The role of incremental parsing in syntactically conditioned word learning

Jeffrey Lidz
University of Maryland

Aaron Steven White
Johns Hopkins University

Rebecca Baier

In a series of three experiments, we use children's noun learning as a probe into their syntactic knowledge as well as their ability to deploy this knowledge, investigating how the predictions children make about upcoming syntactic structure change as their knowledge changes. In the first two experiments, we show that children display a developmental change in their ability to use a noun's syntactic environment as a cue to its meaning. We argue that this pattern arises from children's reliance on their knowledge of verbs' subcategorization frame frequencies to guide parsing, coupled with an inability to revise incremental parsing decisions. We show that this analysis is consistent with the syntactic distributions in child-directed speech. In the third experiment, we show that the change arises from predictions based on verbs' subcategorization frame frequencies.

Keywords: language acquisition, parsing, prediction, thematic roles

1 Introduction

In language acquisition, and indeed in many areas of child development, researchers often find themselves struggling with questions of competence and performance. Do children fail at some task because they lack the relevant knowledge or because that knowledge is masked behind the performance systems used to deploy that knowledge (Hamburger and Crain, 1984; Spelke and Newport, 1998)? Rarely, however, do we face the question of how children's developing performance systems constrain the generalizations that they ultimately make and how errors of interpretation feed forward for subsequent learning (Elman, 1990; Newport, 1990). In this paper, we take up this issue in the domain of syntactic development and word learning. In particular we ask how children's immature parsers lead to the assignment of erroneous grammatical structures and how such errors contribute to the acquisition of unknown words in those structures. This paper thus contributes to discussions of syntactic development, the role of syntax in word learning, and the role of parsing in syntactic development.

In understanding the interaction between parsing and learning, it is important to consider ways that parsing impacts understanding. We can consider two situations. First, the child may have acquired the grammatical rules for some construction without being able to deploy this knowledge consistently and robustly in real time (Hamburger and Crain, 1984; Huang et al., 2013; Omaki et al., 2014; Trueswell et al., 1999). Second, the child may not have acquired a given grammatical construction but nonetheless succeeds in interpreting sentences exhibiting it due to heuristics that promote understanding without relying on precise grammatical knowledge (Gagliardi et al., 2016; Gertner and Fisher, 2012; Yuan et al., 2012).

The case of successful acquisition of the grammatical rules in the absence of a robust deployment system can lead children to fail at accurately interpreting sentences for which they have appropriate grammatical knowledge. This could happen because the construction places high demands on component processes of understanding, such as lexical access, structure building, temporary ambiguity resolution, or retrieval from working memory, making the child's success with the construction dependent on the ease with which these subprocesses can be completed. For example, if a sentence uses low frequency words that are difficult to access from the lexicon, or if it contains a temporary ambiguity, then demands on the parser could be amplified in a way that hinders understanding, despite the child having an appropriate grammar for that construction.

In older children, there is mounting evidence that parsing dynamics shape understanding in a way that gives rise to children behaving in non-adult-like ways (Snedeker and Trueswell, 2004; Trueswell et al., 1999). For example, Trueswell et al. (1999) show that in certain discourse contexts, both adults and 5-year-old children initially interpret the first PP (*on the napkin*) in (1) as if it were the locative argument of the verb. Whereas adults can recover from this initial misinterpretation upon encountering the second PP (*in the box*), children have difficulty doing so.

(1) Put the frog on the napkin in the box.

Similar behavior has been found in at least four other domains: pronoun resolution, WH question interpretation, argument structure construction, and quantifier scope computation.

In the domain of pronoun resolution, Leddon and Lidz (2006) find that 4-year-old children only resolve reflexive pronouns to the closest syntactically licit antecedent, even in the presence of other licit antecedents. For instance, 4-year-olds resolve *herself* in (2) to Janie but not to Miss Cruella.

(2) Janie knew which picture of herself Miss Cruella put on the wall.

They argue that this bias derives from the ballistic nature of the parser, which links the reflexive pronoun to an antecedent as quickly as possible (Sturt, 2003), coupled with children's inability to revise their initial interpretive commitments (Trueswell et al., 1999). Indeed, recent eye-tracking work (Omaki, 2010) shows that adults also initially resolve *herself* in (2) to Janie, but unlike children, are able to revise that initial commitment when necessary.

In the domain of WH question interpretation, Omaki et al. (2014) find that adults and 5-year-old children prefer to associate adjunct WH words like *where* in (3) to the closest verb in terms of linear order (*say*).

(3) Where did Lizzie say that she was going to catch butterflies?

This finding is cross-linguistically robust. In Japanese, which is a head-final language, the order of *say* and *catch* is reversed. Omaki et al. found that the biases displayed by Japanese-speaking adults and 5-year-old Japanese-learning children were concomitantly flipped: both adults and children prefer to associate *where* with *catch* in Japanese.

In the domain of argument structure construction, Huang et al. (2013) find that, upon hearing a subject that is a plausible agent, 5-year-old Mandarin-learning children begin to construct an active interpretation for the sentence, and if they receive information that the sentence is actually passive, they have trouble recovering from this initial misparse. Huang and Arnold (2016) find a similar pattern in a word-learning task with 5-year-old English-learning children.

Finally, in the domain of quantifier scope computation, Musolino et al. (2000); Musolino and Lidz (2003,0) find that 5-year-old children are heavily biased towards interpreting sentences like (4) as meaning (4-a) but not (4-b).

(4) Every horse didn't jump over the fence.
    a. All of the horses failed to jump over the fence.
    b. Not every horse jumped over the fence.

Conroy (2008) and Viau et al. (2010) argue that children's bias results from the interpretation in (4-a) being the first interpretation constructed, paired with children's difficulty to revise their initial parsing commitments. Support for this view comes from several adult on-line parsing studies demonstrating that children's only interpretation corresponds to adults' initial interpretation (Conroy et al., 2008; Lidz and Conroy, 2007).

Given these and other findings showing that preschool aged children's parsers are more brittle than adults' and less able to integrate across multiple information sources (Choi and Trueswell, 2010; Omaki, 2010; Snedeker and Trueswell, 2004), it stands to reason that younger children will be at least as susceptible to failures of understanding due to parsing difficulty as older children are. Moreover, to the degree that parsing derails understanding, we expect that any process of language development that depends on information that would have been gleaned from a successful parse of a given sentence will succeed only to the degree that the parse can be accomplished (Trueswell et al., 2012).

In the domain of word learning, there are several recent findings suggesting that successful parsing of the initial part of a sentence is a prerequisite for learning words downstream. For example, Lidz et al. (2010) examine 24-month-old children's ability to learn a novel intransitive verb as a function of the character of the subject NP. They find that children are able to learn a novel verb when its subject is a pronoun (*it is blicking*) but not when it is a lexical NP (e.g., *the truck is blicking*) (see also Arunachalam et al., 2013; Childers and Tomasello, 2001). They argue that this asymmetry derives from the fact that pronouns are both more frequent and have less complex semantic representations than lexical nouns, making lexical access easier for pronouns.

This hypothesis is further supported by the observation that facilitating lexical access for the full NP subject causes the asymmetry to go away. By using the lexical noun in several sentences prior to the verb learning trial, lexical access for that word is facilitated in the sentence containing the novel verb. In turn, easier lexical access of the subject NP makes verb-learning easier. (See also Arunachalam and Waxman 2011; Yuan et al. 2011 for evidence that arguments with richer lexical content can improve verb-learning for transitive verbs.)

Similarly, Marchman and Fernald (2008) show that infants who are faster in interpreting familiar words in continuous speech are also more successful in learning novel words downstream, suggesting that early fluency in parsing and understanding has cascading consequences for word learning. Such findings are all the more important in light of recent evidence that socioeconomic status correlates with real-time processing ability, both at the level of lexical access (Fernald et al., 2013) and at the level of argument structure construction (Huang et al., 2017).

2 Parsing, syntactic inference and argument structure

Consider sentences (5) and (6).

(5) She's pushing the tiv.

(6) She's pushing with the tiv.

Even without a referential context, one can conclude that in (5) *the tiv* refers to the patient of the pushing event (i.e., the pushee) whereas in (6) *the tiv* refers to the instrument of pushing. This conclusion derives from a link between syntactic position and thematic relations (i.e., the role that an individual plays in an event): direct objects are generally interpreted as patients and the object of the preposition *with* is generally interpreted as an instrument. If we add a referential context to these sentences—for example, a scene in which a woman pushes a block with a truck—then we can use the conclusion about the thematic relation between the novel NP and the verb, to determine the referent of the phrase containing the novel noun, and hence the meaning of that word. *The tiv* refers to the block in (5) but the truck in (6).

Because these inferences depend on syntactic structure, they provide an ideal window into the development of real-time mechanisms for constructing phrase structure representations (Altmann and Kamide, 1999; Gordon and Chafetz, 1990; MacDonald et al., 1994; Snedeker and Trueswell, 2004; Trueswell et al., 1993). Moreover, because lexical differences in subcategorization frequency have been shown to play a critical role in guiding initial parsing decisions in adults and older children, probing the emergence of these effects in the earliest learners can also help us to determine the degree to which links between thematic relations and syntactic positions are acquired via a generalization over individual verbs (Dowty, 1991; Tomasello, 2000) or from more abstract principles of structure mapping (Gleitman, 1990; Pinker, 1989).

A long line of research beginning with Brown (1957, 1973) examines the role of syntax in driving inferences about word meaning. One stream of this research has shown that children can use the syntactic category of a novel word to make inferences about its meaning (see Waxman and Lidz 2006 for a review). For example, Waxman and Markow (1995) show that by 12-months of age English learning infants expect a novel word presented as a noun to refer to a category of objects, but have less specific expectations for a novel word presented as an adjective. By 14-months, the expectations for adjectives become more specific, with infants concluding that a novel adjective refers to an object property (e.g., color/texture) and not to an object category (Waxman and Booth, 2001). At least by 18-months, infants expect a novel verb to refer to a category of events and not to a category of objects (Bernal et al., 2007; He and Lidz, 2017; Waxman et al., 2009).

Beyond inferences from syntactic category to meaning, a number of studies have shown that infants and toddlers can use the number and type of arguments in a sentence to make inferences about the meaning of a novel verb in that sentence (Gleitman, 1990; Yuan and Fisher, 2009). For example, Naigles (1990) shows that 25-month-old children use the transitivity of a clause as a cue to whether a novel verb in that clause refers to a pair of events related by causation (for transitives) or temporal synchrony (for intransitives)—see Fisher et al. 2010 for a review.

It remains a question what these early syntactic representations look like and how they are linked to semantic representations. For instance, Gertner and Fisher (2012) show that, under certain conditions, 21-month-olds map conjoined subject sentences, such as (7), to causal events, just as they do for transitive clauses like (8).

(7) The boy and the girl are gorping.

(8) The boy is gorping the girl.

They argue that this effect derives from children not having acquired adult-like rules for building clause structure. Instead, they simply take the number of nouns in a sentence as a proxy for the number of arguments.

Pozzan et al. (2016) present suggestive evidence for a processing-based account of this phenomenon. They argue that children's error derives from an initial commitment to treat the first noun phrase (*the boy*) as subject of the sentence, and hence the agent of the gorping event—a commitment that is difficult for them to revise.

The idea that parsing plays an important role in argument structure learning comes from Trueswell et al. (2012). Those authors show that morphological cues to argument structure are used more effectively by children learning Tagalog, where the cue comes early, than in Kannada, where the cue comes late.

This idea is further bolstered by Pozzan and Trueswell (2015), who build on a long-circulating draft of the current paper (see *ibid* p 3-4). Using an artificial language learning experiment with adults, they show that the point in a sentence at which one receives morphological cues to an argument's thematic relation has consequences for how well the link between those morphological cues and the thematic information is learned. The later in the sentence that cue comes, the more challenging the link is to learn.

They argue that this arises from (i) an interaction between a baseline bias to map NPs to distinct thematic roles (cf. Gertner and Fisher, 2012) and (ii) a dispreference for revising structures that the parser has already constructed (cf. Trueswell et al., 1999). Then, when the morphological cues about an argument's thematic relation comes later in a sentence, after construction of the initial linking of that argument with some thematic relation, the dispreference for revision makes it difficult to correct the initial default linking.

While this kind of work compellingly demonstrates chil-

Table 1
An example of a single test trial.

Phase	Length	Video	Audio
Pre-trial	2 seconds 5 seconds	Blank screen Smiling baby	*Silence* Baby giggle
Familiarization	15 seconds	Camera being wiped by a cloth	Hey, look at that! She's wiping (with) the tig! Wow, do you see her wiping (with) the tig? Yay, she's wiping (with) the tig!
Test	2 seconds 2 seconds 3 seconds	Blank screen Split screen: camera and cloth	Where's the tig? *Silence* Which one's the tig?

dren's ability to use syntactic information to draw inferences about word meaning, the range of syntactic environments that has been examined to date is relatively narrow. In addition, the kinds of inferences that learners must make from syntactic distribution to verb meaning are somewhat indirect. The syntactic environment provides some evidence about the thematic relations to the verb as well as which NPs bear those thematic relations. The thematic relations provide some evidence about which event is being referred to, which in turn provides evidence about the meaning of the verb. In what follows, we expand the range of syntactic environments that trigger semantic inferences.

In addition, we focus in on the first piece of this process: what do children know about the link between the syntactic context of an NP and its interpretation? There is some evidence that, as early as 19 months of age, children can use a verb's meaning to learn the meanings of nouns that head that verb's subject NPs (Ferguson et al., 2014), but little is known about how robust this ability is across syntactic contexts. By directly exploring this link, we can understand children's knowledge of argument structure without relying on their ability to make complex inferences from argument structure to verb meaning.

3 Experiment 1

Experiment 1 examines how infants use a syntactic context of a noun phrase (NP) to make inferences about its thematic relation. Using a word-learning task in the intermodal preferential looking paradigm (Hirsh-Pasek and Golinkoff, 1999; Spelke, 1976), we tested children's abilities to assign a meaning to a novel noun contained in a direct object NP as compared to a prepositional object NP and a syntactically uninformative control. The particular version of the IPL paradigm that we use here is based on the one used by Waxman et al. (2009).

In adult English, the NP containing the novel word is interpreted as a patient in (9) but as an instrument in (10). In (11), there is no syntactic cue to the meaning of the novel word.

(9) She's pushing **the tiv**.

(10) She's pushing *with* **the tiv**.

(11) It's **a tiv**.

If children are able to use this thematic role information to learn the meaning of a novel noun, in (9), we expect them to be able to link *the tiv* to the object being pushed, or in (10), to the object used to do the pushing.

3.1 Method

3.1.1 Apparatus and procedure. Each infant arrived with his/her parent and was entertained by a researcher with toys while another researcher explained the experiment to the parent and obtained informed consent. The infant and parent were then escorted into a sound proof room, where the infant was either seated on the parent's lap or in a high chair, centered six feet from a 51" television, where the stimuli were presented at the infant's eye-level. If the infant was on the parents' laps, the parents wore visors to keep them from seeing what was on the screen. Each experiment lasted approximately 5 minutes, and infants were given a break if they were too restless or started crying.

In the case that the infant did not complete the experiment or were extremely fussy over the entire course, this was noted for later exclusion from the sample. Crucially, this annotation was done prior to coding and analysis in order to avoid biasing the sample.

Each infant was recorded during the entire experiment using a digital camcorder centered over the screen. A researcher watched the entire trial with the audio off on a monitor in an adjacent room and was able to control the camcorder's pan and zoom in order to keep the infant's face in focus throughout the trial. Videos were then coded offline frame-by-frame for direction of look by a research assistant blind to the experimental condition and without audio using the SuperCoder program (Hollich, 2005).

3.1.2 Design. Participants were presented with eight trials, each involving a different verb and concomitant scene.

Each of these trials was separated into two phases: the familiarization phase and the test phase. These phases are described below and Table 1 gives a sample script.

3.1.2.1 Familiarization Phase. During the familiarization phase, children were shown videos of 15 second dynamic scenes involving three objects: a human hand, an instrument manipulated by the hand, and a patient causally affected via the instrument. A recorded linguistic stimulus of the form either *she's* VERB*ing the* NOVEL NOUN, *she's* VERB*ing with the* NOVEL NOUN, or *it's a* NOVEL NOUN was associated with each scene. Each of these pairing constitute a level in the between-subjects STRUCTURE factor. VERB and NOVEL NOUN in these frames were replaced with a known verb and a novel noun. All linguistic stimuli were recorded by the same adult female. The linguistic stimulus was presented three times as the scene progressed with different lead-in words—e.g. *Look!*. This phase was coded for whether an infant was looking at the screen or not on any particular frame.

3.1.2.2 Test Phase. A blank screen was then shown for two seconds after each scene, during which the question *where's the* NOVEL NOUN*?* was asked once. The test video began at the offset of the novel noun in the first of these questions, when a screen with separate static images of both the instrument and the patient from the previous dynamic scene was displayed. One of these images took up approximately one third both by-width and by-height of the left portion of the screen and the other took up approximately one third by-width and by-height of the right portion, with an approximately one-third by-width separation in the middle of the screen. The side on which the instrument appeared was counterbalanced and pseudorandomized such that the instrument did not show up on the same side more than twice in a row.

Two seconds after the two images were presented, the question—*which one's the* NOVEL NOUN*?*—was played. The split screen was presented for five seconds total, after which the screen went blank. After a two second blank screen, either the next learning phase started or an attention-getting phase involving a picture of an infant and laughter was presented. This phase was coded for the direction the infant was looking (to the left side or to the right side) on any particular frame.

3.1.3 Materials. Eight verbs contained in the MCDI checklist were chosen with the criterion that their associated event concept must support the use of an instrument. Eight novel nouns were constructed and one associated with each verb. Table 1 gives a sample script summarizing the above description. In the *She's Ving with NP* conditions, children heard *with* during the familiarization, while those in the *She's Ving NP* conditions did not, represented in the table by the parentheses. In the *It's a NP* condition, these sentences were replaced with *it's a* NOVEL NOUN.

Table 2
The verbs and novel nouns used in the linguistic stimuli and the objects used in the visual stimuli for Exps. 1 and 2.

Verb	Noun	Instrument	Patient
wipe	*tig*	cloth	camera
throw	*frap*	cup	ball
hit	*tam*	ruler	cone
push	*gop*	bulldozer	block
touch	*pint*	pipe cleaner	pumpkin
wash	*pud*	sponge	toy car
tickle	*seb*	feather	mouse puppet
pull	*wug*	fishing pole	train

Table 2 shows each tuple of verb, novel noun, instrument object, and patient object. To control for possible order effects, we created two presentation orders for the trials by first building one pseudorandomized order according to the above sequencing criterion, then inverting it to create the second order. When crossed with the three linguistic structure levels (STRUCTURE: *She's Ving NP*, *She's Ving with NP*, *It's a NP*), this yielded six stimulus sets.

3.2 Participants

We recruited 48 16-month-olds (24 females) with a median age of 16;19 (mean: 16;18, range: 15;25 to 17;3) and 48 19-month-olds (24 females) with a median age of 19;20 (mean: 19;17, range: 18;29 to 20;5). Fourteen additional 16-month-olds and five additional 19-month-olds were tested but were excluded from the final sample for fussiness or inability to complete the experiment. This exclusion was done prior to analysis and was based on the exclusion annotation described in Section 3.1.1.

Participants were recruited from the greater College Park, MD area and were acquiring English as a native language. All participants heard English at least 80% of the time. Participants within each age group and sex were distributed evenly across the six stimulus sets.

Parents completed the MacArthur-Bates Communicative Development Inventory (MCDI) checklist (Fenson, 2007). By this index, the 16-month-olds' median productive verb vocabulary was 1 verb (mean: 3.1 verbs, IQR: 0–4 verbs), and their median productive total vocabulary was 29 words (mean: 41.2 words, IQR: 16–51.5 words); the 19-month-olds' median productive verb vocabulary was 3 verbs (mean: 12.4 verbs, IQR: 1–9.3 verbs), and their median productive total vocabulary was 59.5 words (mean: 90.7 words, IQR: 25.8–107 words). The parent of one 16-month-old participant in the *She's Ving NP* condition did not submit an MCDI checklist, and for the purposes of analysis, that participant's verb vocabulary value was set to the mean across 16-month-olds.

3.3 Preprocessing

We computed two measures for each trial each infant received. The first measure (FAMILIARIZATION PROPORTION) is the proportion of the time each infant was looking at the screen during the familiarization phase for a given trial. This measure provides a proxy for how well the infant was paying attention to the pairing of the linguistic stimulus with the scene in the video. We expect that the less an infant pays attention during a particular familiarization, the less likely it is that their behavior during the test phase that is associated with that familiarization provides provides evidence about the inferences they make based on the linguistic stimuli.

The second measure (OBJECT COUNT) is the number of frames on which each infant was looking at the instrument (LOOKS TO INSTRUMENT) paired with the number of frames on which they were looking at the patient (LOOKS TO PATIENT) on each trial.[1] This was calculated by converting the left-right coding of the test phase into an instrument-patient coding and then computing the relevant counts by trial for each infant. Note that, unlike the first measure, this second measure is not a proportion, though we can compute a proportion from it. For the purposes of visualization and basic comparisons of means, we work with proportions computed from these counts; for the purposes of more fine-grained analysis, we work with the counts themselves.

3.4 Results

To reiterate our predictions, if children both know the relation between syntactic position—in the current case, presence or absence of a preposition—and thematic relation—in the current case, instrument v. patient—and if they can furthermore deploy that knowledge, they will map the NP in *She's Ving NP* to the patient, and thus look more to the patient when they receive the *She's Ving NP* structure, and they will map the NP in *She's Ving with NP* to the instrument and thus look more to the instrument when they receive the *She's Ving with NP* structure.

Figure 1 plots the mean proportion of looks to instrument by STRUCTURE and AGE. The confidence intervals in Figure 1 are computed from a nonparametric bootstrap of the condition mean with 9,999 iterations. In this bootstrap, infants' mean proportion of looks to instrument across trials, weighted by FAMILIARIZATION PROPORTION, was first computed and then these mean proportions were resampled.

The first thing to note here is that in the mean proportion for the *It's an NP* structure for both the 16-month-olds and the 19-month-olds is significantly below chance ($ps < 0.001$). This suggests that both age groups have a reliable bias to look at the patient absent syntactic or other contextual cues. We refer to this bias as the *patient bias*.

The second thing to note here is that the 16-month-olds show effects in the predicted directions: in the *She's Ving NP*, infants look more to the patient on average than in *It's a NP*, while in the *She's Ving with NP*, infants look more to the instrument on average than in *It's a NP*. In contrast, 19-month-olds do not show such a pattern, tending to look at the patient at about the same rate regardless of structure. This pattern suggests a surprising developmental change: by 16-months, infants show behavior consistent with correctly mapping the structures they were presented to the correct thematic role, while by 19-months, evidence of this behavior is absent.

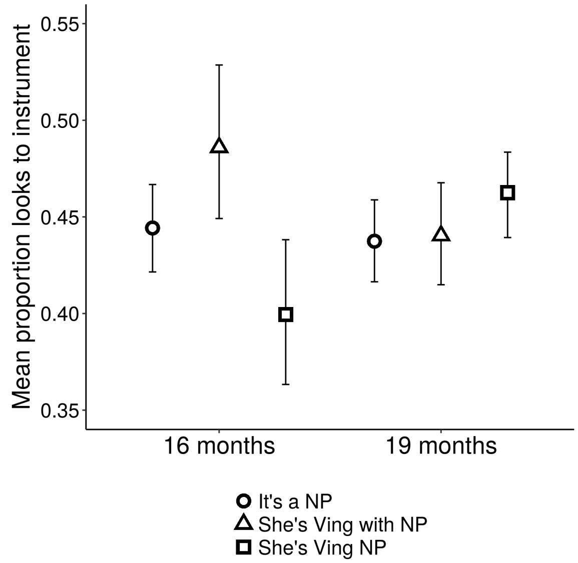

Figure 1. Mean proportion looks to instrument by STRUCTURE and AGE. Error bars show 95% confidence intervals computed from nonparametric bootstrap on participant weighted means.

To assess the reliability of this pattern, we use a logistic mixed effects model with OBJECT COUNT as the dependent variable, random intercepts for infant and item, by-item random slopes for STRUCTURE, AGE, and their interaction, and a loss weighted by FAMILIARIZATION PROPORTION.[2] We begin by fitting a model with fixed effects for STRUCTURE,

[1] Note that, because infants do not necessarily look at the screen during the entire test phase, the sum of LOOKS TO INSTRUMENT and LOOKS TO PATIENT will not necessarily be the number of frames in the test phase. This is actually a feature of OBJECT COUNT as a measure, since it retains information about the relative amount of data from which a probability is computed, where analyzing the proportion directly does not.

[2] An anonymous reviewer questions whether we in fact use a logistic mixed model here, and not a Poisson or negative binomial mixed model. We do in fact use a logistic mixed model here. This model is appropriate because the sum of LOOKS TO INSTRUMENT and LOOKS TO PATIENT is constrained to equal the total number of looks in a particular trial. For this same reason, a Poisson or negative binomial mixed model is inappropriate, since such a model

AGE, and their interaction under a sum coding and then test each effect in turn. We refer to this model as the *maximal model* (cf. Barr et al., 2013).

To test the interaction, we use a log-likelihood ratio test to compare the above full model with one that differs only in that it lacks the fixed interaction (retaining the by-item random slope interaction). This test suggests that the interaction is significant ($\chi^2(1)$ =8.54, p <0.05). A similar test for the by-item random slope interaction (retaining the fixed interaction) also yields significance ($\chi^2(11)$ =63.93, p <0.05). This suggests that the random effects structure that we employ is parsimonious (cf. Matuschek et al., 2017). Further, both fixed interaction terms are significant in the directions suggested by Figure 1 (ps <0.05).

To test the main effects, we compare a model containing the random effects structure from the maximal model but only the fixed main effects against two other models: one containing the random effects structure from the maximal model but only the fixed main effect of AGE and another containing the random effects structure from the maximal model but only the fixed main effect of STRUCTURE. Under these tests, neither the main effect of STRUCTURE ($\chi^2(2)$ =1.63, p = 0.44) nor the main effect of AGE ($\chi^2(1)$ =0.02, p = 0.90) is significant.

3.5 Discussion

The data in Experiment 1 support a developmental change on this task, with the syntactic frame influencing performance for 16-month-olds, but not 19-month-olds. The infants who are influenced by the syntactic context may be computing an inference by which the syntactic position of an NP containing a novel noun determines its thematic relation, which in turn allows the learner to identify the referent of the novel NP and hence to determine the meaning of the novel noun. The decline in performance raises two questions. First, how do the younger infants succeed at this task? Are they succeeding because they know the meaning of *with* or because of a heuristic that leads to correct behavior despite a lack of knowledge? Second, what is responsible for the change in behavior that appears between 16 months and 19 months? We address these questions in Experiments 2 and 3, respectively, but before moving on, we would like to sketch a potential answer for the second question.

Setting aside the 16-month-olds' behavior for the moment, one potential reason we do not see 19 months olds reliably constructing the correct mappings could be that infants have to know at least some verbs to be able to even start using the syntax to infer which thematic relation is assigned to an NP (which is in turn a prerequisite for using the thematic relation to infer the noun's meaning). Then, the failure of infants that don't know enough verbs might be obscuring the success of infants that do know enough verbs. This would make sense under, e.g., Tomasello's (2000) Verb Island Hypothesis, since under that hypothesis all thematic information is verb-specific (at least at this stage in development), and so children should not be able to construct mappings from syntactic structure to thematic structure for verbs they do not know. Of course, this hypothesis does not address why 16-month-olds as a whole appear to be succeeding, but if we find that, among 16-month-olds, those who succeed are those with larger verb vocabularies, we may have at least preliminary evidence for this verb-based hypothesis. Such a finding might in turn shed light on the reason that 19-month-olds fail.

As a suggestive exploratory analysis, then, we ask whether there is a correlation between verb knowledge and success at our task. If more verb knowledge helps infants map NPs to the correct referent in this task, we should see verb knowledge predicting better performance—i.e. there should be a negative correlation between vocabulary and looking to instrument for children who heard *She's Ving NP* and a positive correlation between vocabulary and looking to instrument for children who heard *She's Ving with NP*.

To assess this, we extracted the Best Linear Unbiased Predictors (BLUPs) for the 16-month-olds' participant random intercepts from the maximal model used in the above analysis. A more positive intercept means that the infant looked more to the instrument than the mean in their condition (controlling for item), and a negative intercept means that the infant looked more to the patient.

Figure 2 plots BLUPs for the 16-month-olds' participant random intercepts against VERB VOCAB. The lines show robust linear model fits for the two test conditions *She's Ving NP* and *She's Ving with NP*. What we see here is that, *contra* the hypothesis that more verb knowledge improves infants' ability to learn a noun based on the syntactic structure, more verb knowledge may well be a hindrance: the correlations are exactly the opposite of those predicted by such a hypothesis.

Insofar as it holds up under confirmatory scrutiny—and in ongoing work, we show that it does—this result is quite surprising, since it is unclear why lower verb knowledge should make infants better able to learn novel nouns (White et al., prep). But insofar as 16-month-olds know fewer verbs, it is at least consonant with the finding that 16-month-olds but

---

assumes support on the entirety of the nonnegative integers.

We suspect that this question arises from that the fact that it is common to code the dependent variable **Y** in a logistic regression as a binary number and then optimize $\boldsymbol{\beta}$ against the likelihood $\prod_{i,j} \text{logit}^{-1}(\boldsymbol{\beta}\mathbf{x}_i)^{y_{ij}}(1 - \text{logit}^{-1}(\boldsymbol{\beta}\mathbf{x}_i))^{1-y_{ij}}$. But this likelihood can be equivalently expressed as $\prod_i \text{logit}^{-1}(\boldsymbol{\beta}\mathbf{x}_i)^{\sum_j y_{ij}}(1 - \text{logit}^{-1}(\boldsymbol{\beta}\mathbf{x}_i))^{\sum_j 1-y_{ij}}$, in which case the exponents are nonnegative integers. This second formulation is the analogue of our analysis, where OBJECT COUNT = (LOOKS TO INSTRUMENT, LOOKS TO PATIENT) = $\left(\sum_j y_{ij}, \sum_j 1 - y_{ij}\right)$. This formulation is furthermore made straightforward in R `lme4`'s `glmer` function using the `cbind` syntax for dependent variables in multivariate regression. (To be clear, this is not a multivariate regression; the syntax used by `lme4` is just the same.)

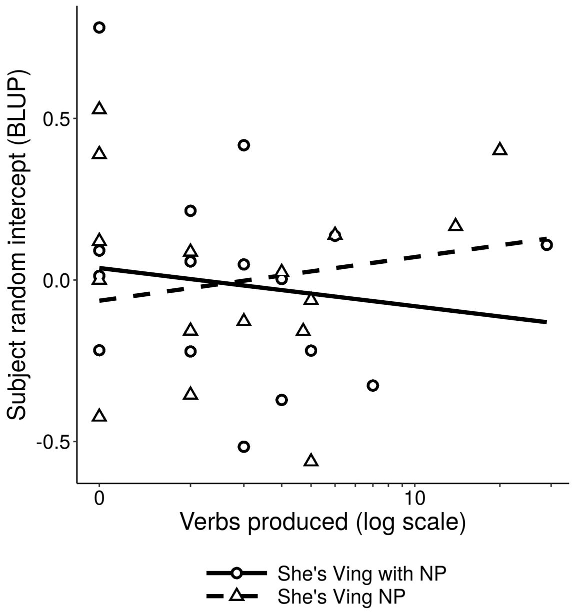

Figure 2. Best linear unbiased predictors for participant random intercepts plotted against VERB VOCAB with robust linear model fits.

not 19-month-olds are able to do this task. That is, if verb vocabulary is the real predictor of ability to correctly map NPs in our task to the correct referent—at least in this age range—then it makes sense that 16-month-olds, who tend to know fewer verbs, would do better on this task in aggregate. We return to this point in the General Discussion.

4 Experiment 2

Experiment 2 examines two groups of 16-month-olds in order to identify the source of their early success in Experiment 1. We consider two hypotheses. First, these children may be exhibiting a heuristic whereby any NP that is directly adjacent to the verb is interpreted as a patient and any NP that is not directly adjacent to the verb is interpreted as bearing some other thematic relation. Given the familiarization materials in Experiment 1, the only possible other role would be instrument. Initial success in this task would therefore be explained not by children representing a link between the object of *with* and the instrumental relation, but rather by a link between being a non-direct object and being a non-patient. Alternatively, these children may be succeeding because they understand the link between syntactic position and thematic relation. In particular, they know that direct objects are patients and that objects of *with* are instruments.

In order to tease these possibilities apart, we tested 16-month-olds in 2 conditions:

(12) She's pushing *on* **the tiv**.

(13) She's pushing *gub* **the tiv**.

These conditions have the following properties: (12) uses a different preposition that assigns a patient-like thematic role to its complement and (13) uses a novel preposition.

If the success of 16-month-olds in Experiment 1 derives from children not knowing the meaning of the prepositions, but using a parsing heuristic whereby any NP that is not adjacent to the verb is interpreted as a non-patient, then we would expect a similar pattern of behavior in Experiment 2 as in Experiment 1. In both conditions, the NP is not adjacent to the verb and so in both conditions we would expect the novel NP to be interpreted as an instrument.

However, if the early success derived from the children having knowledge of the content of the preposition *with*, then we would expect a different pattern of results here. If these children already know the content of the prepositions, then we expect them to look more at the patient in the *on* condition. Moreover, because there is no meaning associated with the novel preposition, children should not know what thematic relation to assign to its object and so we expect chance performance in that condition.

4.1 Method

4.1.1 Apparatus and Procedure. The apparatus and procedure used for this experiment were identical to that used for Experiment 1.

4.1.2 Design. All elements of the design were identical to that found in Experiment 1 except for the linguistic stimuli used in the training phase. Instead of hearing *she's* VERB*ing the* NOVEL NOUN or *she's* VERB*ing with the* NOVEL NOUN, infants heard either *she's* VERB*ing on the* NOVEL NOUN or *she's* VERB*ing gub the* NOVEL NOUN.

4.1.3 Materials. All materials were identical to those from Experiment 1. As noted above, the only change was to the linguistic stimuli during the training phase.

4.2 Participants

32 16-month-olds (16 females) with a median age of 16;19 (mean: 16;19, range: 16;3 to 17;5) were tested on our two new preposition conditions. As in Experiment 1, the new participants were recruited from the greater College Park, MD area and were acquiring English as native language. All participants heard English at least 80% of the time. Parents completed the MCDI checklist. By this index, the median productive verb vocabulary was 1 verb (mean: 6.8 verbs, IQR: 0–4 verbs), and their median productive total vocabulary was 25 words (mean: 48.8 verbs, IQR: 11.5–50.3 verbs). Ten additional infants were tested but were excluded from the final sample for fussiness or inability to complete the experiment. This exclusion was done prior to analysis and was based on the exclusion annotation described in Section 3.1.1.

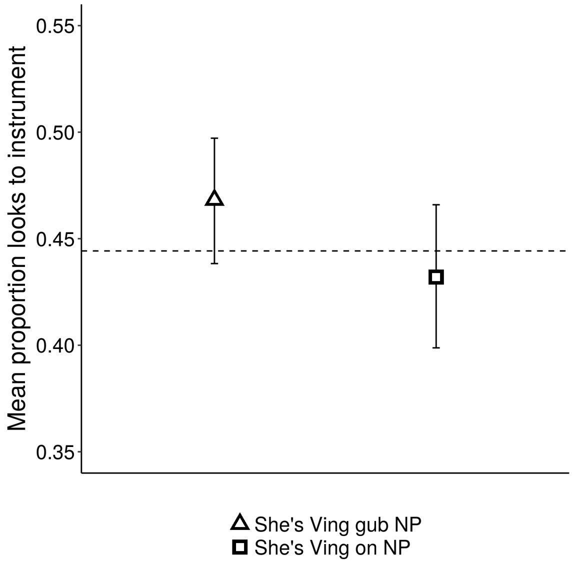

Figure 3. Mean proportion looks to instrument by STRUCTURE. Error bars show 95% confidence intervals computed from nonparametric bootstrap on participant weighted means. Dashed line shows estimated mean for the *It's a NP* structure for 19-month-olds from Experiment 1.

4.3 Preprocessing

All data preprocessing was identical to that conducted for Experiment 1.

4.4 Results

To reiterate our predictions, if children are using a heuristic wherein all prepositions are associated with non-patients, we expect that children will map the NP in both *She's Ving gub NP* and *She's Ving on NP* to the only non-patient in the scene—the instrument—and thus behave similarly to children in the *She's Ving with NP* condition in Experiment 1. But if children know that prepositions differ in the thematic roles they are associated with, we expect one of two outcomes.

On the one hand, if children know that *on* is associated with patients, we expect that they will map the NP in *She's Ving on NP* to the patient and thus behave similarly to children in the *She's Ving NP* condition in Experiment 1. On the other hand, if children do not know that *on* is associated with patients, *on* will not be a useful syntactic cue, and we thus expect that they will behave similarly to children in the *It's a NP* condition in Experiment 1. In both of these scenarios, we expect that children in the *She's Ving gub NP* conditions will behave similarly to children in the *It's a NP* condition in Experiment 1, since *gub* is never a useful syntactic cue.

Figure 3 plots the mean proportion of looks to instrument by STRUCTURE. As in Figure 1, the confidence intervals in Figure 3 are computed from a nonparametric bootstrap of the condition mean with 9,999 iterations. In this bootstrap, infants' mean proportion of looks to instrument across trials, weighted by FAMILIARIZATION PROPORTION, was first computed and then these mean proportions were resampled. The dashed line shows the estimated mean for the *It's a NP* structure for 16-month-olds from Experiment 1.

Since all three of our predictions are stated in terms of similarity with the results of Experiment 1, and not differences between conditions in the current experiment, we cannot use standard null hypothesis significance testing here. Instead, we employ a Bayesian analysis based on the one described in Gallistel 2009.

The basic idea behind this analysis is to ask the following question: given the 16-month-olds' data from Experiment 1, what is the probability that the data from the *She's Ving on NP* and *She's Ving gub NP* conditions arise from the same population as the data from the *It's a NP*, *She's Ving NP*, or *She's Ving with NP* conditions. Thus, by the end of this analysis, we will be left with six probabilities: the probability of the *She's Ving on NP* data given the *She's Ving NP*, the probability of the *She's Ving gub NP* data given the *She's Ving NP*, the probability of the *She's Ving on NP* data given the *She's Ving with NP* data, and so on.

The first step in this analysis is to construct, for each structure in Experiment 1 (*It's a NP*, *She's Ving NP*, *She's Ving with NP*), a posterior distribution over the parameters $\boldsymbol{\theta}$ of a random effects model with random intercepts for item and participant.[3] We denote the data for each structure as $\mathbf{D}_{cont}$, $\mathbf{D}_{DO}$, and $\mathbf{D}_{with}$, respectively, and we denote the respective posterior probability distributions for the parameters $\boldsymbol{\theta}$ of the random effects model as $\mathbb{P}(\boldsymbol{\theta} \mid \mathbf{D}_{cont})$, $\mathbb{P}(\boldsymbol{\theta} \mid \mathbf{D}_{DO})$, and $\mathbb{P}(\boldsymbol{\theta} \mid \mathbf{D}_{with})$.

The second step is to compute the probability of the data from the current experiment, $\mathbf{D}_{gub}$ and $\mathbf{D}_{on}$, given $\mathbf{D}_{cont}$, $\mathbf{D}_{DO}$, or $\mathbf{D}_{with}$. That is, we want to compute $\mathbb{P}(\mathbf{D}_j \mid \mathbf{D}_i)$ for all i in $\{cont, DO, with\}$ and all j in $\{gub, on\}$. To do this, we use the posterior distributions $\mathbb{P}(\boldsymbol{\theta} \mid \mathbf{D}_i)$ that we computed in the first step as priors in the posterior predictive distribution.

$$\mathbb{P}(\,\cdot \mid \mathbf{D}_i) = \int \mathbb{P}(\,\cdot \mid \boldsymbol{\theta})\, \mathbb{P}(\boldsymbol{\theta} \mid \mathbf{D}_i)\, \mathrm{d}\boldsymbol{\theta}$$

We define the likelihood of the hypothesis H_i that data $\mathbf{D}_j$ arose from the same population as dataset $\mathbf{D}_i$ using the above posterior predictive distribution.

$$\mathcal{L}(H_i \mid \mathbf{D}_j) = \mathbb{P}(\mathbf{D}_j \mid \mathbf{D}_i)$$

To compute this likelihood for each pairing of i and j we use Monte Carlo integration over 10,000 samples from

[3] Note that this model is the maximal model for the respective subsets of the data, since we are only looking at the 16-month-olds' data and then further splitting that data by the STRUCTURE variable.

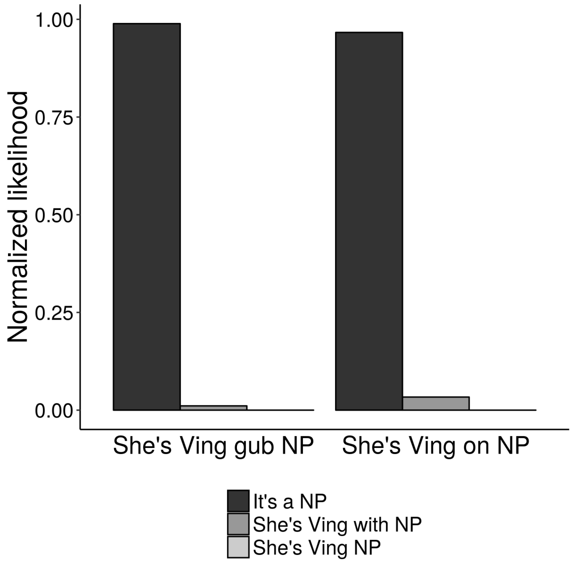

Figure 4. Normalized likelihood of data for each level of STRUCTURE in Experiment 2 given the 16-month-olds' data for each level of STRUCTURE in Experiment 1.

$\mathbb{P}(\boldsymbol{\theta} \mid \mathbf{D}_i)$. And to compare the resulting raw likelihood, we define a normalized likelihood.

$$\hat{\mathcal{L}}(H_i \mid \mathbf{D}_j) = \frac{\mathcal{L}(H_i \mid \mathbf{D}_j)}{\sum_k \mathcal{L}(H_k \mid \mathbf{D}_j)} = \frac{\mathbb{P}(\mathbf{D}_j \mid \mathbf{D}_i)}{\sum_k \mathbb{P}(\mathbf{D}_j \mid \mathbf{D}_k)}$$

These normalized likelihoods are plotted in Figure 4. What we see in this figure is that children's behavior with both *She's Ving gub NP* and *She's Ving on NP* is, far and away, most similar to their behavior with *It's a NP* from Experiment 1.

This suggests two things. First, 16-month-olds in Experiment 1 were not treating *with* as a generalized marker of non-patienthood—i.e. they know that *with* marks instruments. If they were, we should expect the middle bar in Figure 4 to be high and the other bars to be low for at least the *She's Ving gub NP* structure. Second, it does not appear that children know that *on* is a marker of patienthood. If they did, we should expect right bar to be high and the other bars to be low.

4.5 Discussion

Experiment 2 suggests that children do know the content of the preposition *with*. Whereas 16-month-olds in the *She's Ving with NP* condition in Experiment 1 looked to the instrument significantly more than in the *control* condition, children in the novel preposition *V gub NP* position do not.

Experiments 1 and 2 together also provide a novel argument against the view that the links between thematic relation and syntactic position are acquired on a verb-by-verb basis and only generalized after a sizeable verb vocabulary has been acquired (Tomasello, 2000). Because the children who succeed at using syntactic context to determine the thematic relation of the NP in Experiment 1 are reported to have little to no productive verb vocabulary, it cannot be the case that the thematic relations are constructed by a process of generalizing over the distributional and interpretive properties of known verbs. This argument goes through even if the MCDI does not provide a perfect measure of children's verb vocabulary. No version of the exemplar-driven generalization theory predicts that having a larger vocabulary would be detrimental to acquiring the link between syntactic position and thematic relations.

Returning now to the main thread, if we accept the conclusion that the non-verb-knowing 16-month-olds are aware of the relation between syntactic context and thematic relations, then we must determine the source of the dip in performance associated with the onset of a productive verb lexicon. What changes in the child's grammar or parser could cause them to fail to use information that they apparently already have?

We pursue the hypothesis that the dip in performance exhibited by verb-knowing 16-month-olds and 19-month-olds derives from developmental changes in the weighting of predictive vs. bottom-up cues in parsing. As children develop a larger verb vocabulary, they begin to use their knowledge of subcategorization frequencies to anticipate syntactic structure (Altmann and Kamide, 1999; Gordon and Chafetz, 1990; MacDonald et al., 1994; Trueswell et al., 1993). When these predictions conflict with bottom-up information from the sentence itself, they have difficulty resolving this conflict and rely instead on their early commitments. However, relying on early commitments comes at the expense of building a parse that is fully consistent with the bottom-up information.

To make this hypothesis more concrete, consider again sentence (10) from Experiment 1.

(10) She's pushing with the tiv.

Imagine that the child has heard the subject and the verb. At this stage, if the child expects the verb to be used transitively, it is possible to predict that a direct object NP is coming and to build that structure in advance of hearing it (Altmann and Kamide, 1999; Omaki, 2010; Omaki and Lidz, 2015; Sussman and Sedivy, 2003). When the next word turns out to be a preposition and hence is inconsistent with the predicted structure, the parser must revise its initial commitment in order to successfully parse the sentence. However, because this revision is too difficult for children to execute (Trueswell et al., 1999), they treat the object of the preposition as the object of the verb and effectively ignore the preposition for the purposes of parsing and interpretation.

If this hypothesis explains the pattern of data seen in verb-knowing 16-month-olds and 19-month-olds, then we can make several predictions. First, the verbs in our study are

Table 3
Subcategorization frame frequencies for verbs in Experiments 1 and 2 calculated using the Brown corpus from CHILDES (MacWhinney, 2014a,1) parsed by Pearl and Sprouse (2013)

Verb	$[_{VP}$ _ **NP]**	$[_{VP}$ _ **NP PP]**	$[_{VP}$ _ **PP]**	$[_{VP}$ _ $\textbf{PP}_{\textbf{with}}]$	**Total**
hit	136	53	5	0	218
pull	203	28	10	0	291
push	206	30	8	2	302
throw	196	82	9	1	310
tickle	22	0	0	0	39
touch	151	2	2	0	184
wash	144	13	3	0	186
wipe	68	13	0	0	87
Total	1126	221	37	3	1617

predicted to be significantly more likely to be used transitively than intransitively with a PP. This asymmetry is a presupposition of the account based on a differential weighting of predictive vs. bottom-up cues because subcategorization frequency can function as a predictive cue only to the degree that asymmetries in subcategorization frequencies exist. Second, if we could satisfy the verb's subcategorization preference in sentences containing a preposition, then we expect sensitivity to the content of the preposition to re-emerge. Third, if 19-month-olds were given a verb for which they had no subcategorization expectations, sensitivity to the preposition should re-emerge. Fourth, children with no productive verb vocabulary should behave identically with real and novel verbs. Finally, if 19-month-olds were given substantial exposure to a novel verb in one subcategorization frame, then their sensitivity to the preposition should be a function of the degree to which the preposition is consistent with that exposure. We address the first two predictions below, and we address the second two in other work (White et al., prep).

To test the first prediction, we examined the distribution of complement types for each of the 8 verbs used in Experiments 1 and 2 in Pearl and Sprouse's (2013) parsed version of the Brown corpus obtained from CHILDES (MacWhinney, 2014a,1). We asked what proportion of the instances of each verb occurred in a transitive clause not also containing a PP ($[_{VP}$ V _ NP]), an transitive clause also containing a PP ($[_{VP}$ V _ NP PP]), an intransitive clause containing a PP ($[_{VP}$ V _ PP]), or an intransitive clause containing a PP headed by *with* ($[_{VP}$ V _ $\text{PP}_{\text{with}}]$). The results of this search are given in Table 3.

The verbs that we used occurred on average 70% of the time in a transitive clause and .2% of the time in intransitive clauses with PPs headed by with. In addition, we also asked what proportion of all verbs in the corpus occurred in these 4 environments, finding that 33% occurred in transitive clauses with no PP, 7% occurred in transitive frames containing PPs, 13% occurred in intransitive clauses containing PPs, and 2% occurred in intransitive clauses containing PPs headed by with. These data are consistent with the hypothesis that children who fail to use the syntactic context to determine the thematic relation of the novel NP are doing so because they rely on their knowledge of subcategorization frequencies to guide their parsing decisions.

5 Experiment 3

Experiment 3 tests the prediction that satisfying the verb's subcategorization expectations in sentences containing a preposition would allow sensitivity to the content of the preposition to re-emerge. We hypothesized above that the children in Experiment 1 who failed to use syntactic context as a cue to meaning failed to do so because they were relying more on their knowledge of the verb's likely subcategorization than on the verb's actual subcategorization in the experiment. Thus, if we could find a way to test their knowledge of the relation between syntactic context and thematic relation while also putting the verb in its preferred syntactic context, then this knowledge should reemerge.

Consider (14) and (15).

(14) She's pushing that thing *with* **the tiv**.

(15) She's pushing **the tiv** *with* that thing.

Both of these sentences contain two referentially ambiguous expressions (*that thing*, *the tiv*). In (14), the novel word is used as the object of the preposition *with*. In (15) it is used as the direct object of the verb. But, without knowledge of the link between syntactic position and thematic relation, it would be impossible to know what the NP containing the novel word refers to. Hence, to the degree that children can use syntactic context to infer the meaning of the novel word, it follows that they represent the link between syntactic context and thematic structure. Moreover, because these clauses are all transitive, they satisfy the preferred subcategorization frame of the verb, allowing the effect of syntactic context to

Table 4
The verbs and novel nouns used in the linguistic stimuli and the objects used in the visual stimuli in Exp. 3.

Verb	Noun	Instrument	Patient
tap	*pint*	pipe cleaner	train
brush	*seb*	brush	mouse
stop	*frap*	block	ball
hit	*tam*	ruler	cone
wipe	*tig*	cloth	camera
push	*gop*	bulldozer	block

emerge independent of subcategorization preferences.

We use the same visual stimuli as in Experiment 1, with the two audio conditions in (14) and (15). If children are able to use syntactic position as a cue to thematic relation, then they should interpret the novel word as referring to the instrument in (14) but the patient in (15).

We tested 32 19-month-olds. This helps to determine the viability of our hypothesis for their failure in Experiment 1. If the 19-month-olds in Experiment 1 failed to use the preposition as a cue to meaning because they do not know the meaning of the preposition or the link between syntactic position and thematic relation, then they should be unable to identify the meaning of the novel word here. However, if they failed because they were relying on the subcategorization frequency of the verb, then they should succeed here. Because the verb occurs in its preferred syntactic environment, then if children have knowledge of the semantic contribution of the preposition, it should emerge here.

5.1 Method

5.1.1 Apparatus and Procedure. The apparatus and procedure used for this experiment were identical to that used for Experiment 1.

5.1.2 Design. All elements of the design were identical to that found in Experiment 1 except for (i) the form of linguistic stimuli used in the training phase and (ii) the number of trials. Instead of hearing *she's* VERB*ing the* NOVEL NOUN (*She's Ving NP* condition) or *she's* VERB*ing with the* NOVEL NOUN (*She's Ving with NP* condition), infants heard either *she's* VERB*ing the* NOVEL NOUN *with that thing* (*V NP with that thing* condition) or *she's* VERB*ing that thing with the* NOVEL NOUN" (*V that thing with NP* condition). Instead of seeing eight trials, children saw six.

5.1.3 Materials. Three verbs from the original experiment (*hit*, *push*, *wipe*) were retained along with their associated dynamics scenes and novel nouns. Three new verbs from the MCDI were used (*brush*, *stop*, *tap*) along with associated novel nouns. Two new pseudorandomized orders were created. Table 4 shows each tuple of verb, novel noun, instrument object, and patient object.

5.2 Participants

32 19-month-olds (16 females) with a median age of 19;9 (range: 18;27 to 19;29) were tested on our two-argument task. As in Experiments 1 and 2, participants were recruited from the greater College Park, MD area and were acquiring English as native language. All participants heard English at least 80% of the time. Parents completed the MCDI checklist. By this index, the median productive verb vocabulary was 3.5 (range: 0 to 96). Four additional infants were tested but were excluded from the final sample for excessive fussiness. This exclusion was done prior to analysis and was based on the exclusion annotation described in Section 3.1.1.

5.3 Preprocessing

All data preprocessing was identical to that conducted for Experiments 1 and 2.

5.4 Results

To reiterate our predictions, if children both know the relation between syntactic position—in the current case, presence or absence of a preposition—and thematic relation—in the current case, instrument v. patient—and if they can furthermore deploy that knowledge, they will map the NP in *She's Ving NP with that thing* to the patient, and thus look more to the patient when they receive the *She's Ving NP with that thing* structure, and they will map the NP in *She's Ving that thing with NP* to the instrument and thus look more to the instrument when they receive the *She's Ving that thing with NP* structure.

Figure 5 plots the mean proportion of looks to instrument by STRUCTURE. As in Figure 1, the confidence intervals in Figure 5 are computed from a nonparametric bootstrap of the condition mean with 9,999 iterations. In this bootstrap, infants' mean proportion of looks to instrument across trials, weighted by FAMILIARIZATION PROPORTION, was first computed and then these mean proportions were resampled. The dashed line shows the estimated mean for the *It's a NP* structure for 19-month-olds from Experiment 1.

To assess the reliability of this pattern, we use a logistic mixed effects model with OBJECT COUNT as the dependent variable, random intercepts for infant and item, by-item random slopes for STRUCTURE, and a loss weighted by FAMILIARIZATION PROPORTION. We begin by fitting a model with fixed effects for STRUCTURE under a sum coding and then test whether this fixed effect is needed. We refer to this model as the *maximal model*.

To test the main effect of STRUCTURE, we use a log-likelihood ratio test to compare the above maximal model with one that differs only in that it lacks the fixed effect of STRUCTURE. This test suggests that the main effect is significant ($\chi^2(1)$ =4.30, p <0.05). A similar test for the by-item random slope for STRUCTURE (retaining the fixed ef-

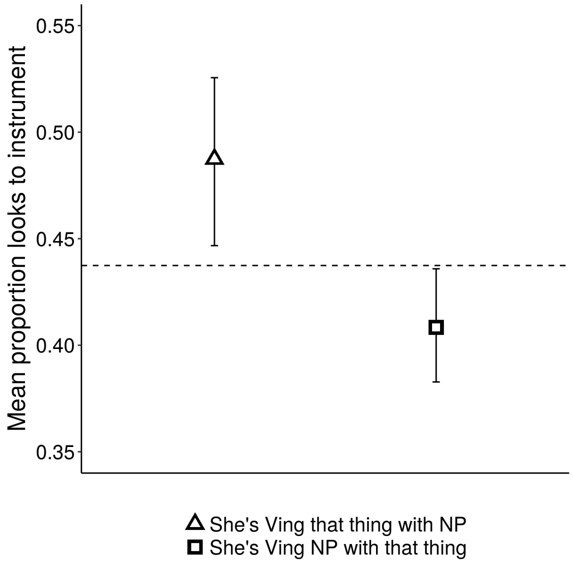

Figure 5. Mean proportion looks to instrument by STRUCTURE. Error bars show 95% confidence intervals computed from nonparametric bootstrap on participant weighted means. Dashed line shows estimated mean for the *It's a NP* structure for 19-month-olds from Experiment 1.

fect) also yields significance ($\chi^2(2)$ =21.04, p <0.05). This suggests that the random effects structure that we employ is parsimonious.

5.5 Discussion

These data support the hypothesis that 19-month-old children know the content of the preposition *with* and can use it as a cue to the thematic relation borne by its object. Moreover, it supports the view that the 19-month-olds' failure in Experiment 1 was caused not by a lack of knowledge, but by interference from the mechanics of parsing. Because these children are better able to use their knowledge of subcategorization frequency to predict upcoming structure, these predictions interfere with children's ability to display their syntactic knowledge. This finding highlights the critical nature of understanding the linguistic input as it is represented by learners. An accurate model of learning must treat the input not as it is intended, but rather as it is represented by immature learners.

6 General discussion

In a series of three experiments we have uncovered the following effects. First, we see a developmental change in children's ability to use the syntactic position of a noun phrase headed by a novel noun to learn the meaning of that noun. Prior to acquiring a substantial verb vocabulary, children are able to distinguish the interpretation of a novel NP when it is a direct object as compared to when it is a prepositional object. Upon acquiring a substantial verb vocabulary, children appear to rely more on their expectations about a verb's syntactic distribution than on the actual sentence it occurs in, blocking the inference from syntactic position to thematic relation, and consequently the inference from thematic relation to lexical meaning. Second, by 16-months children have the ability to use a preposition as a cue to the thematic relation of an NP, and hence as a cue to the meaning of a novel noun in that NP. At this age, children distinguish *with* from other prepositions semantically. Finally, 19-month-olds' difficulty in using *with* as a cue to novel noun meaning can be attenuated by placing the *with* PP after a direct object NP. We have argued that the pattern of findings in 19-month-olds reflects the contribution of their immature parsers. When infants are able to make a prediction about upcoming syntactic structure, but that prediction turns out to be wrong, they have difficulty recovering from the misparse and hence make incorrect inferences about meaning.

These findings leave open a range of interesting questions for future work. Having observed that 19-month-olds make specific syntactic predictions that interfere with bottom-up structure building, we can now ask to what extent their predictions about structure are based on specific lexical statistics? Do these children expect a direct object for all verbs, or only for those verbs that they have ample evidence for that expectation? It is well known that lexical distributional statistics play a prominent role in driving incremental sentence processing in adults (Altmann and Kamide, 2007; Arai and Keller, 2013; Boland, 2005; Thothathiri and Snedeker, 2008a,0; Trueswell and Kim, 1998, *inter alia*), and in older children (Borovsky et al., 2012; Lew-Williams and Fernald, 2007; Snedeker and Trueswell, 2004), but we do not yet know whether such lexically driven incremental processing appears at the earliest stages of language development.

Answering this question can help identify the degree to which children's early syntactic knowledge is associated with specific lexical items and the degree to which it is general across lexical items (Tomasello, 1992). If children's earliest syntactic predictions are driven by properties that are general across the verb lexicon, then that might suggest that their earliest syntactic representations are abstract and not lexically specific. Relatedly, knowing whether children's earliest syntactic structures are lexicalized in this way can also help in building a model of how syntactic information is first acquired. Is syntactic structure acquired by generalizing across features of individual words with similar properties (Lieven, 2016; Theakston et al., 2015; Tomasello, 1992), or are syntactic properties cued by phonological or semantic features (Christophe et al., 2008; Pinker, 1984) and initially acquired as abstract representations (Lidz, 2007; Snyder, 2001; Viau and Lidz, 2011)?

The pattern of effects observed in this paper highlights the importance of identifying and keeping separate the contributions of syntactic knowledge and those of parsing mechanics. The immaturity of a child's parser can lead them to assign erroneous syntactic structures and to consequently make incorrect inferences about the meanings of novel words. In turn, this conclusion emphasizes the importance of separating children's linguistic input, what they are exposed to, from their linguistic intake, how they represent their input.

A great deal of work emphasizes the role of input in language acquisition, especially links between input frequency and the acquisition of words and syntax (Hart and Risley, 1995; Hoff, 2003; Huttenlocher et al., 2002; Rowe, 2012; Weisleder and Fernald, 2013). However, our understanding of the mechanisms that link input frequency with individual differences in acquisition is still lacking. What features of experience make learning possible? Do all experiences with a word contribute equally to its acquisition, or are there special contextual features that promote learning (Cartmill et al., 2013; Medina et al., 2011; White et al., sion)?

Of course, understanding what features of experience promote learning requires understanding the sensitivities of the learner to those features. In characterizing the role of input in shaping language development, we must take care to think of the input not in terms of how it was intended by those who produced it, but rather in terms of the information that children are able to glean from that input. The degree to which an utterance is informative for some learning inference is a function of how that utterance is represented. This representation, as we have seen, can be shaped by properties of the developing parser. An important goal for the future, therefore, is to identify the various ways that children can distort their input as a function of either their current knowledge state (Lidz and Gagliardi, 2015) or their developing information processing mechanisms (Omaki and Lidz, 2015).

Acknowledgments

We are grateful to audiences at the University of Maryland, BUCLD 34, BUCLD 35, LSA 2010, CUNY 2011, and SRCD 2013 for useful comments on this work. This work was supported by NSF BCS-1551629, and White was supported by NSF DGE IGERT-0801465.

7 References

Altmann, G. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.

Altmann, G. T. and Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57(4):502–518.

Arai, M. and Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, 28(4):525–560.

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

Arunachalam, S. and Waxman, S. R. (2011). Grammatical form and semantic context in verb learning. *Language Learning and Development*, 7(3):169–184.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.

Bernal, S., Lidz, J., Millotte, S., and Christophe, A. (2007). Syntax constrains the acquisition of verb meaning. *Language Learning and Development*, 3(4):325–341.

Boland, J. E. (2005). Cognitive mechanisms and syntactic theory. *Twenty-First Century Psycholinguistics: Four cornerstones*, pages 23–42.

Borovsky, A., Elman, J. L., and Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, 112(4):417–436.

Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology*, 55(1):1.

Brown, R. (1973). *A First Language: The early stages*. Harvard University Press.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., and 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.

Childers, J. B. and Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology*, 37(6):739.

Choi, Y. and Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, 106(1):41–61.

Christophe, A., Millotte, S., Bernal, S., and Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech*, 51(1-2):61–75.

Conroy, A. M. (2008). *The Role of Verification Strategies in Semantic Ambiguity Resolution in Children and Adults*. PhD thesis, University of Maryland.

Conroy, A. M., Fults, S., Musolino, J., and Lidz, J. (2008). Surface Scope as a Default: The Effect of Time in Resolving Quantifier Scope Ambiguity.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3):547–619.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2):179–211.

Fenson, L. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual*. Paul H. Brookes Publishing Company, Baltimore.

Ferguson, B., Graf, E., and Waxman, S. R. (2014). Infants use known verbs to learn novel nouns: Evidence from 15-and 19-month-olds. *Cognition*, 131(1):139–146.

Fernald, A., Marchman, V. A., and Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2):234–248.

Fisher, C., Gertner, Y., Scott, R. M., and Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(2):143–149.

Gagliardi, A., Mease, T. M., and Lidz, J. (2016). Discontinuous development in the acquisition of filler-gap dependencies: Evidence from 15-and 20-month-olds. *Language Acquisition*, pages 1–27.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, 116(2):439.

Gertner, Y. and Fisher, C. (2012). Predicted errors in children's early sentence comprehension. *Cognition*, 124(1):85–94.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1):3–55.

Gordon, P. and Chafetz, J. (1990). Verb-based versus class-based accounts of actionality effects in children's comprehension of passives. *Cognition*, 36(3):227–254.

Hamburger, H. and Crain, S. (1984). Acquisition of cognitive compiling. *Cognition*, 17(2):85–136.

Hart, B. and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Paul H. Brookes Publishing Co., Baltimore.

He, A. X. and Lidz, J. (2017). Verb Learning in 14- and 18-Month-Old English-Learning Infants. *Language Learning and Development*.

Hirsh-Pasek, K. and Golinkoff, R. M. (1999). *The Origins of Grammar: Evidence from early language comprehension*. The MIT Press, Cambridge, MA.

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. (2005). Supercoder: A program for coding preferential looking.

Huang, Y. T. and Arnold, A. R. (2016). Word learning in linguistic context: Processing and memory effects. *Cognition*, 156:71–87.

Huang, Y. T., Leech, K., and Rowe, M. L. (2017). Exploring socioeconomic differences in syntactic development through the lens of real-time processing. *Cognition*, 159:61–75.

Huang, Y. T., Zheng, X., Meng, X., and Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, 69(4):589–606.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., and Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, 45(3):337–374.

Leddon, E. and Lidz, J. (2006). Reconstruction Effects in Child Language. In *Proceedings of the 30th annual Boston University Conference on Language Development*, Somerville, MA. Cascadilla Press.

Lew-Williams, C. and Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3):193–198.

Lidz, J. (2007). The abstract nature of syntactic representations: Consequences for a theory of learning. In Hoff, E. and Shatz, M., editors, *Blackwell Handbook of Language Development*, pages 277–303. Blackwell Publishing Ltd., Oxford.

Lidz, J., Bunger, A., Leddon, E., Baier, R., and Waxman, S. R. (2010). When one cue is better than two: lexical vs. syntactic cues to verb learning.

Lidz, J. and Conroy, A. M. (2007). The Role of Working Memory in Explaining Isomorphism. Barcelona, Spain.

Lidz, J. and Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annual Review of Linguistics*, 1(1):333–353.

Lieven, E. (2016). Usage-based approaches to language development: Where do we go from here? *Language and Cognition*, 8(03):346–368.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4):676.

MacWhinney, B. (2014a). *The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs*. Psychology Press.

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

Marchman, V. A. and Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, 11(3):F9–F16.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94:305–315.

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

Musolino, J., Crain, S., and Thornton, R. (2000). Navigating negative quantificational space. *Linguistics*, 38(1):1–32.

Musolino, J. and Lidz, J. (2003). The scope of isomorphism: Turning adults into children. *Language Acquisition*, 11(4):277–291.

Musolino, J. and Lidz, J. (2006). Why children aren't universally successful with quantification. *Linguistics*, 44(4):817–852.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17(2):357–374.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14(1):11–28.

Omaki, A. (2010). *Commitment and Flexibility in the Developing Parser*. PhD thesis, University of Maryland.

Omaki, A., Davidson White, I., Goro, T., Lidz, J., and Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in English and Japanese wh-questions. *Language Learning and Development*, 10(3):206–233.

Omaki, A. and Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition*, 22(2):158–192.

Pearl, L. and Sprouse, J. (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, 20(1):23–68.

Pinker, S. (1984). *Language Learnability and Language Development*. Harvard University Press.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press, Cambridge, MA.

Pozzan, L., Gleitman, L. R., and Trueswell, J. C. (2016). Semantic ambiguity and syntactic bootstrapping: the case of conjoined-subject intransitive sentences. *Language Learning and Development*, 12(1):14–41.

Pozzan, L. and Trueswell, J. C. (2015). Revise and resubmit: How real-time parsing limitations influence grammar acquisition. *Cognitive Psychology*, 80:73–108.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of childâĂŘdirected speech in vocabulary development. *Child Development*, 83(5):1762–1774.

Snedeker, J. and Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49(3):238–299.

Snyder, W. (2001). On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language*, pages 324–342.

Spelke, E. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, 8(4):553–560.

Spelke, E. and Newport, E. L. (1998). Nativism, empiricism, and the development of knowledge. In Damon, W. and Lerner, R. M., editors, *Handbook of Child Psychology: Theoretical models of human development*, volume 1, pages 275–340. John Wiley & Sons Inc, Hoboken, NJ, US, 5 edition.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48(3):542–562.

Sussman, R. S. and Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, 18(2):143–163.

Theakston, A. L., Ibbotson, P., Freudenthal, D., Lieven, E. V., and Tomasello, M. (2015). Productivity of noun slots in verb frames. *Cognitive Science*, 39(6):1369–1395.

Thothathiri, M. and Snedeker, J. (2008a). Give and take: Syntactic priming during spoken language comprehension. *Cognition*, 108(1):51–68.

Thothathiri, M. and Snedeker, J. (2008b). Syntactic priming during language comprehension in three-and four-year-old children. *Journal of Memory and Language*, 58(2):188–213.

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

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74(3):209–253.

Trueswell, J. C., Kaufman, D., Hafri, A., and Lidz, J. (2012). Development of parsing abilities interacts with grammar learning: Evidence from Tagalog and Kannada. In *Proceedings of the 36th annual Boston University Conference on Language Development*, Somerville, MA. Cascadilla Press.

Trueswell, J. C. and Kim, A. E. (1998). How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language*, 39(1):102–123.

Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2):89–134.

Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3):528.

Viau, J. and Lidz, J. (2011). Selective learning in the acquisition of Kannada ditransitives. *Language*, 87(4):679–714.

Viau, J., Lidz, J., and Musolino, J. (2010). Priming of abstract logical representations in 4-year-olds. *Language Acquisition*, 17(1-2):26–50.

Waxman, S. R. and Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, 43(3):217–242.

Waxman, S. R. and Lidz, J. L. (2006). Early Word Learning. In Kuhn, D., Siegler, R. S., Damon, W., and Lerner, R. M., editors, *Handbook of Child Psychology: Cognition, perception, and language*, volume 2, pages 299–335. John Wiley & Sons Inc, Hoboken, NJ, US, 6 edition.

Waxman, S. R., Lidz, J. L., Braun, I. E., and Lavin, T. (2009). Twenty four-month-old infantsâĂŹ interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, 59(1):67–95.

Waxman, S. R. and Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology*, 29(3):257–302.

Weisleder, A. and Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11):2143–2152.

White, A. S., Hacquard, V., Resnik, P., and Lidz, J. (under revision). The contextual modulation of semantic information.

White, A. S., Hirzel, M., and Lidz, J. (in prep). Lexicalization in the developing parser.

Yuan, S. and Fisher, C. (2009). "Really? She Blicked the Baby?" Two-Year-Olds Learn Combinatorial Facts About Verbs by Listening. *Psychological Science*, 20(5):619–626.

Yuan, S., Fisher, C., Kandhadai, P., and Fernald, A. (2011). You can stipe the pig and nerk the fork: Learning to use verbs to predict nouns. In *Proceedings of the 35th annual Boston University Conference on Language Development*, pages 665–677, Somerville, MA. Cascadilla Press.

Yuan, S., Fisher, C., and Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, 83(4):1382–1399.