power law, can emerge from a process of *preferential attachment* as known words “recruit” their neighbors to be added to the network (Steyvers & Tenenbaum, 2005). The short paths observed in small-world networks might facilitate search and recall in semantic memory (Beckage, Steyvers, & Butts, 2012). The combination of psycholinguistic relevance and intriguing structural diversity makes networks a useful way to formalize the interrelatedness of words. More broadly, the application of network representations is leading to advances in multiple scientific fields, often revealing surprising commonalities across domains (Barabasi, 2016; Strogatz, 2001; Watts & Strogatz, 1998).

Previous studies have used networks to characterize the structure of children’s vocabularies. One set of questions concerns the typical network structure and growth patterns of children’s vocabulary networks. As children acquire words in nonrandom orders, their individual vocabulary networks acquire additional structure beyond that inherent in the relationships among random sets of words. One way this can occur is if the earliest-learned words have different relationships than later words. For instance, Engelthaler & Hills (2017) found that words that shared semantic features with fewer other words were likely to be learned earlier. Another study found that children tended to learn words with more connections to other words in the environment (*preferential acquisition*) (Hills, Maouene, Maouene, Sheya, & Smith, 2009). Alternatively, children might learn words differentially based on their relatedness to already known words, either by learning new words that connect to known hub words (*preferential attachment)* or by learning new words that are related to many known words (Beckage & Colunga, 2019), a pattern called *lure of the associates*.

A second set of questions concerns variability in the vocabulary network structures among individual children. Do children with different word learning strategies tend to develop structurally different networks? In one study, children who showed evidence of using mutual exclusivity as a word learning strategy tended to have vocabulary networks with more hubs (Yurovsky, Bion, Smith, & Fernald, 2012). Of particular interest are differences between late talkers and children with typical language acquisition, because delayed growth in productive vocabulary is a strong predictor of later language outcomes (Moyle, Weismet, Evans, & Lindstrom, 2007; Fisher, 2017). In one study, late talkers (sometimes defined as below the 20th percentile for productive vocabulary) learned fewer related words compared to typically-developing children at the same vocabulary size, as shown by network measures including in-degree, clustering coefficient, and geodesic distance (Beckage, Smith, & Hills, 2011). However, with a larger sample size, Jimenez and Hills (2017) failed to replicate this pattern, as late talkers' vocabulary structure was similar to typically developing children. In another study, neural networks trained on the vocabularies of early-talking and late-talking children learned different biases for generalizing novel nouns based on shape and material, suggesting that the children's vocabularies encode different correlational structure (Colunga & Sims, 2017). Tracking longitudinal changes in the network structure of twelve children's vocabularies from 2 to 4 years of age, Ke & Yao (2008) found that vocabulary size and average degree, a measure of word relatedness, were not significantly related. Thus, it remains unclear how the lexical network properties of children with very different vocabulary sizes and word learning skills might qualitatively differ.

Previous studies of children's vocabulary networks suffer from several limitations. First, small sample sizes limit the precision of many findings (e.g. Colunga & Sims, 2017; Beckage, Smith, & Hills, 2011; Ke & Yao, 2008). However, as large databases of children's vocabularies have become available (Frank, Braginsky, Yurovsky, & Marchman, 2016), it is now possible to characterize variability in children's vocabulary structure much more precisely, especially for analyses that do not require longitudinal data. Second, previous studies have represented vocabulary networks using arbitrary or *ad hoc* methods to define links between words. For instance, Hills et al. (2009) and Beckage et al. (2011) used networks with directed links from one word to another if they co-occurred in that order at least once in a corpus of child-directed speech. Ke and Yao (2008) applied ordered co-occurrence separately for children's speech and for caregivers' speech to children. Hills et al. (2009) also tested networks in which two words were linked if they shared at least two features from the McRae, Cree, Seidenberg, and McNorgan (2005) feature norms, which were generated by asking adult participants to list features of nouns. Although these methods clearly align with the intuitive notion of word "relatedness," there is no reason to believe that any of them neatly captures child-relevant aspects of word relations, or is free of influence from factors that are of no relevance to young children (e.g., literary or pop-culture associations). Additionally, co-occurrence-based measures are highly dependent on word frequency, so it is unclear whether results depend on word relatedness or simply the known correlation between word frequency and age of acquisition (Braginsky, Yurovsky, Marchman, & Frank, 2016).

One way to infer word relatedness relies on the observation that children learn groups of related words at a higher rate than would be expected if words were learned independently. In the most direct demonstration of this, a statistical model using the set of words known by a child to predict which words would be acquired by the next month performed better than a model based on normative age of acquisition alone (Beckage & Colunga, 2013). Thus, learning some words is correlated with learning other words. These correlations constitute a rich source of information about language learning. A pair of words might show correlated learning if learning one word facilitates learning the other, if they are associated such that children exposed to one word are likely to be exposed to the other, or if learning them depends on a shared underlying conceptual or grammatical competence. In any case, a strong correlation for a word pair indicates that those words are related in *precisely the way or ways that are most relevant for children's word learning*. Therefore, in the present study we quantify word learning correlations and use them to define a network of relations among words. Correlated learning can be computed for any pair of words for which vocabulary data are available, which in practice means the 600+ words on the MCDI (Fenson et al., 2000). A vocabulary network can thus be constructed by drawing links between words when correlated learning of those words exceeds a certain threshold. Networks can also be evaluated for individual children by taking the subset of words known by each child.

When we define word relatedness using correlated learning, it becomes possible to turn a traditional question about vocabulary networks on its head: rather than assume a given external measure of word relatedness is relevant and use it to evaluate children's vocabulary structure, as has been done in previous studies using co-occurrence or semantic features, we can use our measure of word relatedness that *inherently* reflects relevance to word learning as a ground-truth target, and evaluate whether external measures can predict or explain it. If measures such as co-occurrence are relevant to word learning, they should be significantly correlated with the correlated-learning measure. In this way, we can evaluate the relevance of different candidate measures. Relevant predictors are likely to take the form of similarity measures in the usage or meaning of two words. However, to the extent that some measures explain correlated learning better than others, or differentially predict learning of different subsets of words, we can infer what sources of information are relevant for word learning.

A primary source of information that might explain word learning patterns is the distribution of words in natural speech (Monaghan, Chater, & Christiansen, 2005; Chemla, Mintz, Bernal, & Christophe, 2009; Clair, Monaghan, & Christiansen, 2010). Distributional information refers to quantitative representations of co-occurrence patterns of pairs or groups of words in large samples of language. These include the co-occurrence measures used in previous studies of vocabulary networks, as well as more advanced measures that use word co-occurrence statistics to represent words as vectors, e.g. HAL (Lund & Burgess, 1996), COALS (Rohde, Gonnerman, & Plaut, 2004), and Word2Vec (Mikolov, Chen, Corrado, & Dean, 2013). Thus, if children's word learning is predictable based on the distributional information encoded in the usage patterns of words, we should expect words with similar vector representations to show correlated learning. However, although vector-based distributional representations achieve state-of-the-art performance in many applications, they have some disadvantages as explanatory variables in a developmental context: they conflate syntactic and thematic word features, and in most implementations they make unrealistic demands on memory and require processing of vast amounts of input data (Huebner & Willits, 2018). In previous work, we introduced two distributional representations intended to mitigate these difficulties (Chang & Deák, submitted for publication / Chapter 3). The first representation is based on a word's distribution across "frequent frames," i.e., high-frequency contexts defined by the previous and next word token. Frequent frames are useful units of representation because they primarily encode information about a word's syntactic role (Mintz, 2003), and they do not require learners to represent rare words or long-distance dependencies. The second, complementary representation is based on co-occurrences between words in a sliding window, ignoring the immediately adjacent tokens (which were used to define frequent frames). This representation selectively ignores most syntactic information but preserves more thematic and other semantic information. To maintain child-appropriateness, the two sets of representations were trained on the CHILDES corpora of child-directed speech (MacWhinney, 2014), excluding rare words.

Because our networks depend *only* on word learning and not usage, we can then measure the relevance of distributional word usage by evaluating how closely word relations defined by distributional information correspond to those defined by correlated learning. By predicting the amount of correlated learning based on the similarity between two words' distributional representations, we can determine how well the usage patterns of words in naturalistic input explain which words tend to go together in children's vocabularies. Furthermore, we can roughly partition how much correlated learning of words is predicted by the syntax-like information contained in local frames, versus the additional semantic information encoded in broader-scope word co-occurrences. In addition, as a rough way to determine whether semantic information independent of words' distributions in child-directed speech can predict children's word learning, we used a database of free-association norms (Nelson, McEvoy, & Schreiber, 2004; Steyvers & Tenenbaum, 2005) as an additional predictor. Free associations by adult language users predict performance on psycholinguistic measures including cued recall and intrusions in free recall (Steyvers, Shiffrin, & Nelson, 2002), and they should predict correlated learning if those associations capture relevant dimensions of association between words other than those captured by distributional representations.

The current paper describes findings from two theoretically related but distinct investigations. In the first, we use the Wordbank database of toddlers' productive vocabularies (Frank et al., 2016) to compute correlated learning between pairs of words. We model semantic networks where the value of the log-odds ratio, a measure of correlated learning, defines the strength of connection between words. We describe the structure of this network over all words in the MCDI. We further demonstrate how the network structure differs between children with large vs. small vocabularies for their age, using several quantitative measures of network structure. In the second investigation, we examine how well distributional and subjective relations among words predict their correlated learning. We evaluate regression models predicting the log-odds ratio as a function of frame similarity, word co-occurrence similarity, and subjective associations from the free-association norms.

Study 1: Methods

In this section, we describe how vocabulary data was used to define the connectivity between words in terms of correlated acquisition by American English-learning children, and how the resulting pairwise connections were used to characterize the network structure of each child's productive vocabulary.

Vocabulary Database

CDI data were taken from Wordbank (Frank et al., 2016). The American English Words and Sentences form (word production for children aged 16-30 months) was used. Separate entries for different meanings of the same word were combined. Additionally, three items were dropped because they were family-specific names; eight were dropped because they were multiple-word phrases, and two were dropped because they were sex-specific body parts and were rarely used in the CHILDES corpus. This left 654 words for which production data were analyzed for each child. The database contained 5520 vocabulary checklists for 4867 unique children.

Additionally, we computed the normative age of production of each word using both the Words and Sentences form and the Words and Gestures form (age 8-18 months). We computed the proportion of children at each age who produced each word, then fit a logistic curve to this proportion. The curve was constrained so that the proportion approached 1 as age increased. Normative age of production was calculated as the age at which the curve crossed .5.

Log-Odds Ratio

The strength of correlated learning between each pair of words was calculated using the Log-Odds Ratio (LOR). The LOR of a word pair measures how much more likely a child is to produce both words in the pair over and above what would be expected if words were learned independently. It is defined as the difference between the log odds of producing both words in the pair and the log odds of producing both words assuming independence:

$$LOR_{ij} = logit\left(P\left(w_i \wedge w_j\right)\right) - logit\left(P(w_i)P\left(w_j\right)\right)$$

Equivalently, the probability of producing a word pair can be expressed as:

$$P\left(w_i \wedge w_j\right) = \sigma\left(logit\left(P(w_i)P\left(w_j\right)\right) + LOR_{ij}\right)$$

where $\sigma(x) = \frac{1}{1+e^x}$.

The probabilities $P(w_i)$ and $P\left(w_j\right)$ of producing each word in a pair were not constant across all children in the sample. Instead, for each child we approximated these probabilities using the observed proportion of children with the same total vocabulary size who produced the word. We then approximated the LOR by fitting the value β that maximized the likelihood of the observed data (i.e. the set of children who actually produce both words in the pair). Thus, for a word pair $\left(w_i, w_j\right)$,

$$LOR_{ij} = \begin{matrix} arg\ max \\ \beta_{ij} \end{matrix} \left[\prod_k \left(P_{ijk}\right)^{y_k}\left(1 - P_{ijk}\right)^{1-y_k}\right]$$

where y_k = 1 if the k^{th} child produces both words in the pair and 0 otherwise, and

$$P_{ijk} = \sigma\left(logit\left(p(w_{ik})p\left(w_{jk}\right)\right) + \beta_{ij}\right)$$

where $p(w_{ik})$ is the proportion of children producing word i out of all children with total productive vocabulary equal to that of the k^{th}child.

The form of the above equations is equivalent to logistic regression, so LOR was computed using the glm function in R. For a given word pair, a large positive LOR value indicates that children who produced one word in the pair were much more likely to produce the other one; a large negative LOR value indicates that children who produced one word in the pair were much less likely to produce the other one; and a LOR value near zero indicates that knowing that a child produces one word provides little to no information about the probability that they produce the other word.

Given the strength of connection between any two words, we next define several metrics that characterize the structure of an individual child's vocabulary, i.e. the words that child produces and the connections among them.

Mean edge strength

The mean edge strength is calculated as the average LOR of all word pairs in a child's vocabulary. It takes higher values if the child's vocabulary words are more closely related to each other, and it approaches zero if random sets of words are learned.

Clustering coefficient

Unlike the mean edge strength, which is calculated using the raw LOR values, the other measures are calculated using all-or-nothing edges calculated by thresholding LOR at a given value. This thresholding allows us to use standard measures from network theory to characterize the data. The clustering coefficient measures how locally connected groups of nodes are. For a single node, the clustering coefficient is defined as the proportion of that node's neighbors that are neighbors with each other. The clustering coefficient for the whole graph is computed by averaging over all nodes with at least two neighbors (i.e. wherever it is well-defined). A value of one indicates that word connections are transitive. More generally, in a network with a high clustering coefficient, a word is likely to be connected to most of its neighbors through the same shared feature, which those neighbors in turn share with each other (e.g. *cow* with *horse* and *pig*). By contrast, a low clustering coefficient can arise if many words are connected to different neighbors through different shared features or relations (e.g. *cow* with *horse* and *milk*).

Harmonic path closeness

The harmonic path closeness (harmonic centrality index; Rochat, 2009) measures how easy it is to find paths among nodes in the graph. It is closely related to the mean path length, but is well-defined when the graph is not fully connected (i.e. some node pairs have no path connecting them). Harmonic path closeness is defined as the mean over all node pairs of the reciprocal of the shortest path length (zero if no path exists). Higher values indicate that most pairs of words are connected by short paths. Small-world networks are characterized by high path closeness and high clustering coefficient.

Degree distribution skewness

The degree distribution skewness measures the prevalence of hubs. Whereas a scale-free network should exhibit a degree distribution fit by a power law, other distributions can also be described by the prominence of the right tail, composed of hub nodes. Thus, without making assumptions about the shape of the degree distribution, higher right skewness implies the presence of one or more significant hubs.

Study 1: Results

Log-odds ratio network

The LOR can be interpreted as a pairwise measure of relatedness among words. Structure in these relations can be examined qualitatively by visualizing the LOR values as a matrix and by using hierarchical clustering to identify clusters of words that are related by the LOR metric (Figure 4.1).

Hierarchical clustering using complete linkage in R reveals groupings of words that correspond well with intuitive notions of relatedness. The top-level clusters roughly align with nouns, verbs, function words/relations, and other words. Within these, more focused clusters of thematically and functionally related words are found. Two examples are shown in Figure 4.2.

The LOR can also be represented as an undirected graph, where words are connected if their LOR exceeds a threshold value (Figure 4.3). This threshold can be chosen to fix the overall edge probability to a given value; additional examples of more or less dense graphs constructed using this method are given in Figures C.1, C.2, and C.3.

Differences in network structure between early and late talkers

For each network measure (mean edge strength, degree centralization, clustering coefficient, harmonic path closeness), we investigated whether the measure’s values differed between early and late talkers. Children were divided into four groups based on whether their total vocabulary size was in the first, second, third, or fourth quartile for their age (in months, range: 16-30). Thus, the groups partially overlapped in total vocabulary size.

All measures depended systematically on vocabulary size, which is expected based on the structural properties of graphs as the number of nodes varies, but might also be influenced by differences in the sets of words learned early versus late in vocabulary acquisition. Therefore, it was necessary to correct each measure for vocabulary size before comparing measure values across vocabulary quartile groups. First, to compare across quartile groups, we restricted the analysis to children with vocabulary sizes for which data were available in every quartile group (at various ages). Therefore, we analyzed vocabularies between 74 to 451 words, which included 2510 of the 5520 total vocabularies. The large sample size made it possible to control for vocabulary size nonlinearly. This was done using kernel regression, with Gaussian kernel bandwidth set to 10 words and reweighted so that at each point, the total weights from points in each quartile group were equal. This reweighting prevented differences in the density of points of each group at different vocabulary sizes from distorting the correction. Measure values were adjusted by subtracting the kernel-regression line from the raw values. For multiple measures, the relationship with vocabulary size was strong and nonlinear, and thus any effects of quartile group are detectable only after controlling for vocabulary size. The results were robust to the exact form of this correction; similar results were obtained using kernel regression without reweighting and with a kernel bandwidth of 5, 10, or 20 (see Table C.1).

After correcting for total vocabulary size, measure values were compared across quartile groups using one-way ANOVA (Figure 4.4). The analysis was repeated using thresholds that give 5%, 10%, and 15% edge density. To show differences most clearly, for each metric we report only the threshold that results in the largest effect of vocabulary quartile (i.e. the largest F-score). Therefore, a Bonferroni correction was used to correct for the three comparisons that were run to select the threshold. In addition, linear regression of measure values on the quartile group's numerical value was used to test for ordinal trends. Finally, pairwise tests were conducted using the Holm-Bonferroni procedure on the comparisons between all groups at all thresholds to compute significance.

Results showed that mean edge strength differed significantly by vocabulary quartile group, $F(3, 2506) = 20.1$, $p < .001$, with a significant increasing trend, $F(1, 2508) = 54.5$, $p < .001$. Post hoc comparisons showed that mean edge strength was higher for quartiles 3 and 4 than for quartiles 1 and 2, and higher for quartile 4 than for quartile 3. This suggests that children who were relatively early talkers had vocabularies generally containing more related words than did later talkers. Clustering coefficient fit best with 5% edge density, and differed significantly by vocabulary quartile group, $F(3, 2506) = 8.1$, $p < .001$, with a significant decreasing trend, $F(1, 2508) = 12.1$, $p < .001$. Post hoc comparisons showed that clustering coefficient was lower for quartiles 2, 3, and 4 than for quartile 1. This suggests that children who were late talkers had vocabularies with more dense clusters: that is, they were more likely to learn groups of words that were all related in the same way, such that most of a word's neighbors were also neighbors with each other.

Harmonic path closeness fit best with 15% edge density, and differed significantly by vocabulary quartile, $F(3,2506) = 26.4$, $p < .001$, with a significant increasing trend, $F(1, 2508) = 75.3$, $p < .001$. Post hoc comparisons showed that harmonic path closeness was higher for quartiles 3 and 4 than for quartile 1, and higher for quartile 4 than for quartiles 2 and 3. This suggests that children who were relatively early talkers had vocabulary networks in which shorter paths were available between more word pairs than did later talkers. Degree distribution skewness fit best with 15% edge density, and differed significantly by vocabulary quartile group, $F(3, 2506) = 4.3$, $p < .005$, with a significant increasing trend, $F(1, 2508) = 11.8$, $p < .001$. This suggests that children who were relatively early talkers had vocabularies with more significant hubs than did later talkers. However, none of the pairwise comparisons reached significance.

All comparisons remained significant, with the same pattern of pairwise differences, removing repeated tests of the same child ($n = 11$). Thus, each measure indicates subtle but significant differences in the network structure between children with relatively large or small vocabularies for their age.

Discussion: Study 1

We used CDI data and the log-odds ratio to describe relatedness among words in the context of toddlers' productive word acquisition. This measure grouped words into the major syntactic classes, indicating that individual children in the sample tended to concentrate their word learning disproportionately on nouns, verbs, or other sets of related words. Further exploration of the word learning correlations revealed both strong signatures of more fine-grained taxonomic categories, such as colors and food, as well as a range of miscellaneous functional and thematic associations. Thus, the LOR is effective as a general-purpose word relatedness measure for toddlers. LOR naturally avoids dependence on factors that are irrelevant for children's word learning, and effectively weights various, possibly incommensurate, factors according to the magnitude of their effect on word learning. It also does not depend on a specific set of pre-selected features, but can potentially depend on any distinguishing properties of words or their referents. For the purpose of defining relatedness of word pairs for analysis of children's vocabulary networks, this approach is therefore more likely to identify results that are specifically relevant to word learning, compared to previous approaches using graph edges defined for convenience (e.g., Beckage et al. (2011; Engelthaler & Hills, 2017). However, the log-odds ratio also has some limitations: it can only be computed for word pairs where vocabulary data are available, and it is affected by any bias in parents' report of their children's vocabulary (Feldman et al., 2000; Oliver et al., 2003; Eriksson, Westerlund, & Berglund, 2002; Tomasello & Mervis, 1994). It also focuses solely on production, and therefore does not account for receptive vocabulary or graded levels of word knowledge.

Using the LOR graph, we investigated differences in the vocabulary composition of late-talking and early-talking children. At a given vocabulary size, early-talking children's vocabulary networks had higher mean edge strength, higher degree distribution skewness, lower clustering coefficient, and higher harmonic path closeness. In other words, children with large productive vocabulary for their age tended to know words that were more closely related among each other, more organized around hub words, less organized into locally coherent clusters, and better connected by short paths or chains of relations. Previous studies investigating network structure differences between late talkers and typical children have reached different conclusions. Beckage et al. (2011) reported that late talkers had networks with less coherent structure than typical children. However, Jimenez and Hills (2017), using a much larger sample – essentially the same vocabulary database analyzed here – found no differences between late talkers and typical children. Both of these studies used word co-occurrences as network links. Using correlated word learning instead of word co-occurrences to define links, we find support for the idea that late talkers' words are less coherent. Nevertheless, the small magnitude of our effects cast doubt on the reliability of detecting such differences in samples of fewer than 100 children, which might partly explain discrepancies across previous investigations. The small effect sizes also limit the clinical usefulness of the findings.

What can we learn from the differences in vocabulary structure between early and late talkers? One general finding is that the effects generally show a consistent directional trend across all successive quartile groups, rather than specific differences between "late talkers" or "early talkers" and others. This suggests that the differences reflect parametric variability in factors that are reflected in scalar differences in vocabulary acquisition rate across most or all children, and cannot be attributed specifically to a disordered subpopulation. The observed differences can be summarized by saying that compared with late talkers, early-talking children learn words that have many connections to other words, but do not simply fixate on one category or a few categories to the exclusion of other words. Early-talking children might learn more related words because they are more sensitive to the contextual distribution of words. For example, if children are more sensitive to taxonomic relations (Colunga & Sims, 2017) or use heuristics such as a 'one-word-per-concept' bias to disambiguate groups of associated words (Yurovsky et al., 2012) or if they know more syntactic constructions that pick out groups of words related by their common role in the sentence, they might build relatively large vocabularies that are well connected but not narrowly clustered. Thus, sensitivity to certain patterns in language input might explain why some children acquire more words. However, to the extent that children build a resulting network with strong connections and small-world structure, that structure might feed back into further word learning by accruing more diverse words to serve as potential associates for even more novel words, and thus progressively structure the lexicon to enable more efficient processing.

Study 2

Having established that correlated word learning reveals both patterns in connections among words and individual differences in children's lexical development, we next asked whether other measures of word relatedness were consistent with the structure in the word learning network. Correlations and regressions between pairwise similarity matrices have been used previously to evaluate the quantitative agreement between different sets of pairwise structure on words (Sadeghi, McClelland, & Hoffman, 2015; Devereux, Clarke, Marouchos, & Tyler, 2013; Peelen & Caramazza, 2012; Kenett et al., 2013). We evaluated the extent to which LOR was correlated with two distributional measures of similarity word use – frame overlap and context-word overlap, described below. We also evaluated whether LOR was correlated with adults' subjective word associations as measured by free-association norms (Nelson, McEvoy, & Schreiber, 2004). Free association data serve as a useful counterpoint to the distributional measures because they depend directly on the words' conventional meanings and are not specific to the information discernible from linguistic inputs to children. The Nelson et al. free association norms (Steyvers & Tenenbaum, 2005) and similar association datasets (De Deyne & Storms, 2008; Kenett, Kenett, Ben-Jacob, & Faust, 2011) have been used directly to construct semantic networks. Therefore, understanding the relationships between the word connections defined by subjective associations, distributional measures, and LOR is important for the interpretation of existing and future network analyses of children's vocabulary.

Methods: Study 2

Corpus

The corpus for computing distributional word representations was the same one used by Chang & Deák (submitted / Chapter 3), which contains all caregiver speech directed to typically-developing children aged up to 48 months in the CHILDES database (MacWhinney, 2014). Special codes and punctuation were removed except for utterance boundary tokens. Dialectal, spelling, word segmentation, contractions, and transcription-style variants were standardized, so that each word, and especially each CDI item, corresponded to a single word type in the corpus. Finally, we removed inflections and contractions (*-ing, -ed, -s, ‘ll, ‘ve, ‘m, ‘re, -n’t*). The corpus contained 952,564 utterances and 5,083,634 tokens from 3377 transcribed sessions (range: 2 - 11,722 tokens) of 2093 children (range: 2 - 462,442 tokens).

Frame Overlap

The frame overlap (FO) was designed to be a quantitative measure of similarity in the distributions of two words across utterance contexts, reflecting primarily syntactic constructions but without making specific assumptions about syntactic structures. For each CDI word, we recorded its frequency of appearance in each frequent frame, following Chang & Deák (submitted / Chapter 3). Frequent frames were defined as pairs of tokens (i.e. words or utterance boundary) that appeared surrounding a single word at least 1000 times in the corpus. 403 such frequent frames occurred in the corpus. Each word was represented as a 403-dimensional vector, where the components represented the proportion of tokens of that word that were in each frame. Then, for each pair of words, we computed the overlap by taking the sum over all 403 frames of the proportion shared between the two words (i.e. the lesser of the two values: if the two words appeared in a given frame 5% and 3% of the time, respectively, then that frame contributed 0.03 to the total overlap). Finally, FO was normalized so that each word had an average FO of 0 with respect to all other words, and the standard deviation of FO over all word pairs was 1. Thus, for a given word pair, a positive FO indicates that the two words tended to occur in the same frames (i.e. contexts defined by the immediately preceding and following words) whereas a negative FO indicates that the words tended to occur in disjoint sets of frames.

Context-Word Overlap

The context-word overlap (CWO) was also designed to measure similarity in the contextual distributions of two words. However, complementary to FO, CWO was designed to reflect primarily thematic semantic information while being relatively independent of syntax. This was achieved by representing each word's distribution using the co-occurrence with a set of context words, while selectively removing information about word transitions. Following Chang and Deák (submitted), word vectors were generated using the COALS model (Rohde et al., 2004). All word types that occurred at least 1000 times in the corpus were used as context words. 1376 context words were retained. For each CDI word and context word, we computed the number of times they co-occurred within five words of each other, but not adjacent. Ignoring adjacent co-occurrences allows the CWO to avoid redundancy with FO and limits its dependence on specific syntactic constructions. The raw co-occurrence counts were then pre-processed to improve their properties as a measure of association between words. First, to remove dependence on word frequency, co-occurrence counts were normalized to represent correlations. Next, to select specifically for positive associations, negative values were set to zero. Finally, to increase the weight of weaker but significant associations relative to the strongest ones, values were square-root transformed. Each word was thus represented as a 1376-dimensional vector, where the components represented the degree of positive association between the word and each context word. Then, similarly to FO, we computed the overlap for each word pair by taking the sum over all 1376 context words of the value shared between the two words (i.e. the lesser of the two values, representing the weaker of the two words' associations with that context word). Finally, CWO was normalized so that each word had an average CWO of 0 with respect to all other words, and the standard deviation of CWO over all pairs was 1. Thus, for a given word pair, a positive CWO indicates that the two words tended to occur in the vicinity of the same context words, whereas a negative CWO indicates that the two words tended to occur with disjoint sets of context words.

Subjective Associations

As a measure of salient subjective associations between words that is not derived from the words' usage patterns in child-directed speech, we used the free-association norms of Nelson, McEvoy, and Schreiber (2004). To construct these norms, participants were given one of over 5000 stimulus words and were asked to write the first word that came to mind that was meaningfully associated with the stimulus word. Each word was shown to an average of 149 (SD = 15) adult participants, and responses produced by at least 2 participants were recorded as connections. Free-association norms were available for 559 of the 654 words in the study (85%), which implies that connections could be evaluated for 73% of word pairs. For the current study, forward and backward connections were treated as equivalent, and words were considered connected if either stimulus produced the other as a response. We defined both the direct association (DA), which is equal to 1 if the words are connected and 0 otherwise, and the indirect association (IA), which is equal to the reciprocal of the length of the shortest path between the words, and 0 if no path exists. Mean DA and IA values were imputed for pairs where norms were not available. Finally, DA and IA were normalized the same way as other variables.

Results: Study 2

Frame overlap and word overlap were calculated for all word pairs. Figure 4.5 shows the FO and CWO as matrices, where the order of the rows and columns was set according to the hierarchical clustering on the LOR, as in Figure 4.1. The blue squares along the diagonal of Figure 4.5A show that the high-level clusters in the LOR matrix correspond closely with FO. This is expected because the high-level clusters are organized around major word classes, which are strongly reflected in frames but less so in broader contexts (e.g., occurring during the same activity). More generally, the three matrices (LOR, FO, and CWO) appear to be positively correlated, indicating that word usage patterns can explain departures from word-wise independence in children's word learning. The following analysis examines this hypothesis quantitatively.

The simple correlations among the three matrices are positive: *r* (LOR, FO) = .41, *r* (LOR, CWO) = .35, *r* (FO, CWO) = .30. To address the question of how well distributional word usage explains children's word learning, we regressed the LOR values on the FO and CWO values. Visual inspection of the relationships between LOR and the two predictor variables show that they are approximately linear. Because the pairwise structure of the values violates independence, ordinary regression gives valid estimates for the regression coefficients but not for the standard errors. Thus, standard errors were computed using bootstrapping: we repeatedly sampled 654 words with replacement to approximate the sampling distribution of the regression coefficients. Regression results are summarized in Table 4.1. For regression of LOR on FO, β (change in expected LOR per standard deviation of FO) = .0217 (R^2 = .166, 95% CI [.146 - .184]). For regression of LOR on CWO, β = .0184 (R^2 = .120, 95% CI [.103 - .137]). Multiple regression of LOR on FO and CWO achieves R^2 of .222 (CI [.200 - .241]). Thus, CWO is a significantly better predictor than chance, FO is significantly better than CWO, and FO and CWO jointly are better than FO alone.

Better model fits can be achieved by accounting for the fact that LOR is negatively correlated (*r* = -.17) with the difference in normative age of acquisition between two words (∆AOA), where normative age of acquisition is the estimated age at which half of children produce the word (Frank et al., 2016). This is expected both for fundamental reasons and for methodological reasons: because words that are learned at widely disparate ages are likely to be learned under very different conditions and conceptual foundations, and because LOR cannot detect relations between word pairs where children start learning the harder word only after the age at which all children know the easier word. Including ∆AOA as a main effect and interaction term increased the R^2 of the FO, CWO, and full models to .196, .154, and .257, respectively, and also increased the coefficients for the main effects of FO and CWO to represent the estimated effect sizes in the ideal case where the word pairs are at the same developmental level.

Next, we investigated whether adults' subjective associations among words improved prediction of LOR over and above distributional measures alone (Table 4.2). Multiple regression of LOR on FO, WO, and direct subjective associations (DA) did not significantly improve fit over the model using FO and WO alone, either without including effects of ∆AOA (R^2 = .224, CI [.202 - .244], p = .45) or with those effects (R^2 = .259, CI [.234 - .282], p = .46). Including subjective associations slightly lowered the estimated effect of CWO but not FO, suggesting that the relevance of adults' subjective associations to children's word learning is mostly through thematic relations between words. Similarly, indirect subjective associations (IA) did not improve the fit over FO and CWO alone, either without including effects of ∆AOA (R^2 = .226, CI [.204 - .245], p = .40), or with those effects (R^2 = .261, CI [.235 - .284], p = .42).

Finally, we investigated whether relationships between distributional measures and LOR differed by word class (Table 4.3). Word classes were defined as the 22 sections on the CDI, which are split into semantic groups by both functional distinctions and, for object nouns, thematic distinctions. For each word class, we re-fit the two models, first with FO, ∆AOA, and their interaction, and then with CWO, ∆AOA, and their interaction, using only word pairs in which at least one of the words was a member of the word class. Each model was evaluated using both R^2 and the coefficient on FO or CWO (Table 4.3). Although a detailed statistical analysis of differences in effect sizes across word classes is outside the scope of the current study, a few patterns are evident. First, β_{FO} and β_{CWO} are greater than 0 for all word classes, indicating that the observed effects are not simply due to the learning patterns of a specific subset of words. β_{FO} tends to be higher for word classes made up of closed-class words than for open-class words, and β_{FO} is lowest for word classes that are not strongly embedded in sentences (sounds, games and routines). Compared to FO, the effect of CWO did not vary as strongly or as systematically across word classes.

Discussion: Study 2

Using distributional semantic representations derived from a corpus of child-directed speech, we were able to explain up to 26% of the variability in the tendency of word pairs to be learned in a correlated way. Words that occurred in overlapping frames tended to align with major syntactic and functional word classes, such as nouns, verbs, and several types of closed-class words, which were also the strongest clusters in LOR. A smaller but still significant effect was observed for words that occurred in the vicinity of the same context words, a distributional measure of thematic semantics. The frame overlap and word overlap each contributed separately to predicting children's learning patterns, as the model using both predicted LOR significantly over and above either measure alone. The accuracy of these predictions was improved by accounting for the fact that words with widely disparate normative ages of acquisition tend to have smaller correlations. However, augmenting the model in this way did not change the overall pattern of results. Thus, the systematic correlations in children's word learning are well explained by the usage patterns of words in natural language samples. Neither syntactic nor thematic relations dominated; rather, children seem able to flexibly use whatever type of salient association is available to organize their word learning. This pattern of complementary effects of the two measures is consistent with previous research indicating that word similarity measures that integrate multiple sources of information (features and word co-occurrences) fit behavioral data better than models using only one source (Andrews, Vigliocco, & Vinson, 2009). The overall patterns were not driven by one or a few word classes, but held generally across word categories as defined on the MCDI, although frames were more predictive of correlated learning for word pairs that included closed-class words.

We also addressed the question of whether children's correlated word learning is sensitive primarily to the word associations present in their linguistic environment, or to lexical associations that adults find highly available. To do this, we used the free-association norms of Nelson, McEvoy, and Schreiber (2004) to compute both the adjacency matrix and the path distance between words. However, these subjective associations did not significantly improve the LOR predictions. The free-association norms contain a rich amount of information beyond that encoded the distributional measures, such as taxonomic associations, object properties, linguistic collocations, and cultural associations. Nevertheless, word associations encoded in adults' semantic associations did not predict children's word learning correlation above the word associations implicit in children's actual language exposure. It is possible that free-association norms were ineffective because they selected mostly strong associations involving later-learned semantic (e.g., idiomatic or literary) usages, whereas children's early learning might depend primarily on broader associations such as membership in the same word class or co-occurrence in an activity context. Such loosely connected word pairs might be expected to occur in similar linguistic contexts and be amenable to learning by similar strategies, despite not appearing in free associations.

Another relevant question is whether lexical distributional measures directly index important patterns of input to children's learning, or whether they are proxies for other environmental associations or semantic factors. A few studies have investigated statistical associations between words or objects and the spatial, temporal, and perceptual environments in which they occur (Roy, Frank, DeCamp, Miller, & Roy, 2015; Sadeghi et al., 2015; Clerkin, Hart, Rehg, Yu, & Smith, 2017). These studies find that nonlexical environmental distributions are correlated with lexical distributions, and also independently predict learning. Thus, though it is difficult to assign causality to the lexical distributional measures, the current results add to a body of results suggesting that associative structure in children's everyday environmental inputs influences learning (Goldstein et al., 2010). Further characterization of environmental inputs other than speech, from naturalistic datasets involving large samples of children and multiple, representative social-interactive contexts, would be likely to explain additional systematicity in children's learning.

General Discussion

Firth (1957:11) famously said, "you shall know a word by the company it keeps," referring to the power of co-occurrence between words in language corpora to deliver insight about their meaning. Another way words "keep each other company" is by co-occurring in the vocabulary of an individual. In this chapter, we use patterns in the tendencies of words to keep each other's company as a source of insights about conceptual and lexical development in children. We showed that the network of words, linked by their correlated learning patterns, not only reflects common-sense knowledge about the structure of children's semantic universe, but also makes it possible to explore that structure in vivid detail. In addition to mapping this word learning network, we also develop two of its applications: first, we use subnetworks defined by the words known by individual children to show that, at a given absolute vocabulary size, children with large vocabularies for their age (early talkers) know sets of words that are more closely related, connected by shorter paths, and contain more hubs compared to children with small vocabularies for their age (late talkers). Second, we use the network as a target and investigate to what extent it can be explained ("predicted") by the distribution of words over linguistic contexts in large natural samples of parents' speech to children.

In Study 1, we identified systematic differences in the network structure of children's vocabularies as a function of children's age-adjusted vocabulary size. We did not find a specific effect for the slowest-learning group ("late talkers"), but rather approximately linear effects spanning the whole range of vocabulary sizes. What might differ between the learning mechanisms of faster-learning and slower-learning children to account for these patterns? Faster-learning children might learn more connected words because they are better able to harness semantic and syntactic connections to learn new words. This might occur because they have greater syntactic knowledge or a more well-connected and richer conceptual system.

Some children might also learn sets of related words if they discover generalizable word learning strategies or “biases,” such as an object-shape/object-name bias (Landau, Smith & Jones, 1988; Colunga & Sims, 2017), a limited assumption that novel words map onto non-synonymous concepts (Markman & Wachtel, 1988; Yurovsky, Bion, Smith, & Fernald, 2012; but see Deák & Maratsos, 1998; Deák, Yen & Petit, 2001) or syntactic bootstrapping (Gleitman, 1990), which could promote learning many words from the same class.

Alternatively or in addition, faster-learning children might more reliably learn a specific set of “fundamental” words that are more densely connected among each other than to words outside that set. A related but distinct possibility is that causation goes in the other direction: learning a set of well-connected words might accelerate further vocabulary growth. Another complicating factor is that faster-learning children reach a given vocabulary size at a younger age, so age-related differences in children’s interests, environments, and non-vocabulary knowledge might additionally influence their vocabulary growth patterns (Storkel, 2009). Ultimately, explaining differences between fast- and slow-learning children might depend on a better understanding of how individual differences in language development emerge from variability in genetic factors (Tomblin & Buckwalter, 1998; Dale, Dionne, Eley, & Plomin, 2000; Stromswold, 2017) and environmental factors such as socioeconomic status, caregiver responsiveness, quantity of speech input (Hart & Risley, 1995; Tamis-LeMonda, Bornstein, & Baumwell, 2001; Hoff, 2003; Tamis-LeMonda, Kuchirko, & Song, 2014).

Study 2 demonstrated that word learning correlations are consistent with the tendency of word pairs to occur in shared contexts in natural child-directed speech – both the immediate context defined by the two adjacent words, and the broader context defined by words that appeared in close proximity but not adjacent. These large-scale measures of children’s situated exposure to words explained a substantial proportion of the variance in correlated word learning, whereas adults’ word association norms did not. Despite the caveat that the different measures might have varied in sensitivity and validity, this suggests that rich structure in children's environment, rather than word semantics measured independently of children's experience, might ultimately be the best way to account for children's learning.

A large proportion of variance still remains to be explained. Two factors limit our accuracy: first, our language samples come from a different sample of children than the children whose vocabularies we studied. Second, our analysis is limited to orthographic transcriptions of parental speech, so we cannot observe effects involving the embedding of words in children's spatiotemporal or sensory environments. Studies using video recording suggest that individualized, multimodal inputs do have a strong influence on word learning. For instance, toddlers are more likely to learn words they hear during moments when the referent object is visually dominant (Yu & Smith, 2012) or during joint attention with parents (Tomasello & Farrar, 1986). An open question concerns to what extent this embodied referent selection depends on explicitly social communication (Csibra & Gergely, 2009) as opposed to relying on general processes of attention and associative learning (Szufnarowska, Rohlfing, Fawcett, & Gredebäck, 2014; Yu, Suanda, & Smith, 2019).

Whereas the aforementioned studies used short recording sessions and evaluated learning of a few target words, one study used large-scale recording of the spatiotemporal distribution of words in one toddler's environment and revealed that for that child, words with distinct contextual distributions were learned earlier (Roy et al., 2015). Technical and practical limitations have prevented collecting this type of dataset for large samples of children. In recent years, however, emerging technologies as well as a greater appreciation for the role of environmental structure have produced several efforts that might provide data sources richer than the speech transcripts used in the current study. For example, the LENA system provides hardware for all-day audio recording for infants and software for speaker recognition, allowing detailed assessments of the amount and timing of speech (Xu et al., 2008; Greenwood, Schnitz, Irvin, Tsai, & Carta, 2018; Bergelson et al., 2019). Head-mounted cameras can be fitted to infants and toddlers, and platforms such as Amazon Mechanical Turk enable crowdsourcing of image analysis, e.g. object tagging (Clerkin et al., 2017; Fausey, Jarayaman, & Smith, 2016). Head-mounted eye tracking systems make it possible to record infants' gaze targets during naturalistic behavior (Franchak, Kretch, Soska, & Adolph, 2011; Yu & Smith, 2013; Franchak, Kretch, & Adolph, 2018), and advances in computer vision might enable more automated video coding (e.g. Bambach, Lee, Crandall, & Yu, 2015; Mirsharif, Sadani, Shah, Yoshida, & Burling, 2017). Thus, in the near future it might be possible to create datasets that enable large-scale analyses of not only the linguistic distribution of words, but also their distributions over time, space, visual stimulation, and parent social cues.

These studies only scratch the surface of the potential applications for mapping the tracks along which children grow their semantic networks - or the trails and ruts where those tracks will be more readily laid. We computed the pairwise relations among words for only a single sample of convenience of American English-learning children aged 16 to 32 months. One issue is that this sample is not representative of all populations of word learners. Without studying diverse populations, we cannot explore the similarities and differences between semantic networks trained on populations that vary by language, culture, ethnicity, socioeconomic status, gender, technology exposure, or neurodevelopmental status. Studying varied populations might reveal alternative ways of organizing words in a network as a function of different experience. Word relations might vary in part because of structural differences between languages. For example, Korean sentences typically include verbs in salient positions and often omit noun information. As a result, Korean-learning toddlers are better than English-learning toddlers at learning verbs without supporting noun information (Arunachalam, Leddon, Song, Lee, & Waxman, 2013). This might cause Korean verbs to show less correlated learning with semantically related nouns than in English. Another population that might provide an interesting comparison is children with ASD. Although children with ASD have delayed vocabulary growth, their vocabulary composition has not been found to significantly deviate from typically developing children (Rescorla & Safyer, 2013). Yet more detailed analyses might reveal subtler effects of language-specific encoding of certain word classes (e.g., spatial predicates) on acquisition. Free-association norms in adults with high-functioning ASD reveals that their semantic networks are more strongly “modular,” or organized around rigid categories (Kenett, Gold, & Faust, 2016). Thus, one might expect the patterns of correlated word learning to be more affected in ASD than the probabilities of learning individual words or the distribution over word classes. Finally, by studying older children in these and other populations, we might learn how exposure to the more structured pedagogical environment of school contributes to the developing organization of children’s vocabulary (Dubossarsky, De Deyne, & Hills, 2017).

Another potential direction to explore is the practical applications of models predicting word learning via network and corpus properties. It might be possible to accelerate word learning by ensuring children are exposed to words that are likely to be learned given the words that they already know. Another possible strategy is to enrich words that have many connections to unknown words, to maximize the number of target words that can connect to the growing network. Finally, it might be useful to enrich words that reduce the path distances within the network, under the hypothesis that the short path distances characteristic of small-world networks are beneficial.

This investigation introduced a method to compute compact pairwise representations of the correlations in learning between words, without relying on longitudinal data. The same method could be applied to study the learning patterns of any set of discrete items. For example, one could use networks to study correlated mastery of classroom learning objectives. Similarly to vocabulary, it might be possible to identify differences in structure between the networks of items learned by children with different levels of overall performance, or to identify features of learning items that explain their correlated learning. Another possible application is to batteries of cognitive tests (e.g. Bayley, 2006). Network analysis of correlated performance on test items might reveal structures missed by existing divisions into subscores or techniques such as factor analysis.

Overall, the current study represents an effort to use “big data” not to predict developmental outcomes of direct practical interest, but rather to mine the data for secondary patterns that reveal the structure of cognitive development. With the rapid growth in collection of large and detailed observational datasets, there will be many more opportunities to discover subtle traces of cognitive and developmental mechanism within the complexity of human behavior.

Acknowledgements

Chapter 4, in part, is currently being prepared for submission for publication of the material, and is co-authored with Gedeon Deák. The dissertation author was the primary investigator and author of this paper.

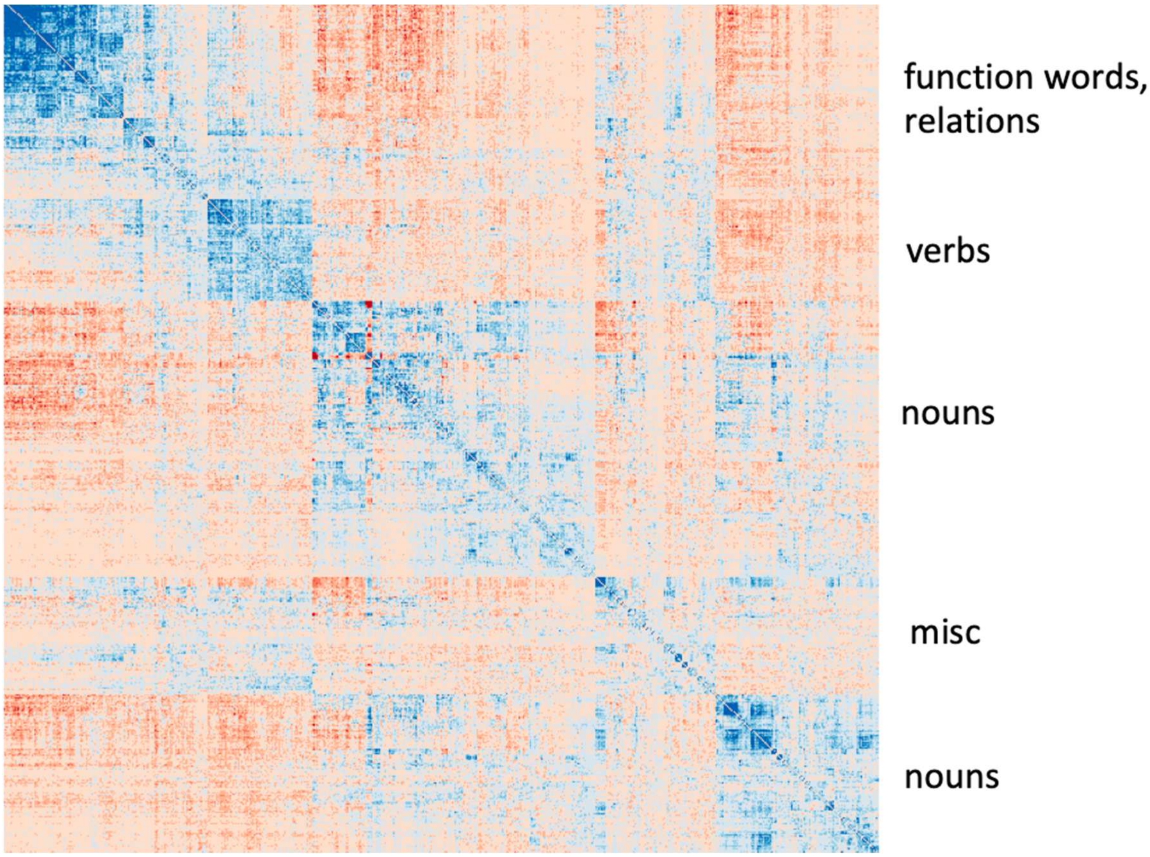

Figure 4.1: Log-odds ratio (LOR) matrix organized using hierarchical clustering. Each row and each column of this 654x654 matrix represents a single word, and each entry represents the LOR of a word pair (entries below the diagonal are redundant). Blue values indicate positive LOR, and red values indicate negative LOR. Hierarchical clustering was used to organize related words into groups visible as blue squares along the diagonal. Labels on the right describe typical words in each cluster.

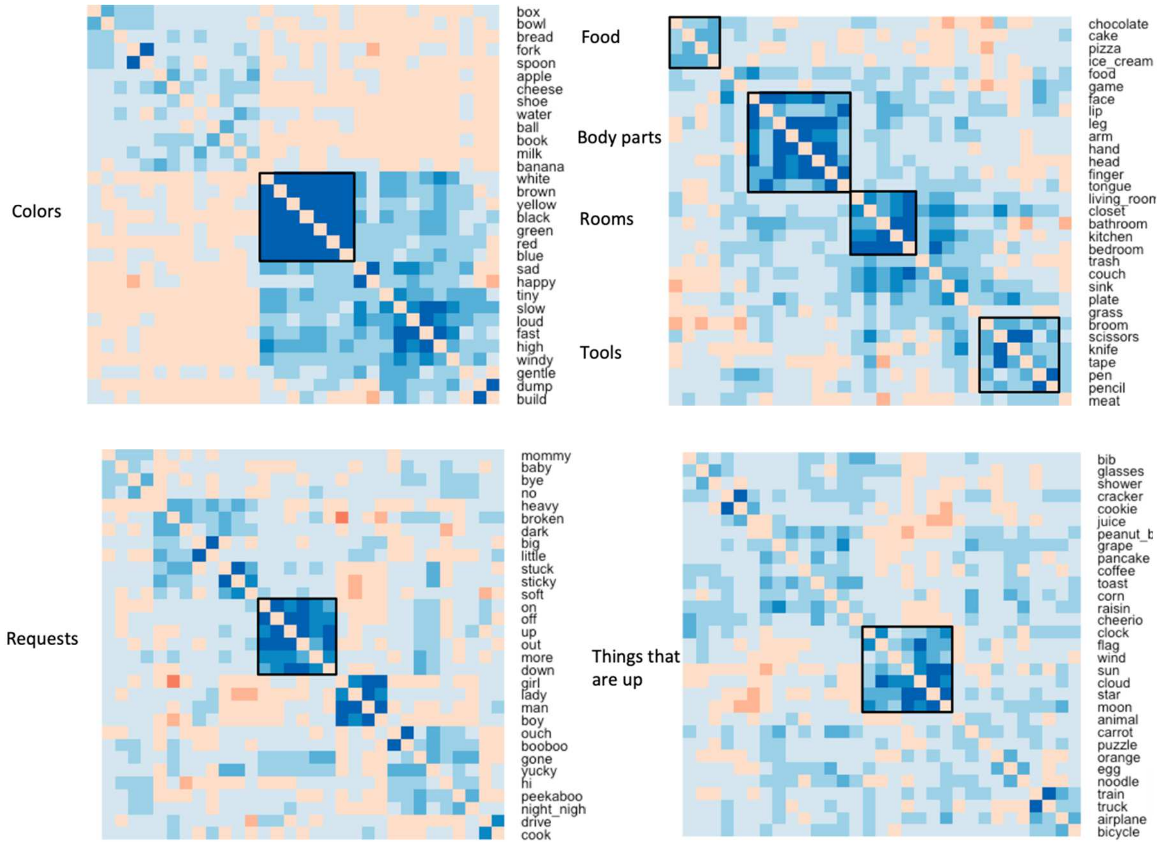

Figure 4.2: Selected small-scale clusters of related words.

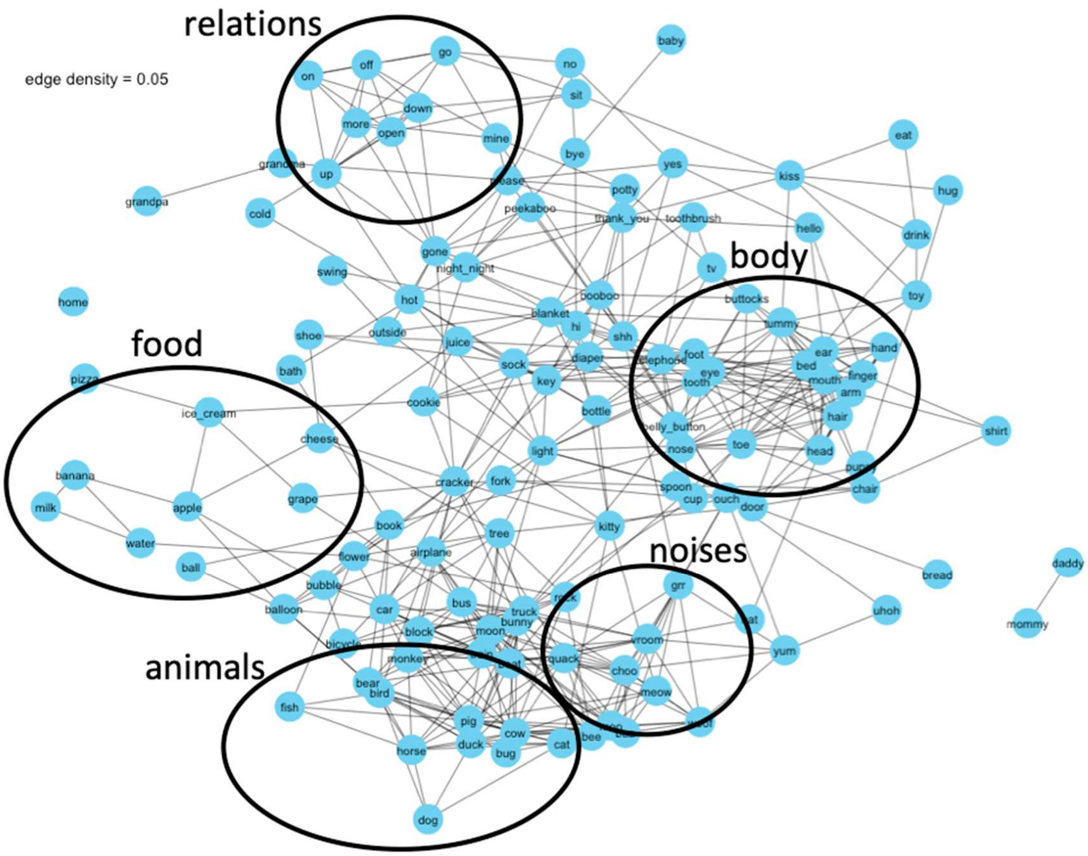

Figure 4.3: Log-odds-ratio subnetwork consisting of the 125 earliest-learned words, with connections shown when LOR was greater than the 95th percentile of word pairs. Nodes were placed using the Fruchterman-Reingold force-directed algorithm (Fruchterman & Reingold, 1991) and resulting clusters of qualitatively related nodes are indicated.

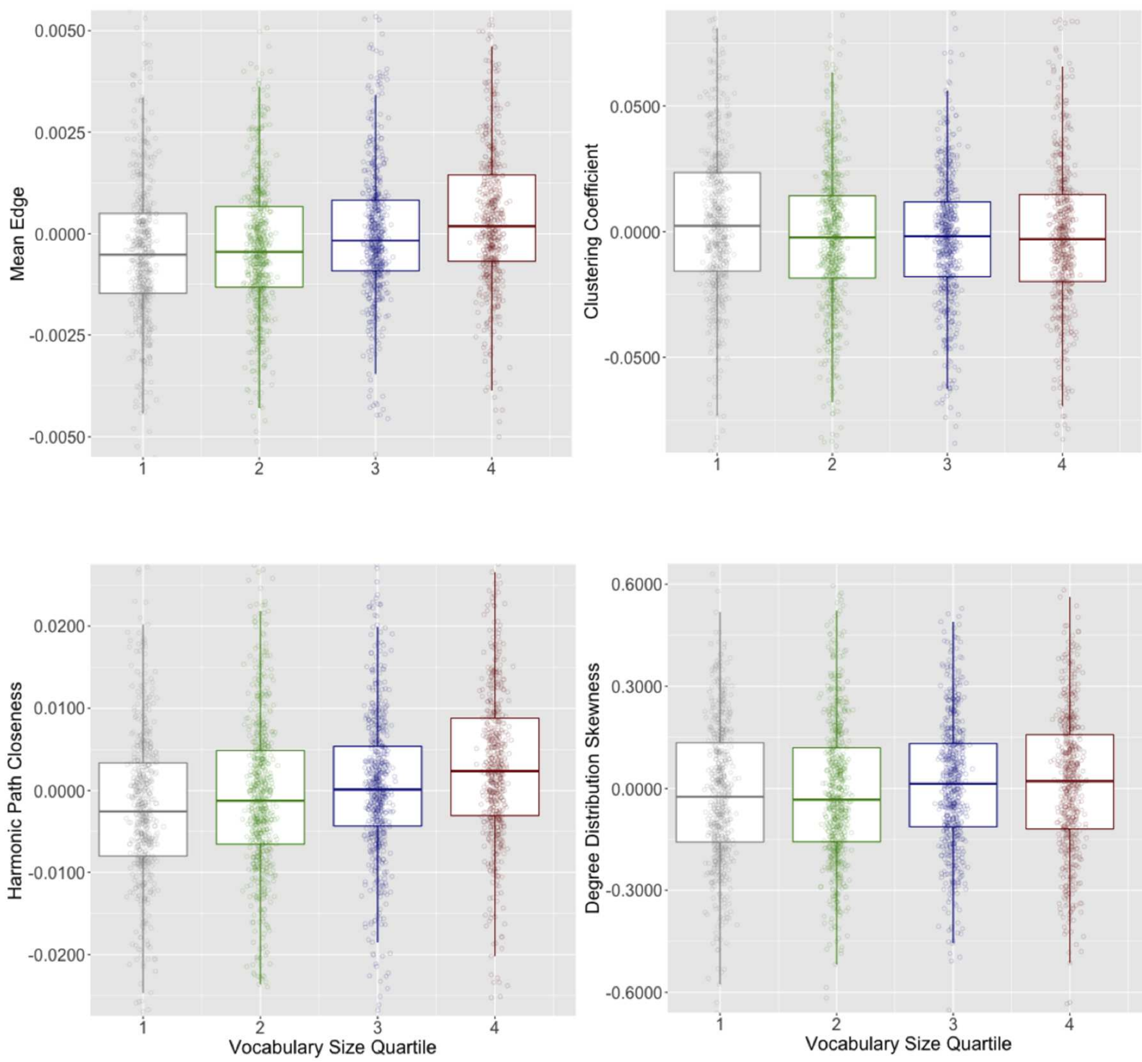

Figure 4.4: Differences in network structure metrics as a function of age-adjusted vocabulary size quartile group. Y-axis shows metrics corrected for absolute vocabulary size as described in the text, so that 0 represents the mean for a given vocabulary size. Box plots indicate the 5th, 25th, 50th, 75th, and 95th percentile, and small circles indicate the values (with random horizontal jitter) for individual children (maximum and minimum values do not necessarily fall within the plot area).

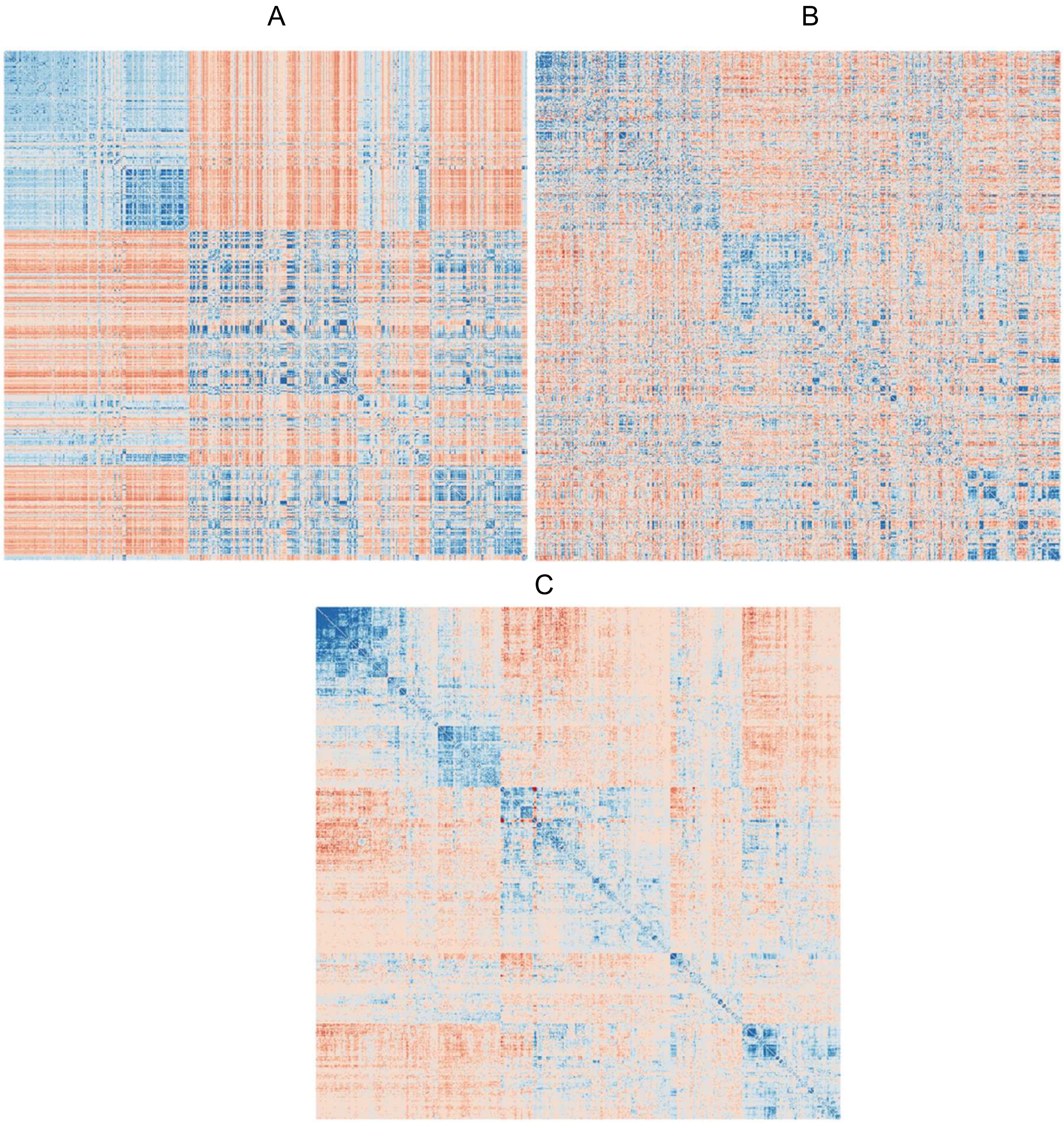

Figure 4.5: Word relation matrices organized using the same hierarchical clusters based on LOR. Blue indicates positive values, and red values indicate negative values. A: FO; B: CWO; C: LOR (reproduction of Figure 4.1, for comparison)

Table 4.1: Regression models using frame and context-word overlap to predict log-odds ratio.

	β_{FO}	β_{CWO}	R^2	CI
FO	.0217	-	.166	.146 - .186
CWO	-	.0184	.120	.102 - .136
FO, CWO	.0178	.0131	.222	.200 - .242
FO x ∆AOA	.0310	-	.196	.174 - .218
CWO x ∆AOA	-	.0286	.154	.132 - .173
FO x ∆AOA, CWO x ∆AOA	.0243	.0201	.257	.233 - .281

Table 4.2: Subjective associations do not significantly improve prediction of log-odds ratio over and above the regression on frame and context-word overlap.

	β_{FO}	β_{CWO}	β_{DA}	β_{IA}	R^2	CI
FO, CWO	.0178	.0131	-	-	.222	.200 - .242
FO, CWO, DA	.0178	.0127	.0184	-	.224	.202 - .244
FO, CWO, IA	.0177	.0122	-	.0034	.226	.204 - .245
FO, CWO x ∆AOA	.0243	.0201	-	-	.257	.233 - .281
FO, CWO, DA x ∆AOA	.0243	.0193	.0033	-	.259	.234 - .282
FO, CWO, IA x ∆AOA	.0241	.0186	-	.0047	.261	.235 - .284

Table 4.3: Model fits split by CDI category (ordered by average R^2 across the two models).

		FO x ΔAOA		WO x ΔAOA	
Word category	Number of words (pairs)	β_{FO}	R^2	β_{WO}	R^2
Connecting words	6 (3903)	0.106	0.482	0.069	0.378
Helping verbs	20 (12870)	0.073	0.358	0.047	0.223
Pronouns	25 (16025)	0.08	0.379	0.039	0.194
Question words	7 (4550)	0.073	0.417	0.033	0.144
Time words	12 (7770)	0.055	0.288	0.032	0.246
Locations	24 (15396)	0.063	0.331	0.031	0.136
Quantifiers	17 (10965)	0.08	0.321	0.033	0.132
Animals	43 (27176)	0.033	0.202	0.032	0.186
Outside words	31 (19778)	0.023	0.196	0.032	0.167
Vehicles	14 (9051)	0.023	0.211	0.03	0.136
Body parts	25 (16025)	0.024	0.181	0.022	0.143
Action words	103 (62006)	0.03	0.211	0.019	0.088
Furniture and rooms	33 (21021)	0.017	0.157	0.028	0.138
Descriptive words	63 (39196)	0.032	0.174	0.023	0.109
Sounds	12 (7770)	0.011	0.063	0.022	0.212
Household objects	50 (31425)	0.018	0.131	0.021	0.098
Food and drink	68 (42126)	0.027	0.123	0.018	0.1
Clothing	28 (17906)	0.021	0.089	0.025	0.113
Toys	18 (11601)	0.019	0.13	0.018	0.063
Places	22 (14135)	0.012	0.053	0.025	0.138
People	26 (16653)	0.02	0.057	0.021	0.09
Games and routines	19 (12236)	0.011	0.046	0.01	0.045

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, *22*(3), 261-295.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological review*, *116*(3), 463.

Arunachalam, S., Leddon, E. M., Song, H. J., Lee, Y., & Waxman, S. R. (2013). Doing more with less: Verb learning in Korean-acquiring 24-month-olds. *Language acquisition*, *20*(4), 292-304.

Bambach, S., Lee, S., Crandall, D. J., & Yu, C. (2015). Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions. In *Proceedings of the IEEE International Conference on Computer Vision* (pp.1949-1957)

Barabási, A. L. (2009). Scale-free networks: a decade and beyond. *Science*, *325*(5939), 412-413.

Barabási, A. L. (2016). *Network science*. Cambridge University Press.

Bayley, N. (2006). *Bayley scales of infant and toddler development*. PsychCorp, Pearson.

Beckage, N., & Colunga, E. (2013). Using the words toddlers know now to predict the words they will learn next. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.

Beckage, N. M., & Colunga, E. (2019). Network Growth Modeling to Capture Individual Lexical Learning. *Complexity*, 2019.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS one*, *6*(5).

Beckage, N., Steyvers, M., & Butts, C. (2012). Route choice in individuals—semantic network navigation. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*.

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental science, 22*(1), e12724.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child development*, *87*(6), 1893-1908.

Borovsky, A., & Elman, J. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of child language*, *33*(4), 759-790.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. (2016, August). From uh-oh to tomorrow: Predicting age of acquisition for early words across languages. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using 'frequent frames': what cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental science*, *12*(3), 396-406.

Clair, M. C. S., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*, *116*(3), 341-360.

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.

Colunga, E., & Sims, C. E. (2017). Not only size matters: Early-talker and late-talker vocabularies support different word-learning biases in babies and networks. *Cognitive science*, *41*, 73-95

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148-153.

Dale, P. S., Dionne, G., Eley, T. C., & Plomin, R. (2000). Lexical and grammatical development: A behavioural genetic perspective. *Journal of child language*, *27*(3), 619-642.

Deák, G. O., & Maratsos, M. (1998). On having complex representations of things: Preschoolers use multiple words for objects and people. *Developmental Psychology*, *34*(2), 224.

Deák, G. O., Yen, L., & Pettit, J. (2001). By any other name: when will preschoolers produce several labels for a referent? *Journal of Child Language*, *28*(3), 787-804.

De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213-231.

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*, *33*(48), 18906-18916.

Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Developmental psychology, 53*(8), 1560.

Engelthaler, T., & Hills, T. T. (2017). Feature biases in early word learning: Network distinctiveness predicts age of acquisition. *Cognitive science*, *41*, 120-140.

Eriksson, M., Westerlund, M., & Berglund, E. (2002). A screening version of the Swedish communicative development inventories designed for use with 18-month-old children. *Journal of Speech, Language, and Hearing Research, 45*, 948–960.

Erten, İ. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, *36*(3), 407-422.

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.

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child development*, *71*(2), 310-322.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics*, *21*(1), 95-116.

Firth, John R. 1957. A synopsis of linguistic theory 1930–1955. In *Studies in linguistic analysis*, 1-32. Oxford: Blackwell.

Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2935-2948.

Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant–caregiver social looking during locomotor free play. *Developmental science*, *21*(4), e12626.

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*(6), 1738-1750.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. Journal of Child Language. doi: 10.1017/S0305000916000209.

Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and experience*, *21*(11), 1129-1164.

Gleitman, L. (1990). The structural sources of verb meanings. *Language acquisition*, *1*(1), 3-55.

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in cognitive sciences*, *14*(6), 249-258.

Greenwood, C. R., Schnitz, A. G., Irvin, D., Tsai, S. F., & Carta, J. J. (2018). Automated language environment analysis: A research synthesis. *American Journal of Speech-Language Pathology*, *27*(2), 853-867

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

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition?. *Psychological science*, *20*(6), 729-739.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child development*, *74*(5), 1368-1378.

Huebner, P. A., & Willits, J. A. (2018). Structured semantic knowledge can emerge automatically from predicting word sequences in child-directed speech. *Frontiers in Psychology*, *9*, 133.

Jimenez, E., & Hills, T. (2017). Network Analysis of a Large Sample of Typical and Late Talkers. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Ke, J., & Yao, Y. (2008). Analysing language development from a network approach. *Journal of Quantitative Linguistics*, *15*(1), 70-99.

Kenett, Y. N., Gold, R., & Faust, M. (2016). The hyper-modular associative mind: a computational analysis of associative responses of persons with Asperger syndrome. *Language and Speech*, *59*(3), 297-317.

Kenett, Y. N., Kenett, D. Y., Ben-Jacob, E., & Faust, M. (2011). Global and local features of semantic networks: Evidence from the Hebrew mental lexicon. *PloS one*, *6*(8), e23912.

Kenett, Y. N., Wechsler-Kashi, D., Kenett, D. Y., Schwartz, R. G., Ben Jacob, E., & Faust, M. (2013). Semantic organization in children with cochlear implants: computational analysis of verbal fluency. *Frontiers in Psychology*, *4*, 543.

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory & Cognition*, *9*(6), 587-598.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive development*, *3*(3), 299-321.

Lerner, A. J., Ogrocki, P. K., & Thomas, P. J. (2009). Network graph analysis of category fluency testing. *Cognitive and Behavioral Neurology*, *22*(1), 45-52.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research methods, instruments, & computers*, *28*(2), 203-208.

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, *20*(2), 121-157.

Martin, N., Roach, A., Brecher, A., & Lowery, J. (1998). Lexical retrieval mechanisms underlying whole-word perseveration errors in anomic aphasia. *Aphasiology, 12*(4-5), 319-333.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, *37*(4), 547-559.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Milgram, S. (1967). The small world problem. *Psychology today*, *2*(1), 60-67.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91-117.

Mirsharif, Q., Sadani, S., Shah, S., Yoshida, H., & Burling, J. (2017). A Semi-Automated Method for Object Segmentation in Infant's Egocentric Videos to Study Object Perception. In *Proceedings of International Conference on Computer Vision and Image Processing* (pp. 59-69). Springer, Singapore.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143-182.

Moyle, M. J., Weismer, S. E., Evans, J. L., & Lindstrom, M. J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research*.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3), 402-407.

Oliver, B., Dale, P.S., Saudino, K.J., Petrill, S.A., Pike, A., & Plomin, R. (2003). The validity of a parent-based assessment of cognitive abilities in three-year olds. *Early Child Development and Care, 172*, 337–348.

Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *Journal of Neuroscience*, *32*(45), 15728-15736.

Rescorla, L., & Safyer, P. (2013). Lexical composition in children with autism spectrum disorder (ASD). *Journal of Child Language*, *40*(1), 47-68.

Rochat, Y. (2009). *Closeness centrality extended to unconnected graphs: The harmonic centrality index*. ASNA.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Rohde, D. L., Gonnerman, L. M., & Plaut, D. C. (2004). An improved method for deriving word meaning from lexical co-occurrence. *Cognitive Psychology*, *7*, 573-605.

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*, *112*(41), 12663-12668.

Sadeghi, Z., McClelland, J. L., & Hoffman, P. (2015). You shall know an object by the company it keeps: An investigation of semantic representations derived from object co-occurrence in visual scenes. *Neuropsychologia*, *76*, 52-61.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2002). Semantic spaces based on free association that predict memory performance. *Triple Festschrift honoring Lyle Bourne, Walter Kintsch, and Tom Landauer*.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41-78.

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of child language*, *36*(2), 291-321.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, *410*(6825), 268.

Stromswold, K. (2017). Genetic specificity of linguistic heritability. In *Twenty-First Century Psycholinguistics* (pp. 141-156). Routledge.

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? *Scientific Reports*, *4*, 5304.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child development*, *72*(3), 748-767.

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

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, *21*(3), 371-380.

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

Tomasello, M., & Mervis, C.B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, *59*, 174–179.

Tomblin, J. B., & Buckwalter, P. R. (1998). Heritability of poor language achievement among twins. *Journal of Speech, Language, and Hearing Research*, *41*(1), 188-199.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, *393*(6684), 440.

Xu, D., Yapanel, U., Gray, S., Gilkerson, J., Richards, J., & Hansen, J. (2008). Signal processing for young child speech language development. In *First Workshop on Child, Computer and Interaction*.

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language learning and Development*, *4*(1), 32-62.

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

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS one*, *8*(11), e79659.

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental science*, *22*(1), e12735.

Yurovsky, D., Bion, R. A., Smith, L. B., & Fernald, A. (2012). Mutual exclusivity and vocabulary structure. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society.*

CONCLUSION

In this work, I took two angles to the study of how interacting contextual factors shape infants' and toddlers' word learning: a micro-level approach and a macro-level approach. Within the micro-level approach, I studied recordings of naturalistic infant-parent interactions using hand-annotated transcriptions of infants' and parents' vocal, gaze, and manual behaviors. Building on previous research, I found that parents' utterances were embedded in multiple layers of regularities determined by infants' object-oriented and person-oriented activity, the physical situation, parents' responsiveness, and the structure of typical language/action routines. In Chapter 1 I found that, between the ages of 4 and 9 months, parents tended to name objects in response to onsets and offsets of infants' manual object exploration of multiple objects. Developmental changes and individual difference in the amount and maturity of infants' manual activity thus determined the availability of word-learning opportunities. In Chapter 2, I found that parents tended to produce descriptive and object-naming utterances as part of multi-utterance sequences that also depended on successfully recruiting and sustaining infant's attention. Infants' exposure to parents' descriptive and naming utterances about objects was therefore not a static affordance of the environment. Rather, it was co-constructed by both partners in the interaction, modulated by individual differences in the infants' developmental status and the parents' interaction style. Overall, these results illustrate how the inputs to word learning go far beyond co-occurrences between word forms and potential referents, by framing object-labeling events in a spatial, temporal, linguistic, and developmental context.

At the macro level, I studied how the usage patterns of English words in a large corpus of child-directed speech relate to regularities in children's production of those words. I introduced two distributional measures of word usage: frame-based features, and non-adjacent co-occurrence-based features. I found that the co-occurrences between words in child-directed speech, when filtered and processed at these two levels, produced rich, interpretable representations of words that effectively segregated two types of information: the frame-based features primarily reflect information about grammatical class and usage in linguistic constructions, whereas the non-adjacent co-occurrence-based features primarily reflect broader semantic associations between words and the objects and activities they co-occur with. In Chapter 3, I showed that these distributional representations help predict normative age of acquisition over and above semantic and frequency variables that have previously been used to explain age of acquisition. Moving on from population-average ages of acquisition, in Chapter 4 I investigated the tendency of words to be learned in a correlated manner—that is, whether individual children learned both words in a usage-correlated pair more often than would be expected if words were learned independently. I found that the similarity between two words' distributional representations is a strong predictor of their tendency to be learned together. I also used the observed tendency for each word pair as a novel way to define the edges in a vocabulary network. Using this network representation I found modest but significant differences between the network structures of children with above- and below-average productive vocabulary: higher-vocabulary children produced words that were, on average, more closely related, connected by shorter-distance paths, and organized with more significant hubs. These results demonstrate that children learn words in predictable orders that follow naturally from similarities and differences in the words' contextual distributions.

The individual studies in this dissertation contain several novel findings that might be primarily of interest to researchers studying several different areas within the field of first language acquisition, including word learning but also early syntactic development, categorization of objects and other concepts, and statistical learning in general. Taken together, they support theories of word learning that ascribe a major role to the multimodal, embodied contexts in which words are experienced (e.g., Suanda, Smith, & Yu, 2016). Although word exposures are at first embedded in contexts provided by parents' scaffolding (Gogate, Maganti, & Bahrick, 2015; Gogate, Bahrick & Watson, 2000; Meyer, Hard, Brand, McGarvey, & Baldwin, 2011), infants and toddlers gradually develop behavioral and cognitive competences that allow them both to actively participate in generating richly embedded linguistic inputs (Smith, Jayaraman, Clerkin, & Yu, 2018) and to efficiently make valid inferences about word meaning (Hockema & Smith, 2009). Rather than simply referring to specific entities, words might be best understood as parts of enacted communicative events that unfold according to conventional “protocols” such as *interactional formats* (Bruner, 1983) or *pragmatic frames* (Rohlfing, Wrede, Vollmer, & Oudeyer, 2016). In sum, we see that parents’ object-referring utterances to infants are embedded in multimodal, responsive sequences, and that patterns in the words children learn suggest that infants and toddlers learn groups of words organized around habitual ways of enacting their meaning in a consistent context.

Several limitations of the work presented in this dissertation also illustrate pervasive and ongoing challenges for research on word learning. Notably, our ability to understand how children learn words is constrained by the difficulty of determining causality, and by the diversity of sociocultural contexts in which children are raised, as described below. However, I will explain how these challenges also present opportunities for future research directions and discoveries.

None of the studies in this work were expected to produce causal evidence for the claim that any specific environmental feature increases the probability that a child will learn a word. It is extremely difficult (or unethical) to manipulate a child’s linguistic environment in ways that maintain experimental control without oversimplifying the developmental task. In particular, stimulus characteristics cannot be controlled without severely constraining the freedom of the child-adult dyad to engage in mutually responsive interaction. Therefore, experiments and randomized controlled trials of word learning have been limited to small-scale laboratory experiments, usually involving a few exposures to a few words, and to enrichment interventions, which apply experimental control at too high a level to answer questions about representative mechanisms. Instead of the experimental method, I have followed many other developmental scientists in adopting an observational, descriptive, and/or correlational approach: documenting behavioral patterns that are outside our control but are of theoretical relevance for word learning, and measuring the extent to which they correlate with word learning. Using new statistical techniques being developed for data science, econometrics, and other scientific fields, we tap into a growing repertoire of tools for finding patterns and quantifying trends in complex, multivariate datasets. These techniques can even generate quasi-causal inferences under certain conditions (Imbens & Lemieux, 2018, Pearl, 2009). On the data collection side, innovations include, for example large-scale data annotation via crowdsourcing (Clerkin, Hart, Rehg, Yu, & Smith, 2017), egocentric video and gaze tracking (Smith, Yu, Yoshida, & Fausey, 2015), and mobile brain-body imaging (Liao, Acar, Makeig, & Deák, 2015). Such methods could be used to replicate and extend the new findings reported in Chapters 3 and 4. At the same time, researchers are increasingly compiling or creating large, open datasets such as the CHILDES and Wordbank databases, with the explicit intent of enabling more “reusable research” (Gilmore, 2016).

Our approach to studying word learning depends on the concrete behavioral instantiations of developmental processes, rather than on abstract universal principles. The current studies analyzed data derived from the specific sequences of visual, manual, and verbal behaviors that occurred in our study populations, i.e. American English-learning children and their parents, within a limited range of times and social classes. Thus, the relevance of our results depends on what one can learn by studying this population. Child-directed speech and play practices vary significantly between populations. In some populations, caregivers rarely address prelinguistic infants directly (Lieven, 1994; Cristia, Dupoux, Gurven, & Stieglitz, 2019), and the physical and social contexts in which child-directed speech practices occur vary significantly across populations (Bornstein et al., 1992; Ogura, Dale, Yamashita, Murase, & Mahieu, 2006; Kärtner et al., 2008). Nevertheless, the structured information generated by these practices correlates with language acquisition outcomes even in populations with minimal child-directed talk (Shneidman & Goldin-Meadow, 2012; Weber, Fernald, & Diop, 2017). Future research might therefore document similarities and differences in the relations between the detailed structure of language input–both child-directed and overheard–and word learning trajectories in children both within and outside commonly-studied WEIRD (Western, Educated, Industrial, Rich, Democratic) populations (Henrich, Heine, & Norenzayan, 2010; Nielsen, Haun, Kärtner, & Legare, 2017). In making such comparisons, it will be important to view each study population carefully rather than simply applying methods that make implicit Western-centric assumptions (Kline, Shamshudheen, & Broesch, 2018). Truly general principles might then follow from an effort to develop models that can account for both the similarities and differences in the word learning patterns observed across different populations.

Cognitive development in general, and word learning in particular, is the result of complex systems that unfold over time and space across multiple scales, between multiple individuals and with abundant crosstalk between different cognitive domains and sensorimotor modalities. Although this complexity can make it difficult to characterize these systems in the abstractions of scientific theory, it also generates learning opportunities by embedding words in predictable interactive sequences and linguistic and non-linguistic contexts. In this dissertation, I have attempted to show how complexity can also be an asset for researchers, as detailed recordings of the events occurring within a complex developmental system and the resulting learning outcomes can provide clues to its function.

References

Bornstein, M. H., Tal, J., Rahn, C., Galperin, C. Z., Pecheux, M. G., Lamour, M., ... & Tamis-LeMonda, C. S. (1992). Functional analysis of the contents of maternal speech to infants of 5 and 13 months in four cultures: Argentina, France, Japan, and the United States. *Developmental Psychology*, *28*(4), 593.

Bruner, J. (1983). Child’s Talk: Learning to Use Language. New York, NY: Norton.

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.

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-Directed speech is infrequent in a forager-farmer population: a time allocation study. *Child development*, *90*(3), 759-773.

Gilmore, R. O. (2016). From big data to deep insight in developmental science. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(2), 112-126.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child development*, *71*(4), 878-894.

Gogate, L., Maganti, M., & Bahrick, L. E. (2015). Cross-cultural evidence for multimodal motherese: Asian Indian mothers' adaptive use of synchronous words and gestures. *Journal of experimental child psychology*, *129*, 110-126.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, *33*(2-3), 61-83.

Hockema, S. A., & Smith, L. B. (2009). Learning your language, outside-in and inside-out. *Linguistics*, *47*(2), 453-479.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, *142*(2), 615-635.

Kärtner, J., Keller, H., Lamm, B., Abels, M., Yovsi, R. D., Chaudhary, N., & Su, Y. (2008). Similarities and differences in contingency experiences of 3-month-olds across sociocultural contexts. *Infant Behavior and Development*, *31*(3), 488-500.

Kline, M. A., Shamsudheen, R., & Broesch, T. (2018). Variation is the universal: making cultural evolution work in developmental psychology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1743), 20170059.

Liao, Y., Acar, Z. A., Makeig, S., & Deák, G. (2015). EEG imaging of toddlers during dyadic turn-taking: Mu-rhythm modulation while producing or observing social actions. *NeuroImage*, *112*, 52-60.

Lieven, E. V. (1994). Crosslinguistic and crosscultural aspects of language addressed to children.

Meyer, M., Hard, B., Brand, R. J., McGarvey, M., & Baldwin, D. A. (2011). Acoustic packaging: Maternal speech and action synchrony. *IEEE Transactions on Autonomous Mental Development*, *3*(2), 154-162.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. Journal of Experimental Child Psychology, 162, 31-38.

Ogura, T., Dale, P. S., Yamashita, Y., Murase, T., & Mahieu, A. (2006). The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-playing contexts. *Journal of Child Language*, *33*(1), 1-29.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics surveys*, *3*, 96-146.

Rohlfing, K. J., Wrede, B., Vollmer, A. L., & Oudeyer, P. Y. (2016). An alternative to mapping a word onto a concept in language acquisition: pragmatic frames. Frontiers in psychology, 7, 470.

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental science*, *15*(5), 659-673.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences*, *22*(4), 325-336.

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.

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.

Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: Effectiveness of a parenting program in rural Senegal. *Child Development*, *88*(5), 1513-1526.

APPENDIX A: Supporting information for Chapter 2

Table A.1. Conditional probability that an utterance contains content type in column given that it contains the content type in row.

	Dec	**Imp**	**Que**	**Dsc**	**Att**	**Act**	**ONm**	**INm**	**Aff**	**Soc**
Dec		.05	.06	.30	05	.16	.38	.02	.03	.03
Imp	.08		.07	.03	.53	.51	.20	.10	.02	.02
Que	.06	.05		.09	.09	.36	.33	.04	.01	.02
Dsc	.67	.04	.20		.04	.02	.25	.01	.03	.01
Att	.10	.80	.20	.04		.07	.22	.11	.02	.01
Act	.20	.44	.44	.01	.04		.25	.05	.02	.04
ONm	.41	.15	.36	.12	.11	.22		.03	.03	.04
INm	.09	.33	.19	.02	.23	.20	.12		.01	.17
Aff	.10	.04	.04	.04	.02	.05	.07	.00		.03
Soc	.07	.03	.05	.01	.01	.07	.06	.09	.03	

Table A.2. Cohen's Kappa for coded variables

Content type	κ
DECLARATIVE	.89
IMPERATIVE	.91
QUESTION	.96
DESCRIPTION	.76
ATTENTION	.94
ACTION	.86
OBJECT NAME	-- *
INFANT'S NAME	.93
AFFIRMATION	.92
SOCIAL ROUTINE	.80
Gaze target	.76
Object handling	.81

* This was tagged using a word list, so it did not differ across coders

Table A.3. Representative maternal speech sequence

Utterance	Content types	Gaze target	Toys handled
Hey [name]	SOCIAL ROUTINE, INFANT'S NAME	TOY	1
Look at that block	IMPERATIVE, ATTENTION, OBJECT NAME	OTHER	1
Can you give it to me?	QUESTION, ACTION	TOY	2
Give it to me.	IMPERATIVE, ACTION	TOY	2
Oh thank you.	SOCIAL ROUTINE	TOY	2
Thank you.	SOCIAL ROUTINE	TOY	1
How about the box?	QUESTION, OBJECT NAME	TOY	1
I don't, you don't like these blocks do you?	DECLARATIVE, OBJECT NAME	OTHER	0
You just want me to play with them, huh.	DECLARATIVE	TOY	0
[Toys exchanged]			
I know but we got new toys to play with.	DECLARATIVE	TOY	0
Hey look	IMPERATIVE, ATTENTION	TOY	2
Look at this	IMPERATIVE, ATTENTION	TOY	1
Look at what we can do, we can make a ball.	IMPERATIVE, ATTENTION, DECLARATIVE, ACTION, OBJECT NAME	TOY	1
We can make a ball.	DECLARATIVE, ACTION, OBJECT NAME	TOY	1
We can make a bigger ball.	DECLARATIVE, ACTION, OBJECT NAME, DESCRIPTION	TOY	1
wow.	SOCIAL ROUTINE	TOY	1

Table A.4. Coefficients for repetition of utterance content types. The first column reproduces the diagonal of Figure 2. The second column shows re-computed coefficients when exact repetitions are removed.

Type	β with exact repeats	β without exact repeats
DECLARATIVE	.63**	.38**
IMPERATIVE	.60**	.34†
QUESTION	.83**	.55**
DESCRIPTION	1.89**	1.71**
ATTENTION	.94**	.63**
ACTION	1.27**	1.10**
OBJECT NAME	1.66**	1.42**
INFANT'S NAME	.90**	.48†
AFFIRMATION	.41*	-.20
SOCIAL ROUTINE	1.07**	.62**

Table A.5. Proportion of utterances with each content type that are followed by exact repetitions.

Type	Proportion exact repetitions	Proportion overlapping words	Proportion not overlapping words
DECLARATIVE	.08	.58	.35
IMPERATIVE	.15	.52	.33
QUESTION	.14	.58	.29
DESCRIPTION	.09	.56	.35
ATTENTION	.15	.61	.24
ACTION	.12	.59	.30
OBJECT NAME	.14	.71	.15
INFANT'S NAME	.28	.62	.10
AFFIRMATION	.41	.23	.36
SOCIAL ROUTINE	.32	.39	.30

APPENDIX B: Supporting Information for Chapter 3

Table B.1. Variance Inflation Factors for distributional features

	Production			Comprehension		
Variable	**Coefficient**	**SE**	**VIF**	**Coefficient**	**SE**	**VIF**
Syntactic PC1	-.53	.17	3.24	-.52	.25	3.18
Syntactic PC2	-.27	.14	2.17	.27	.22	2.42
Syntactic PC3	-.09	.18	3.83	.06	.28	3.11
Syntactic PC4	-.81	.13	1.86	-.47	.22	2.30
Syntactic PC5	-.48	.13	2.03	-.47	.21	1.87
Syntactic PC6	-.02	.11	1.43	.21	.18	2.02
Syntactic PC7	-.50	.11	1.25	-.93	.20	1.90
Syntactic PC8	-.10	.12	1.54	-.51	.23	2.46
Syntactic PC9	-.29	.12	1.50	-.14	.21	1.83
Syntactic PC10	-.03	.12	1.56	.10	.21	2.07
Thematic PC1	-.48	.22	5.30	-.38	.35	3.94
Thematic PC2	-.21	.14	2.10	.18	.19	2.26
Thematic PC3	-.03	.13	1.80	.68	.19	2.11
Thematic PC4	-.01	.15	2.64	.38	.24	3.29
Thematic PC5	-.45	.11	1.29	-44	.16	1.54
Thematic PC6	-.10	.11	1.33	.34	.16	1.54
Thematic PC7	-.67	.10	1.13	-.1.01	.15	1.23
Thematic PC8	-.22	.10	1.17	-.45	.15	1.36
Thematic PC9	-.14	.10	1.07	-.16	.14	1.19
Thematic PC10	-.06	.10	1.16	-.34	.15	1.28

Table B.2. Variance Inflation Factors for baseline features

	Production			Comprehension		
Variable	**Coefficient**	**SE**	**VIF**	**Coefficient**	**SE**	**VIF**
MLU	.60	.17	3.43	.37	.32	5.11
Frequency	-1.82	.25	7.34	-.76	.39	5.98
Final frequency	-.89	.15	2.43	-.84	.26	2.59
Solo frequency	-1.59	.18	3.48	-1.31	.32	4.39

Table B.3. RMSE for different values of minimum frame/context word frequency

Min Frequency		500	1000	2000
Production				
	Full	2.47	2.49	2.54
	Baseline	2.84	2.84	2.84
	Syntactic+Thematic	3.24	3.31	3.30
	Baseline+Thematic	2.69	2.68	2.68
	Baseline+Syntactic	2.53	2.57	2.63
Comprehension				
	Full	2.81	2.83	2.92
	Baseline	3.31	3.31	3.31
	Syntactic+Thematic	3.15	3.16	3.24
	Baseline+Thematic	3.00	2.97	2.95
	Baseline+Syntactic	3.00	3.01	3.08

Table B.4. RMSE for different numbers of PCs

PCs		5	10	20
Production				
	Full	2.60	2.49	2.46
	Baseline	2.84	2.84	2.84
	Syntactic+Thematic	3.49	3.31	3.45
	Baseline+Thematic	2.84	2.68	2.67
	Baseline+Syntactic	2.61	2.57	2.60
Comprehension				
	Full	3.06	2.83	2.91
	Baseline	3.31	3.31	3.31
	Syntactic+Thematic	3.44	3.16	3.15
	Baseline+Thematic	3.22	2.97	3.00
	Baseline+Syntactic	3.09	3.01	3.12

Table B.5. RMSE for different winsorization thresholds

Winsorization		5 MAE	10 MAE	20 MAE
Production				
	Full	2.49	2.49	2.52
	Baseline	2.84	2.84	2.84
	Syntactic+Thematic	3.24	3.31	3.35
	Baseline+Thematic	2.69	2.68	2.68
	Baseline+Syntactic	2.59	2.57	2.60
Comprehension				
	Full	2.78	2.83	2.85
	Baseline	3.31	3.31	3.31
	Syntactic+Thematic	3.11	3.16	3.20
	Baseline+Thematic	2.97	2.97	2.97
	Baseline+Syntactic	3.00	3.01	3.03

APPENDIX C: Supporting Information for Chapter 4

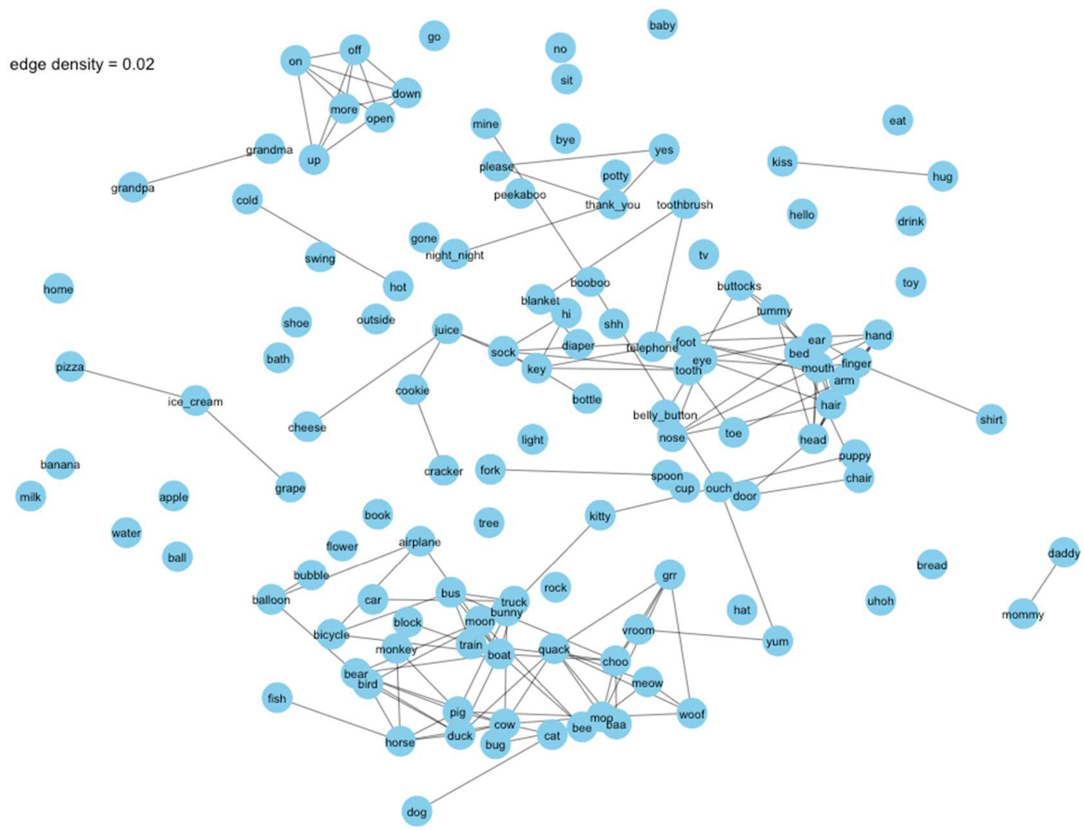

Figure C.1. Log-odds-ratio subnetwork consisting of the 125 earliest-learned words, with connections shown when LOR was greater than the 98th percentile of word pairs.

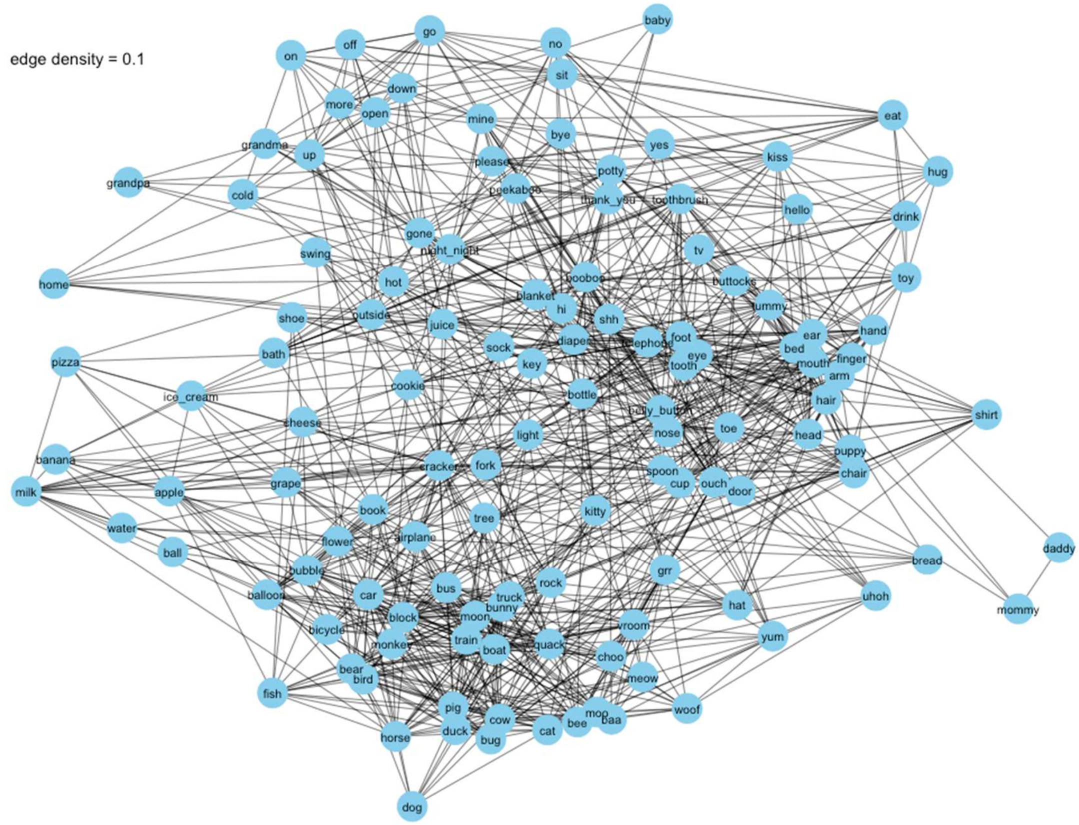

Figure C.2. Log-odds-ratio subnetwork consisting of the 125 earliest-learned words, with connections shown when LOR was greater than the 90th percentile of word pairs.

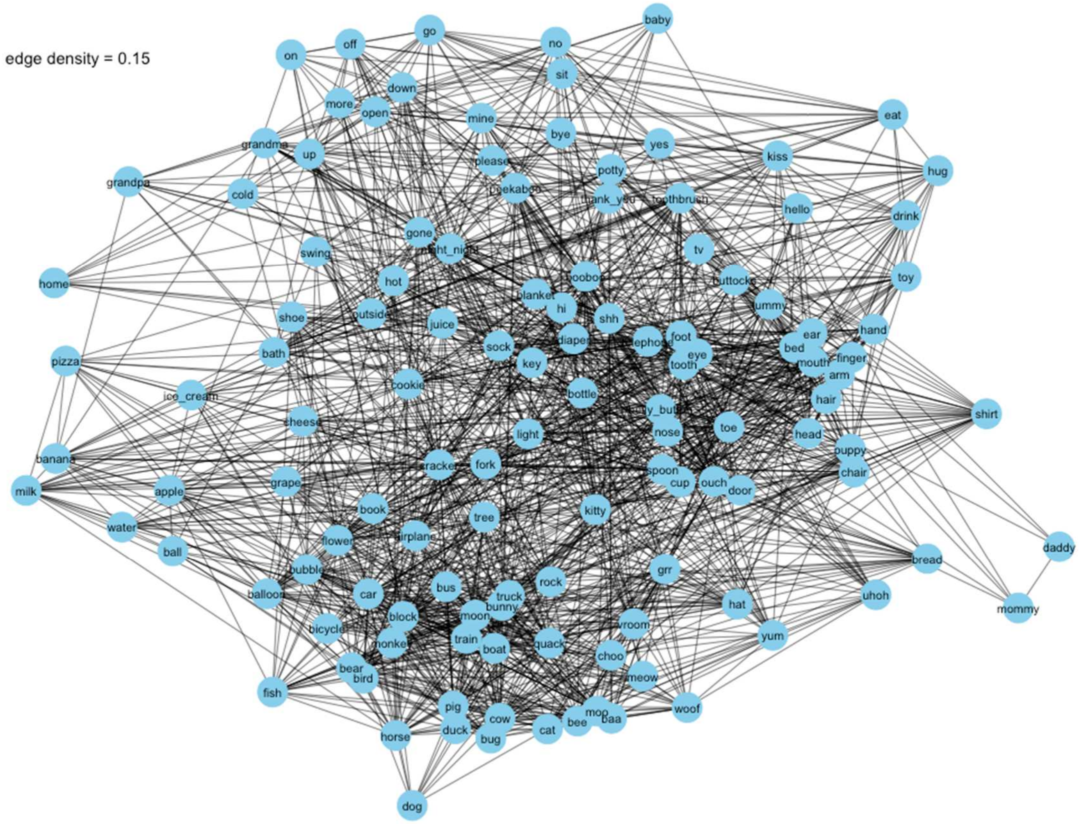

Figure C.3. Log-odds-ratio subnetwork consisting of the 125 earliest-learned words, with connections shown when LOR was greater than the 85th percentile of word pairs.

Table C.1. F-statistics for models predicting network measures as a function of vocabulary quartile group (both as categorical variable and as linear trend) are robust to different methods of correcting for vocabulary size. All models reached statistical significance by the methods described in the text.

	Network measure	Reweighting by quartile	No reweighting	Bandwidth 5	Bandwidth 20
Categorical models	Mean Edge	20.1	19.9	20.2	19.9
	Clustering Coef.	8.1	8.0	8.4	7.5
	Harmonic Path	26.4	26.1	26.5	26.7
	Deg. Dist. Skewness	4.3	4.3	4.3	4.3
Linear Trend models	Mean Edge	54.5	54.1	53.9	55.0
	Clustering Coef.	12.1	12.0	12.5	11.6
	Harmonic Path	73.3	74.3	75.3	75.1
	Deg. Dist. Skewness	11.8	11.8	11.4	11.7