Cognitive Psychology 60 (2010) 190–217

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases

Anne Fernald *, Kirsten Thorpe, Virginia A. Marchman

Department of Psychology, Stanford University, Stanford, CA 94305, United States

A R T I C L E I N F O

Article history:
Accepted 23 December 2009

Keywords:
Language development
Lexical development
Sentence processing
Eye movements
Online comprehension
Processing speed
Adjectives
Prenominal adjectives
Adjective–noun integration

A B S T R A C T

Two experiments investigated the development of fluency in interpreting adjective–noun phrases in 30- and 36-month-old English-learning children. Using online processing measures, children's gaze patterns were monitored as they heard the familiar adjective–noun phrases (e.g. *blue car*) in visual contexts where the adjective was either informative (e.g. blue car paired with red car or red house) or uninformative (e.g. blue car paired with blue house). Thirty-six-month-olds processed adjective–noun phrases incrementally as adults do, orienting more quickly to the target picture on informative-adjective trials than on control trials. Thirty-month-olds did not make incremental use of informative adjectives, and experienced disruption on trials when the two potential referents were identical in kind. In the younger children, difficulty in integrating prenominal adjectives with the subsequent noun was associated with slower processing speed across conditions. These findings provide evidence that skill in putting color word knowledge to use in real-time language processing emerges gradually over the third year.

1. Introduction

In understanding spoken language, adults effortlessly integrate linguistic knowledge with information from the visual scene, apprehending meaning incrementally as speech is heard (e.g., Henderson & Ferreira, 2004; Trueswell & Tanenhaus, 2005). Given the virtuosity of adult performance, how does the young language learner progress toward such efficiency? For many years, studies of early receptive language development relied primarily on parent-report measures and offline assessments of

* Corresponding author.
E-mail address: afernald@stanford.edu (A. Fernald).

0010-0285/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2009.12.002

children's comprehension, typically characterizing developmental change in terms of all-or-none measures of lexical and grammatical knowledge. Such studies showed, for example, that young children produce more nouns than adjectives in their early vocabularies (e.g., Nelson, 1973), and that 4-year-olds can interpret novel words correctly as adjectives rather than nouns, while 2-year-olds cannot (e.g., Hall, Waxman, & Hurwitz, 1993; Taylor & Gelman, 1988). With the refinement of online measures of understanding suitable for use with young children, such developmental change can now be delineated in finer detail. Research using eye-tracking methods with very young language learners reveals gradual, continuous development in children's processing efficiency in some domains, and discontinuity in others. For example, the speed of infants' recognition of familiar object names increases steadily over the second and third years of life (e.g., Fernald, Perfors, & Marchman, 2006; Zangl & Fernald, 2007). However, in a more challenging online processing task in which children could make use of both lexical and referential information, 4- and 5-year-olds rely on interpretative strategies that differ from those of adults (Snedeker & Trueswell, 2004). Here we use measures of children's speed and accuracy in speech processing to investigate when young language learners begin to show adult-like proficiency in interpreting adjective–noun phrases in real time.

How listeners interpret adjectives and nouns in combination illustrates the challenges of understanding spoken language as quickly as it is heard. Adjectives often depend on the nouns they modify for the full interpretation of their meaning—consider the different implications of the same adjective used in different contexts, e.g., *good boy*, *good meal* and *good idea*. Even a less abstract property term like *tall* can only be interpreted in relation to the set of things denoted by the noun it modifies, e.g. when used to describe a child versus a tree (Barner & Snedeker, 2008). Although all adjectives depend on the nouns they modify to a degree, different semantic subclasses of adjectives vary in how they combine with nouns and how much context is necessary for interpretation (e.g., Partee, 1995; Vendler, 1963). Subsective adjectives, including scalar terms such as *big* or *hot* and evaluative terms such as *good*, rely to a greater extent on context provided by the nouns they modify for their interpretation. In contrast, intersective adjectives such as words for color and material denote relatively independent properties, combining with nouns via the intersection of sets, e.g., a *plastic cup* is a member of the set of things that are plastic as well as the set of things that are cups. But even adjectives such as color words, considered to have relatively stable independent meanings, are still to some extent interpreted with respect to the nouns they describe. For example, *red* indicates different hues depending on whether it modifies *apple* or *sunset*. And understanding even a simple noun phrase such as *blue car* involves determining both what property *blue* refers to and what type of object *car* refers to, and then combining the meanings of these two words to identify a single object that is both blue and a car. The research presented here focuses on the time-course of children's interpretation and integration of prenominal color words with the nouns they modify.

1.1. Why are adjectives difficult for young children?

Adjectives comprise a diverse class of words that are challenging for young language learners for several reasons. Gentner (1982) argues that because of their "relational relativity", adjectives, prepositions, and other relational terms are more cross-linguistically variable in how they map from concept to word than are concrete and proper nouns, which refer to things that are perceptually coherent and easily individuated. The conceptual complexity of adjectives relative to concrete nouns could account in part for the fact that children are typically slower to learn property terms than object names (Gentner & Boroditsky, 2001). Adjectives may also be difficult for young learners because they can occur in potentially ambiguous syntactic frames. The novel word in the sentence *That's zavvy* could be an adjective, a mass noun or a proper noun. If a child hears this word in prenominal position following an article, as in *That's a zavvy car*, it is identifiable as an adjective as soon as the next word is heard, but a reasonable first bet is that a new word following *a* or *the* will be a noun rather than an adjective. Given this temporary ambiguity, a child could initially misinterpret a novel prenominal adjective as an object name rather than a property term, potentially disrupting processing of adjective–noun phrases (Thorpe & Fernald, 2006). Property terms may also be challenging for young learners because adjectives occur less frequently in speech to children than do nouns and verbs (Sandhofer, Smith, & Luo, 2000) and children have less experience with these kinds of words. This constellation of factors helps to explain why 2- and 3-year-olds with several hundred words in their expressive vocabularies still have difficulty in language tasks involving adjectives.

Studies of early adjective acquisition have focused extensively on how children learn the meanings of property terms, exploring which kinds are easiest to understand first (e.g., Bornstein, 1985; Clark, 1972; Johnson, 1977; Pitchford & Mullen, 2005; Sandhofer & Smith, 1999), and when and how children correctly interpret novel adjectives as property terms rather than as object labels (e.g., Booth & Waxman, 2009; Gelman & Markman, 1984; Hall & Graham, 1999). Other studies have also looked at how young children interpret novel adjectives in interaction with the nouns they modify. For example, learning a novel property term is easier for 2- to 3-year-olds when the object it relates to is labeled with a familiar category name, e.g., *stoof horse* vs. *stoof one/thing* (Mintz, 2005; Mintz & Gleitman, 2002). In addition, when the kind of object described is held constant, a novel property term is easier to map; that is, 2- and 3-year-olds are more successful in identifying the appropriate property of a *blickish horse* when the novel word *blickish* was used previously in reference to other horses, rather than other kinds of animals (e.g., Klibanoff & Waxman, 2000; Waxman & Klibanoff, 2000). Such findings are interpreted as evidence that knowledge of the object described can provide support for a child trying to interpret a novel adjective, by helping the child focus on the relevant property more easily. However, the processes involved in learning novel adjectives and mapping a single word to a relevant property may be quite different from the processes involved in rapidly interpreting familiar adjective–noun phrases in continuous speech.

Many earlier studies with very young children have focused on when the first adjectives are learned and how novel adjectives are interpreted. Here we focus on a different question: Given that by age 30 months most English-learning children can correctly produce at least a few property terms, including color words like *red* and *blue*, how efficiently can they use this knowledge receptively when encountering these words in continuous speech? That is, how rapidly and reliably can children interpret the color word in the noun phrase *blue car* in relation to the visual scene? We know that 24-month-olds interpret familiar object names incrementally, making use of phonetic information as soon as it becomes available without waiting to hear the entire word (Fernald, Swingley, & Pinto, 2001; Swingley, Pinto, & Fernald, 1999). The present study examines incremental processing at a different level, asking when young children develop the ability to interpret a prenominal adjective as it is heard and to integrate it efficiently with the subsequent noun, in order to rapidly identify a referent.

1.2. Integrating adjectives and nouns

While developmental studies have asked at what age and under what conditions children can correctly identify the property indicated by an adjective, the adult literature on adjective use has explored a wider range of questions, including how property concepts are combined with object concepts in interpreting an adjective in relation to a noun (e.g., Smith, Osherson, Rips, & Keane, 1988; Kamp & Partee, 1995). Several researchers have argued that in adjective–noun phrases, interpretation of the modifier is constrained by the interpretation of the head noun (e.g., Murphy, 1988, 1990). To make the point that an adjective is linked to the noun it modifies, Medin and Shoben (1988) asked adults to rate the similarity of phrases in which the same adjectives were combined with different nouns. As an example, adults rated the phrase *grey clouds* as more similar to *black clouds* than to *white clouds*, although *grey hair* was rated as more similar to *white hair* than to *black hair*. Thus, even adjectives referring to relatively independent properties like color are interpreted variably in relation to particular nouns. A central question in such studies with adults is how a reader or listener integrates the meanings of different kinds of words that are presented sequentially, such as *red* and *ball* in relation to a single referent, a ball that is red (e.g., Smith & Osherson, 1984).

In developmental studies, children's difficulty in interpreting adjectives has not typically been framed in terms of processing challenges at the level of integration. One exception is research by Ninio (2004), who raised the possibility that it is just this need for integration that causes difficulty for young language learners in processing adjectives in conjunction with nouns. Ninio assessed children's skill in interpreting adjective–noun phrases in a picture-pointing task. While children were fairly accurate in identifying a *black shoe* in a two-choice task where the relevant property (black or white) was the only difference between items of the same kind, it was harder for children to choose the *black* *shoe* from a set of four items that crossed the relevant property with the object kind (e.g., black and white shoes and socks), a task that required finding the intersection of meanings for both the adjective and noun. Ninio proposed that interpreting adjective–noun combinations may be challenging for children because it requires a two-step process: first, identification of the object category labeled by the noun, and second, subcategorization within the object category, using the adjective to identify the appropriate referent. According to this hypothesis, the child must first interpret the object name *shoe* and then interpret the property term to find the *black* one. If this two-step process is too challenging, a child might focus only on identifying a suitable referent of *shoe*, and omit the second step.

In Ninio's (2004) study, children were tested in Hebrew, a language in which adjectives are positioned postnominally rather than in the prenominal position characteristic of English. Therefore the sequential steps of interpretation proposed by Ninio are aligned with the order in which words are actually heard in Hebrew noun phrases that contain adjectives. However, for the English-speaking child, this model could involve an additional processing challenge. If noun phrases containing adjectives are interpreted in two steps—the noun first and the adjective second—then the English-speaking child would have to wait until hearing the noun before beginning to interpret the phrase. This would entail holding the prenominal adjective in working memory while performing the first step of noun identification, then retrieving the adjective to identify the appropriate referent. This demand on memory might make interpreting adjective–noun phrases relatively more difficult for young English learners than for Hebrew learners.

As a first step in exploring adjective–noun integration for English-speaking children, Thorpe, Baumgartner, and Fernald (2006) conducted a replication of Ninio's (2004) study using the same offline testing methods. To investigate developmental differences, children were tested in three age groups: 1.5-, 2.5-, and 3.5-year-olds. Although Ninio did not group children by age, in her study the children spanned this same range. Similar to Hebrew-learners, English-speaking children responded less accurately overall in a four-choice condition in which both words in an adjective–noun phrase were essential for identifying the target ($M = 60\%$ correct), as compared to a condition in which the adjective alone was sufficient ($M = 72\%$ correct). However, children's success in integrating adjective–noun combinations varied dramatically by age. The youngest children had difficulty with adjectives regardless of condition, while children older than 3 years were close to ceiling on this task. Only the 2.5-year-olds showed evidence of integration difficulty above and beyond adjective-only comprehension, and it was only at this age that English-learning children showed a noun-biased error pattern similar to that observed in Hebrew learners. This pattern of errors was interpreted by Ninio as evidence that children process noun phrases containing an adjective by first determining the noun category, but then failing to complete the process of identifying the correct picture based on the adjective. However, because the offline task used in this research only captured children's final interpretation of the phrase, we cannot know if their error patterns were the result of processing the noun *first* and then leaving off the adjective, or whether the errors resulted from being more familiar with or remembering the nouns better than the adjectives they heard.

1.3. Interpreting adjectives in real time

If an adjective is interpreted in relation to the noun it modifies, then the typical positioning of the adjective *before* the noun in English would not seem to be an optimal arrangement. Indeed, Clifton and Ferreira (1989) have proposed that given the dependent nature of property terms, the prenominal adjective in a phrase such as *red face* can only be understood after the noun is encountered. Thus, under certain circumstances a listener would have to delay interpreting an adjective until after the noun to fully apprehend the intended meaning of the phrase. However, depending on the perceptual accessibility of the property denoted by the particular adjective, and also on constraints from the discourse or visual context, there are differences in how early the adjective can be used to narrow down or identify the referent of a phrase. Studies using eye-tracking techniques have shown that English-speaking adults in certain situations do begin to interpret a prenominal adjective as soon as it is heard, even before they have heard the subsequent noun (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Sedivy, Tanenhaus, Chambers, & Carlson, 1999). When participants heard instructions such as *Touch the plain blue square*, their responses differed depending on what objects other than the target were present. With several plain blue objects in view, they waited to make their choice until they heard the word *square*; however when there was only a single object present that was both plain *and* blue, listeners looked at the referent within 300 ms of hearing *blue*. Eberhard et al. showed that when a prenominal adjective was present that enabled the listener to uniquely identify one referent within a larger set, adults used this information in online comprehension before the object name was spoken.

Can young children also interpret adjectives incrementally? Eye-tracking studies with infants have shown that efficiency in language understanding increases substantially over the second year (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998), and that children's speed and accuracy in online processing are related to traditional offline measures of lexical and grammatical development (Fernald et al., 2006; Hurtado, Marchman, & Fernald, 2007; Zangl, Klarman, Thal, Fernald, & Bates, 2005). By the age of two, children can rapidly identify familiar words before the target word has been fully spoken (Fernald et al., 2001; Swingley et al., 1999). Moreover, by this age English-learning children can also process a determiner-noun phrase incrementally, not just the noun itself, reflecting some awareness that the article *the* is typically followed by an object name. When 18-month-olds heard sentences in which this common co-occurrence pattern was violated by substituting a nonce word for the familiar article (e.g., *Where's po ball?* vs. *Where's the ball?*), their ability to identify the target word was disrupted (Zangl & Fernald, 2007). If 2-year-olds have developed expectations that determiners are typically followed by object names, what happens when the article *the* is followed directly by an adjective instead of a noun? Will this also disrupt processing of the target object name that follows?

Thorpe and Fernald (2006) addressed one aspect of this question by presenting 2-year-olds with a familiar but uninformative adjective followed by a familiar object name (e.g., *Where's the nice BALL*?). They found that 24-month-olds could listen through such prenominal adjectives, responding as rapidly to the subsequent target noun as if no adjective were present. But if the adjective was novel and/or received primary stress (e.g., *Where's the ZAVVY ball*?), children were more likely to "false alarm" by shifting prematurely away from the target picture. That is, they responded as if the novel word following the article *the* was a potential object name, disrupting processing of the familiar target noun that followed. These findings suggest that 2-year-olds use a combination of syntactic, lexical, and prosodic cues to anticipate what is coming next in the speech stream. Thorpe and Fernald found that young children ignored an uninformative adjective preceding a familiar noun, showing that they can efficiently delay responding until the head noun in an adjective–noun phrase is heard; however this study did not investigate responses to adjectives that *are* informative and thus could potentially help in establishing reference. Do children wait for the noun in that case as well, or do they respond more quickly by making use of the adjective as they hear it? When an adjective is potentially useful for distinguishing which object is the appropriate referent, do young listeners, like adults, pay attention to the adjective and use it incrementally to narrow down the set of referents?

The goal of the present study was to explore the early development of children's ability to interpret familiar adjectives in online sentence comprehension. In two experiments we tested young children's identification of pictures labeled with familiar adjective–noun phrases, using an eye-tracking procedure that allows us to monitor the time course of children's interpretations with fine-grained temporal resolution. In both experiments, children were presented with pictures of objects labeled with the familiar color words *red* and *blue*. These adjectives were chosen for two reasons: First, parents report that by the age of 30 months, 85% of English-speaking children produce the words *red* and *blue*, according to norms for the MacArthur-Bates Communicative Development Inventory (Dale & Fenson, 1996). And second, because color terms refer to perceptually accessible properties, they are more likely than other types of adjectives to be interpretable without depending as crucially on information from the noun. For example, on hearing *Where's the blue*..., a child could potentially identify the correct picture without needing to hear the object named, as long as color was sufficient to distinguish between the alternatives.

In Experiment 1, participants were 30 months old, an age when children can typically produce a number of descriptive words and tend to respond correctly to *red* and *blue* in offline tests of color word comprehension (e.g., Bartlett, 1978; Johnson, 1977), although they still have difficulty learning and interpreting adjectives (e.g. Sandhofer & Smith, 1999). The participants in Experiment 2 were

6 months older. By 36 months many children have begun to overcome the challenges of producing and understanding common adjective–noun combinations, as shown in several studies using offline measures (e.g. Mintz, 2005). Given that children at both ages can correctly use words like *red* and *blue* in their spontaneous speech, how efficient are they in interpreting color words in prenominal position to rapidly identify a referent? Comparing the performance of children at 30 and 36 months enabled us to investigate developmental changes in processing efficiency as children become more experienced in comprehending adjective–noun constructions.

2. Experiment 1

In Experiment 1, 30-month-old English-learning children were tested in the looking-while-listening procedure (e.g., Fernald et al., 1998, 2006) in which they looked at pairs of pictures while listening to speech naming one of the pictures. Frame-by-frame analyses of eye movements yielded a record of the time course of looking behavior as each stimulus sentence unfolded (see Fernald, Zangl, Portillo, & Marchman, 2008). From this record we derived measures of speed and accuracy in response to the adjective–noun phrase on every trial. To determine whether children could make use of potentially informative prenominal adjectives in establishing reference, we varied the visual array as they heard the same stimulus sentences (e.g., *Can you find the blue car*?) in three conditions:

1. *Same Color/Different Objects (control)*: In this condition, children saw, for example, a blue car and a blue house while hearing *Can you find the blue car*? Here, only the noun identified the correct referent because the adjective was not informative. To avoid confusion among the names of the three conditions, we will refer to the control condition as the *Blue Car/Blue House* condition. Note, however, that two objects and two colors were fully crossed as stimuli in these experiments, so this designation is used to refer to all four adjective noun pairings in which the objects differed in kind but the color was the same.
2. *Different Colors/Different Objects*: In this condition, children saw, for example, a blue car and a red house while hearing *Can you find the blue car*? The adjective was sufficient as a cue to identify the correct referent, although not necessary since the noun was also sufficient. This will be referred to as the *Blue Car/Red House* condition.
3. *Different Colors/Same Object*: Here, children saw, for example, two cars, one red and one blue, while hearing *Can you find the blue car*? The adjective was the only cue to the correct referent; hence, the adjective was both necessary and sufficient. This will be referred to as the *Blue Car/Red Car* condition.

In the Blue Car/Blue House condition, the prediction was clear: since both objects were of the same color, the adjective could not possibly be used to identify the correct referent. Thus this condition served as a control, enabling us to determine how rapidly children could interpret the noun phrase when they had to rely on the noun itself to identify the target object. In the second and third conditions, both *informative-adjective* conditions, the stimulus objects differed in color, and thus children could potentially make use of the prenominal adjective to identify the target in advance of hearing the object label. If children can process adjective–noun phrases as efficiently as adults, they should look to the target faster on both Blue Car/Red Car trials and Blue Car/Red House trials than on control trials, because in both cases the adjective uniquely identifies the target. If they are also able to integrate the adjective efficiently with the following noun, they should continue looking at the picture that is both blue *and* a car as the noun is subsequently heard.[1]

However, given findings suggesting that children may not perform this task with optimal efficiency, three alternate predictions should be considered. First is Ninio's (2004) proposal that young children process adjective–noun phrases in two steps, first the noun and then the adjective, and have difficulty integrating the two. If the head noun is always processed before the adjective, then English-

[1] In other studies using this procedure, children as young as 24 months typically maintain fixation on the target picture for at least 2 s after correctly matching the target word with its referent (e.g., Fernald et al., 1998; Zangl & Fernald, 2007).

learning children should wait to respond until after hearing the noun. taking just as long to orient to the target in the two informative-adjective conditions as in the control condition.

Second, children in this age range might find it harder to interpret *blue car* when choosing between a blue car and a red house than when choosing between two cars that differ in color, based on findings that they have difficulty interpreting novel property terms used to distinguish objects labeled by different nouns (e.g., Klibanoff & Waxman, 2000; Waxman & Markow, 1998). This outcome could occur if the property labeled by a novel adjective becomes relatively less salient when objects also differ in kind, i.e., if children attend less to a property difference when there is also a type difference. If comprehension of familiar adjectives is parallel in difficulty to the extension of new adjectives, then young children might also be slower to identify the target on Blue Car/Red House trials than on Blue Car/Red Car trials because the two objects differ in kind as well as in color.

Third, if young children do have difficulty integrating adjective and noun meanings, then even if they could rely solely on the color word in the informative-adjective conditions, there might still be evidence of disruption in the Blue Car/Red Car condition as the noun is heard. In this condition, unlike the other two, the noun alone does not distinguish the target, since both pictures are of the same object type. If children wait to integrate the adjective until after hearing the noun, they might tend to look back and forth between the two cars as they hear *car* before integrating the adjective. Such momentary difficulty with integration would not necessarily lead to a failure to locate the target picture eventually, since in all three conditions a child could pick out the referent without having to combine the adjective and noun. But by tracking the time-course of interpretation, we should be able detect a temporary problem with integration in the form of disruption as the noun is heard in the Blue Car/Red Car condition, when the noun alone could refer to either picture.

2.1. Method

2.1.1. Participants

Participants were 32 30-month-old children ($M = 30.1$ mos; range = 28–32); 19 were female. All were reported to be typically developing children from families where English was the dominant language. One additional participant was excluded due to inattentiveness during testing (i.e., failure to look at either of the pictures on more than half of the trials). Parents also reported that all the 30-month-olds were producing the words *red* and *blue* in their spontaneous speech.

2.1.2. Visual stimuli

The visual stimuli were digitized full-color photographs of objects named by the target words. Pictures were presented in pairs according to condition and matched approximately for size and brightness. Each of the four stimulus types (blue car, red car, blue house, red house) served once as the target object and once as distracter in each of the three conditions. Side of presentation of the target picture was counterbalanced across trials.

2.1.3. Speech stimuli

As shown in Table 1, the same speech stimuli were used in all three conditions, with different configurations of visual stimuli. A female native speaker of American English first recorded multiple tokens of each sentence. Exemplars chosen for the final stimulus set were selected based on acoustic measurements of the carrier phrase, the prenominal adjective, and the noun. Adjectives in the final stimulus set were comparable in duration ($M = 273$ ms, $SD = 35$ ms), as were target nouns ($M = 594$ ms, $SD = 67$ ms) and carrier phrases ($M = 958$ ms, $SD = 188$ ms). In a few instances, minor editing of the stimuli was required to achieve a satisfactory match, using Peak™ LE 2.62 waveform editing software. One of four attentional sentences (e.g., *Look at that!*) followed each target sentence to maintain children's interest. Sixteen filler trials naming additional familiar pictures in carrier phrases without adjectives were interspersed among the 12 critical test trials.

2.1.4. Apparatus and procedure

Caregiver and child were escorted into a playroom where the study was explained. The caregiver signed a consent form and completed a checklist indicating whether the child produced the words

Table 1
Overview of visual and auditory stimuli in Experiments 1 and 2.

Condition	Picture pair	Speech stimuli
Same Color/Different Objects (control)	Blue car – blue house	*Can you find the blue car*? *Can you find the blue house*?
	Red car – red house	*Can you find the red car*? *Can you find the red house*?
Different Colors/Different Objects	Blue car – red house	*Can you find the blue car*? *Can you find the red house*?
	Blue house – red car	*Can you find the blue house*? *Can you find the red car*?
Different Colors/Same Object	Blue car – red car	*Can you find the blue car*? *Can you find the red car*?
	Blue house – red house	*Can you find the blue house*? *Can you find the red house*?

red and *blue*. Caregiver and child then proceeded to the testing room, which contained a rear projection screen (1.5 × 2.1 m). The child sat on the caregiver's lap approximately 1.2 m from the screen. Cloth-covered panels (1 × 2 m) were positioned on both sides to reduce distraction. During testing the caregiver wore sunglasses with lenses covered in black tape to block the view of the pictures and prevent the caregiver from influencing the child's responses.

When the child was attentive, the trials began. During testing, picture pairs were presented on the screen in 36 × 50 cm rectangles, 64 cm apart, using an NEC™ WT600 rear projection system. Speech stimuli were played over a central loudspeaker below the screen. The child's gaze patterns were recorded with a Sony™ SSC-M383 video camera positioned at eye level. Order and timing of picture and sentence presentation were controlled using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). On each trial, the pictures appeared for 2 s before the sound stimulus and remained on the screen for 3 s from the onset of the speech stimulus. Trials were separated by a 1-s pause. Trial types within each of two block were presented in a quasi-random order such that trials of the same type did not occur consecutively and the target picture did not appear on the same side more than three times in a row. Children were randomly assigned to one of four different orders.

2.1.5. Coding

Participants' gaze patterns were video-recorded during testing. A digital time-code accurate to 33 ms was recorded onto the video, with visual markers indicating onset and offset of the pictures and speech stimulus on each trial. Each video record was pre-screened to eliminate unusable trials, e.g. if a child looked away during an entire trial, or if the parent or child was talking during sentence presentation. Highly trained coders, blind to trial type, target word and side of target picture, coded all usable trials frame-by-frame. Using custom software, coders indicated on each frame whether the participant's eyes were oriented to the left picture, to the right picture, between the pictures, or away from both. Eye movement data were then temporally aligned with the onset of the speech signal on each trial. This yielded a high-resolution record of the time course of each child's gaze patterns to the target and distracter pictures as the speech stimulus unfolded.

Inter- and intra-observer reliability checks were conducted for all coders. For 25% of the subjects, eight trials with two or more gaze shifts were coded independently by two coders. Two measures of inter-observer reliability were assessed. The first was the proportion of frames (33-ms units) on each trial on which two coders agreed. In this case, agreement was 99%. However, because this analysis included many frames on which the child was maintaining fixation on one picture, we also calculated a more stringent test of reliability. This second measure focused only on shifts in gaze, ignoring steady-state fixations in each trial on which agreement was inevitably high. By this more conservative measure, coders agreed within one frame on 96% of all shifts.

2.1.6. Measures of efficiency in online speech processing

The looking-while-listening procedure enables us to measure children's interpretation of spoken language from multiple vantage points, examining response speed as well as patterns of correct and incorrect fixation in relation to particular regions of interest in the unfolding sentence (Fernald et al., 2008). Because children cannot know in advance which picture will be labeled, about half the time they will by chance be looking at the distracter picture at target-word onset (distracter-initial trials), and half the time they will already be looking at the target picture (target-initial trials). On distracter-initial trials, the correct response is to shift to the target picture as soon as sufficient phonetic information is available to identify the correct referent, while on target-initial trials, the correct response is to continue looking at the target picture without shifting away. Thus, a child with perfect accuracy would shift quickly to the target picture on 100% of the distracter-initial trials, and would never shift away to the distracter picture on target-initial trials.

By measuring children's looking patterns across the entire sentence, we capture not only their initial responses, but also their patterns of fixation over time, examining gaze patterns that are time-locked to particular words and phrases as they are heard. The looking-time data in Experiments 1 and 2 are presented graphically in three ways: First, we present plots of overall *accuracy*, showing mean proportion of looking time to target across the unfolding sentence. Second, we present *onset-contingent plots* that show fixation patterns on distracter- and target-initial trials grouped separately, contingent on where the child was looking at the onset of the adjective–noun phrase. From these, we assess two further measures of speech processing efficiency, *reaction time* and *off-target responses*. Reaction time is based on distracter-initial trials, to measure response speed on correct shifts from the distracter to the target picture. Off-target responding is based on children's mean proportion of looking time to the distracter on target-initial trials, to capture *incorrect* shifting away from the target to the distracter picture as evidence of disruption or confusion. Third, we present a *difference-score analysis* to capture the *relative* efficiency of children's responses on informative-adjective trials as compared to control trials at 30 and 36 months.

2.1.7. Accuracy

The overall accuracy of children's responses on each trial is operationalized as proportion of looking time to the named target picture. Combining the data from both target- and distracter-initial trials, this measure represents the total amount of time the infant spent fixating the target picture as a percentage of total time fixating either picture during the 300–2000 ms window from adjective onset. Accuracy was assessed over two regions of interest in the speech stimulus: First, the Adjective/Noun window (300–1167 ms after the onset of the adjective) corresponded to the mean duration of the adjective–noun sequence in each stimulus sentence, taking into account the 300 ms required to program and launch a saccade in response to verbal stimuli. Second, the Post-Noun Phrase (Post-NP) window captured responses during an 800 ms window after the adjective–noun phrase had been heard (i.e., 1200–2000 ms from adjective onset).

2.1.8. Reaction time

To investigate whether children respond more rapidly using information from the prenominal adjective to identify the target noun, a mean reaction time (RT) to shift from distracter to target picture was calculated for each child in each condition. Response latencies less than 300 ms were excluded from the analyses. Determination of this cutoff in previous studies has been based on estimates of the minimum time it takes listeners to process adequate phonetic information from the relevant word and then launch an eye movement to the correct referent (see Allopenna, Magnuson, & Tanenhans, 1998; Canfield, Smith, Brezsnyak, & Snow, 1997; Fischer, 1992; Haith, Wentworth, & Canfield, 1993). In studies using eye-tracking methods with participants of different ages, this estimate has varied from 200 to 400 ms (e.g., Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Fernald et al., 2001; Swingley & Aslin, 2000) with shorter intervals typically used with adults and older children (e.g., Tanenhaus, Magnuson, Dahan, & Chambers, 2000). This measurement window extended to the end of the trial at 2500 ms. Note that this was longer than the 1800 ms window used in previous studies investigating children's responses to nouns alone (e.g., Fernald et al., 1998, 2006), to allow the additional time they might need to integrate an adjective with a noun after hearing the entire noun phrase.

2.1.9. Off-target responses

In addition to calculating mean RT based on correct first shifts to the target on distracter-initial trials, we also examined incorrect responding on target-initial trials, in the form of looks away from the target to the distracter. To assess the extent to which each child false-alarmed by moving incorrectly to the distracter, we calculated the proportion of looking time to the distracter on target-initial trials, during both Adjective/Noun and Post-NP windows. In both coding windows, we would expect that children who were responding efficiently would only rarely shift away from the correct picture on target-initial trials; thus an increase in proportion of time on the distracter in either of these windows was interpreted as evidence of disruption or confusion.

Because the referent was always named with a familiar adjective and noun, children might wait until after hearing the entire noun phrase to identify the target. However, if they can use the adjective to facilitate identification of the referent, children should look to the target picture as soon as they have sufficient information, resulting in faster reaction times on shifts from distracter to target object in the two informative-adjective conditions, as compared to the control condition in which the adjective was not informative. Additionally, if children at this age can efficiently interpret a noun phrase with a prenominal adjective, then incorrect shifting away from target to distracter should occur infrequently, and the proportion of off-target looking should stay low in all three conditions.

2.2. Results of Experiment 1

Fig. 1A provides an overview of the time course of 30-month-old children's looking patterns to the target picture in three conditions. This graph tracks changes over time in the mean proportion of trials on which children fixated the correct referent, averaged over participants for each 33-ms interval as the adjective–noun phrase unfolds. The dashed vertical lines demarcate the two regions of interest in the unfolding sentence, the Adjective/Noun and Post-NP windows. As an additional point of reference, the dotted line indicates the acoustic onset of the target noun. These response windows are offset by 300 ms from the onset of the adjective, allowing for time required to mobilize a signal-driven eye movement.

The curve indicated by a solid line depicts performance of 30-month-olds in the Blue Car/Blue House condition, when the adjective was uninformative. On these control trials, the mean proportion of looking time to the target was at chance until halfway through the Adjective/Noun window, when information from the noun became available. In contrast, looking time to the target began to rise earlier in the Adjective/Noun window in the two informative-adjective conditions, suggesting that some of the children were responding to the color word without waiting for the noun. However, Fig. 1A also reveals a noteworthy difference in performance between the two informative-adjective conditions: In the Blue Car/Red House condition, when both adjective and noun were informative, 30-month-olds responded more accurately overall than in the Blue Car/Red Car condition, when they had to rely entirely on the adjective to identify the target.

2.2.1. Accuracy

To assess these differences statistically, mean accuracy scores were calculated for each child in each condition in the Adjective/Noun and Post-NP windows, and these were compared in a 2 (window) $\times$ 3 (condition) repeated measures ANOVA. Overall accuracy was higher in the Post-NP window ($M = .82$, $SD = .14$), after children had heard the entire sentence, than in the earlier window ($M = .65$, $SD = .13$), leading to a significant main effect of window, $F(1, 31) = 28.89$, $p < .0001$. Although there was no main effect of condition, the condition $\times$ window interaction approached significance, $F(2, 30) = 2.91$, $p < .07$. To examine this interaction, mean accuracy scores in each of the three conditions were compared within each window. Performance in the Adjective/Noun window did not differ

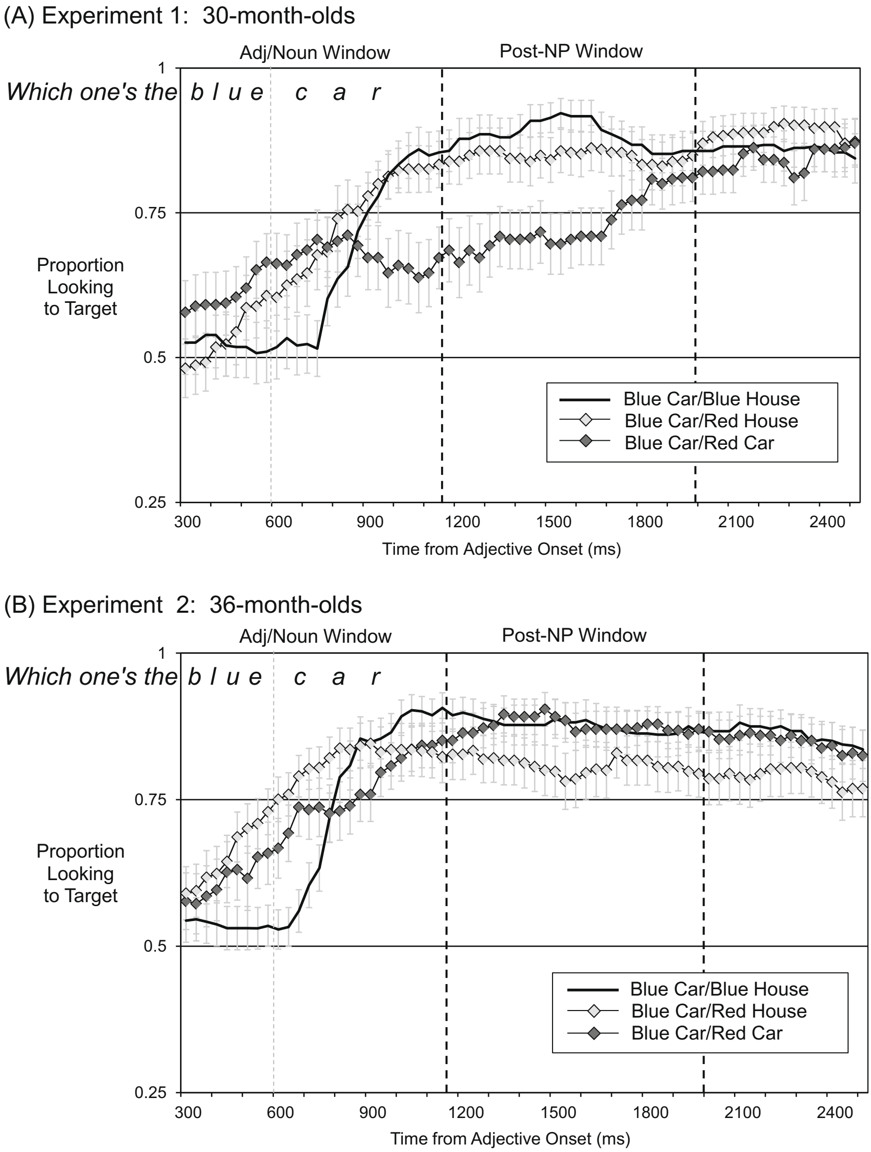

Fig. 1. Overall accuracy of responses to adjective–noun phrases in three conditions by children in two age groups: (A) 30-month-olds in Experiment 1; (B) 36-month-olds in Experiment 2. Curves show the mean proportion of trials on which children are fixating the target picture at each 33-ms interval as the stimulus sentence unfolds, measured from the acoustic onset of the prenominal adjective. Data from all 32 participants in each of the two age groups are included in these graphs. Error bars represent standard errors of the mean over participants.

among the three conditions. However, in the Post-NP window, planned pairwise comparisons indicated that 30-month-olds were significantly *less* accurate in the Blue Car/Red Car condition ($M = .72$, $SD = .25$) than in either the Blue Car/Red House ($M = .84$, $SD = .21$) or the control ($M = .88$, $SD = .15$) conditions, $F(2, 30) = 5.3$, $p < .02$.

When 30-month-olds heard *Where's the blue car?* while looking at a blue car and a red house, their performance was no different than on control trials, where the adjective was uninformative. Thus they did not appear to benefit from the potentially informative prenominal adjective as adults do. However when they heard the same sentence while looking at two cars, one blue and one red, the performance of 30-month-olds was significantly worse. What could account for this unexpected asymmetry? Were children simply slower to shift correctly from distracter to target on Blue Car/Red Car trials, or did they shift incorrectly away from the target more often? Either or both of these response patterns could account for the lower level of overall accuracy in this condition.

2.2.2. Onset-contingent plots

To begin to explore these questions, we examined the data more closely. While the time course plot of overall accuracy in Fig. 1A shows changes in the proportion of looking time to the correct referent as the adjective–noun phrase unfolds, collapsing across target- and distracter-initial trials, the onset-contingent plot in Fig. 2A offers a different view of the same data at a higher level of detail. For each of the three conditions, target- and distracter-initial trials are plotted separately, contingent on which picture the child was fixating 300 ms from the onset of the Adjective/Noun window. Plotted on the y-axis at each 33-ms interval is the mean proportion of trials on which children at that point are looking at a picture that is *different* from where they started at adjective onset. Because onset-contingent plots capture children's tendency to shift away from the original starting point, both correct and incorrect responses are tracked separately in the same figure: on distracter-initial trials, a shift away to the target picture is a *correct* response, while on target-initial trials, a shift away to the distracter picture is an *incorrect* response. The top three lines in Fig. 2A depict distracter-initial trials in each condition, showing changes in the mean proportion of fixation to the named picture as 30-month-olds shifted from distracter to target after the onset of the adjective. The three lower lines depict target-initial trials, showing changes in mean proportion of looking to the distracter when children shifted incorrectly away from the target picture.

The curves indicated by solid lines show children's responses in the Blue Car/Blue House control condition, when the color word could not be used to identify the target. On distracter-initial trials in this condition, shifting to the target picture did not begin to increase until mid-way through the Adjective/Noun window, when enough phonetic information from the noun had accumulated to enable identification of the target picture. Note that on target-initial control trials, 30-month-olds also tended to maintain fixation correctly; that is they did not often false alarm by shifting away from the target to the distracter picture.

In the two informative-adjective conditions, by comparison, it appears that some children began to shift from the distracter to the target picture earlier in the Adjective/Noun window than on control trials. However, in the Blue Car/Red Car condition, we see signs of disruption in the trajectories of *both* distracter-initial and target-initial trials. On distracter-initial trials, mean proportion looking to the target picture dropped off towards the end of the Adjective/Noun window. And on target-initial trials, 30-month-olds who were already looking at the correct referent at adjective onset tended to shift away incorrectly during the Post-NP window, a false-alarm response that was not evident in the other conditions. Thus the lower level of overall accuracy in the Blue Car/Red Car condition, apparent in Fig. 1A, resulted from a combination of two different kinds of inefficient response – shifting more slowly on distracter-initial trials and false-alarming on target-initial trials. It is interesting to note that the 30-month-olds eventually reached a high level of accuracy in identifying the correct referent, converging on the named target picture within 2 s from the start of the adjective–noun phrase. But in the Blue Car/Red Car condition, in which both target and distracter objects had the same noun label, they showed signs of disruption followed by recovery.

2.2.3. Reaction time

For each child, the mean RT to shift to the target picture on distracter-initial trials was calculated for each condition of Experiment 1.[2] Although in Fig. 2A it appeared that on some trials 30-month-olds

[2] RTs could only be calculated for trials on which children were by chance initially looking to the distracter picture and shifted between 300 and 2500 ms from the onset of the adjective; thus not every child had an RT in every condition. In Experiment 1, the analysis of mean RTs included the 19 of 32 participants who made distracter-to-target shifts in all three conditions. Mean RTs for each child in each condition were based on an average of 1.9 trials.

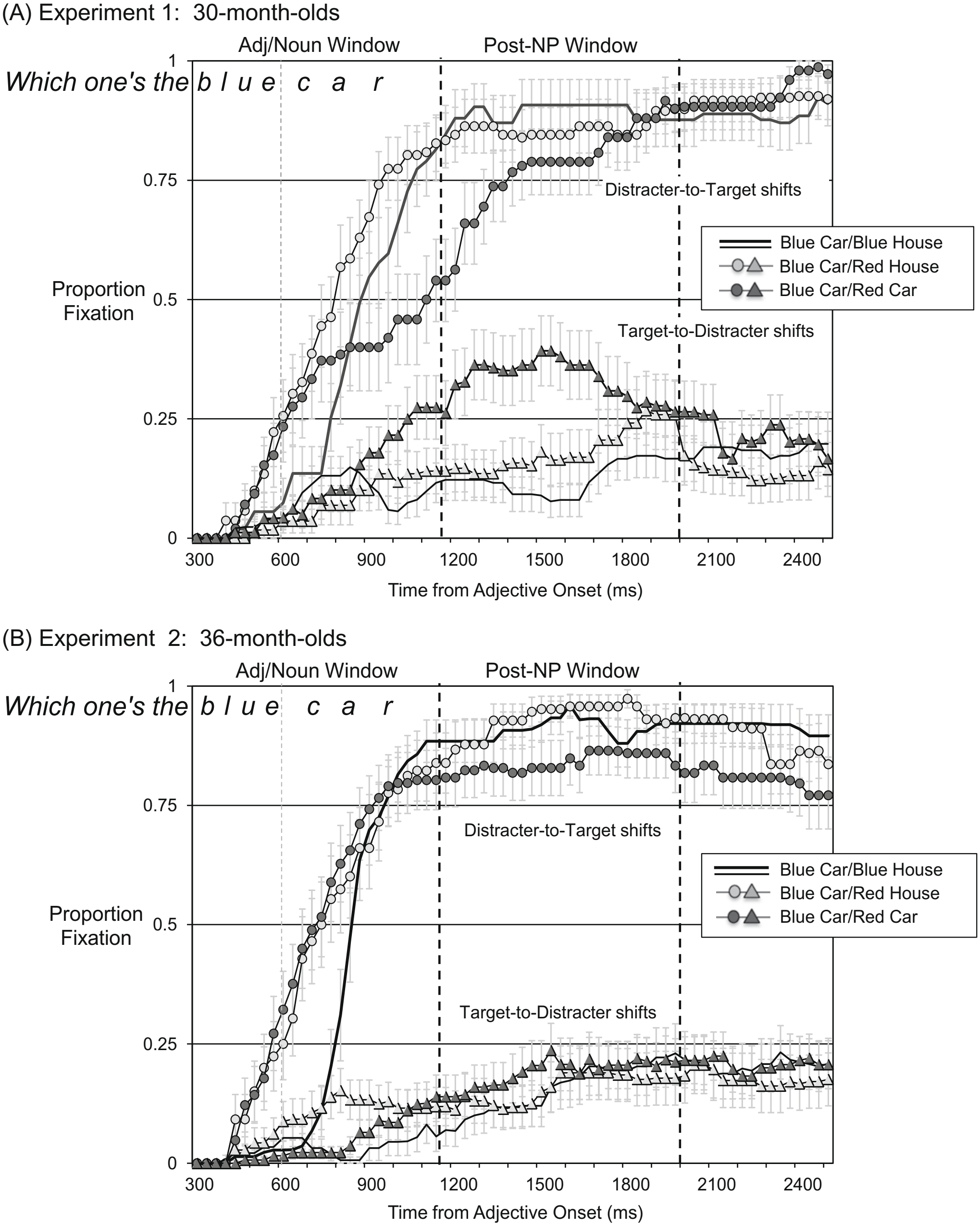

Fig. 2. Onset-contingent plot of distracter-initial and target-initial trials by children in two age groups: (A) 30-month-olds in Experiment 1; (B) 36-month-olds in Experiment 2. At each 33-ms interval, plots show the mean proportion of trials on which children have shifted away from the picture on which they started at adjective-onset, and are now fixating the other picture. In each graph, the top three curves show *distracter-initial* trials in the three conditions; on these trials a shift away from the distracter to the target picture is a *correct* response. The bottom three curves in each graph show *target-initial* trials, on which a shift away to the distracter picture is an *incorrect* response. Data from all 32 participants in each age group are included in these graphs. Error bars represent standard errors of the mean over participants.

Table 2
Mean proportion off-target for 30-month-olds in Experiment 1 and 36-month-olds in Experiment 2 by window.

Condition	30-Month-olds		36-Month-olds	
	Adj./Noun, *M* (*SD*)	Post-NP, *M* (*SD*)	Adj./Noun, *M* (*SD*)	Post-NP, *M* (*SD*)
Blue Car/Blue House (control)	.08 (.12)	.15 (.20)	.04 (.08)	.15 (.28)
Blue Car/Red House	.09 (.15)	.22 (.28)	.13 (.20)	.19 (.29)
Blue Car/Red Car	.13 (.17)	.35 (.32)	.06 (.10)	.23 (.26)
Overall	.09 (.09)	.23 (.22)	.07 (.08)	.19 (.16)

shifted early to the target on informative-adjective trials, a one-way repeated measures ANOVA on participant RT means on Blue Car/Blue House (M = 542 ms, SD = 218), Blue Car/Red House (M = 460 ms, SD = 336) and Blue Car/Red Car trials (M = 679 ms, SD = 462) revealed no significant main effect of condition, $F(2, 16) = 1.3$, $p = .30$. Consistent with the analysis of overall accuracy, the 30-month-olds as a group showed no RT advantage on trials when they heard a potentially informative adjective in advance of the noun.

2.2.4. Off-target responses

The next analysis focused on target-initial trials, when at adjective onset 30-month-olds were already looking at the correct referent. To assess the extent to which children showed confusion or disruption by shifting away from the correct picture on target-initial trials, the mean proportions of incorrect looking to the distracter during the Adjective/Noun and Post-NP windows were calculated in each condition.[3] If children were efficiently using the adjective to identify the referent, then looking away to the distracter should be minimal, with no differences between the two informative-adjective conditions. Mean proportions of off-target looking time during each window and each condition are shown in Table 2. Because these distributions were skewed toward zero, non-parametric Wilcoxin Signed Ranks tests were used to compare proportion off-target looking across windows and across conditions. Off-target looking was more likely to occur in the Post-NP window (M = .23) than in the Adjective/Noun window (M = .09), $Z = -3.5$, $p < .002$. In the Adjective–Noun window, paired comparisons revealed no reliable differences between conditions. In the Post-NP window, in contrast, mean proportion of off-target looking was significantly higher in the Blue Car/Red Car condition (M = .35) than in the control condition (M = .15), $Z = -2.5$, $p < .02$. However, off-target looking in the other informative-adjective condition, Blue Car/Red House (M = .22), was not reliably higher than in the control condition ($p = .23$). Thus, when 30-month-olds started out looking at the target object on Blue Car/Red Car trials, they were more likely to shift away from the correct picture *after* hearing the noun phrase, as compared to the other two conditions. This pattern of incorrect shifting is reflected in the "bump" in the curve representing target-initial trials in the Blue Car/Red Car condition in Fig. 2A, providing further evidence that trials in this particular condition were problematic for the younger children after the noun phrase had been heard.

2.2.5. Variability in efficiency of processing at 30 months

The results from Experiment 1 appear somewhat contradictory. On the one hand, correct early shifting to the target on distracter-initial trials is evident in the fixation patterns shown in Fig. 2A, suggesting that on some occasions 30-month-old children made use of the prenominal adjective to identify the referent. Yet the analysis of mean RTs showed that these children as a group did *not* identify the target any faster overall in informative-adjective conditions than in the control condition. Moreover, the analysis of off-target looking revealed that 30-month-olds were more likely to experience disruption on target-initial trials, shifting incorrectly to the distracter picture, particu-

[3] The measure of mean proportion off-target responding was based on target-initial trials, i.e. the subset of trials on which children were by chance initially looking to the target picture at 300 ms from adjective onset. In Experiment 1, the analysis of mean off-target response was based on the 25 of 32 participants who had target-initial trials in all three conditions. Means for each child in each condition were based on an average of 2.1 trials.

larly in the Blue Car/Red Car condition. Recall however, that children who looked briefly away had returned to the correct picture by 2500 ms after the onset of the adjective; thus they eventually succeeded in identifying the referent, even in the condition where use of the color word was essential. Thus the overall pattern of responses by 30-month-olds as a group in this condition reflects an inconsistent mix of efficient responses early on, combined with poor performance and disruption evident about 700 ms after the noun phrase, followed ultimately by success in correctly identifying the target picture.

The mixed findings in the Blue Car/Red Car condition could result from various patterns of responses by individual children. One possibility is that all the 30-month-olds were variable in their performance on both distracter- and target-initial trials, shifting rapidly to the target picture on some trials but more slowly on others, and maintaining fixation correctly on some of the target-initial trials while shifting back and forth indecisively on others. An alternative possibility is that individual children were relatively consistent in their patterns of response on target- and distracter-initial trials. Several studies have shown that children who are faster overall to respond to familiar words at a given age also tend to be more advanced in expressive vocabulary size and other measures of language development (Fernald et al., 2006; Hurtado et al., 2007). Thus one plausible hypothesis is that children in Experiment 1 who were faster overall on distracter-initial trials were also more efficient in their responses on target-initial trials, and conversely, that children who tended to shift more slowly to the target on distracter-initial trials were also more likely to shift back and forth between pictures more frequently on target-initial trials in the challenging Blue Car/Red Car condition.

To explore the paradoxical pattern of responses in the Blue Car/Red Car condition in Experiment 1, we examined whether the results observed in this group of 30-month-olds reflected contributions from individual children who showed characteristically different response patterns as the sentence unfolded. This analysis took advantage of the fact that our measure of processing disruption – in the form of incorrect shifting to the distracter on target-initial trials – was derived from a different subset of trials than the measure of response speed, which was based on correct shifts to the target on distracter-initial trials. Thus these measures provided independent sources of information reflecting different aspects of efficiency in online interpretation of adjective–noun phrases, allowing us to ask whether individual differences in the tendency to false alarm on Blue Car/Red Car trials were related to differences in processing speed across conditions. The 30-month-olds were divided into two groups based on a median split of the mean proportion of off-target shifting during the Post-NP window on Blue Car/Red Car trials. Children in the Low False Alarm (Low-FA) group either never or rarely shifted off the correct picture on target-initial trials ($M = .07$, $SD = .12$), while children in the High False Alarm (High-FA) group shifted frequently away from the target picture ($M = .62$, $SD = .33$). The question of interest was whether those children who responded more efficiently on target-initial trials, by maintaining fixation on the correct picture, were also more efficient on distracter-initial trials, shifting quickly to the referent named in the adjective–noun phrase.

Mean RTs for 30-month-olds in the Low-FA and High-FA groups in the three adjective conditions were compared in a 3 (condition) $\times$ 2 (FA group) mixed ANOVA, with FA group as the between-subjects factor. Children in the High-FA group were slower overall to shift correctly from distracter to target ($M = 743$ ms, $SD = 147$) than children in the Low-FA group ($M = 457$ ms, $SD = 147$), $F(1, 13) = 14.1$, $p < .002$. However, neither the main effect of condition nor the condition $\times$ group interaction was significant, suggesting that those 30-month-olds who were more likely to false alarm on target-initial trials in the Blue Car/Red Car condition were also slower in processing speed on distracter-initial trials across all three of the adjective conditions.

The onset-contingent plots for the Low-FA and High-FA groups shown in Fig. 3A and B reveal striking differences among 30-month-olds in their performance on Blue Car/Red Car trials. By definition, children in the High-FA group shifted incorrectly away from the target picture much more frequently than children in the Low-FA group, who only rarely false alarmed on target-initial trials. More importantly, these curves illustrate differences in the slopes of the distracter-initial trials, indicating that those children who experienced disruption on target-initial trials were also slower, on average, across conditions than children who were less likely to false alarm. The finding that

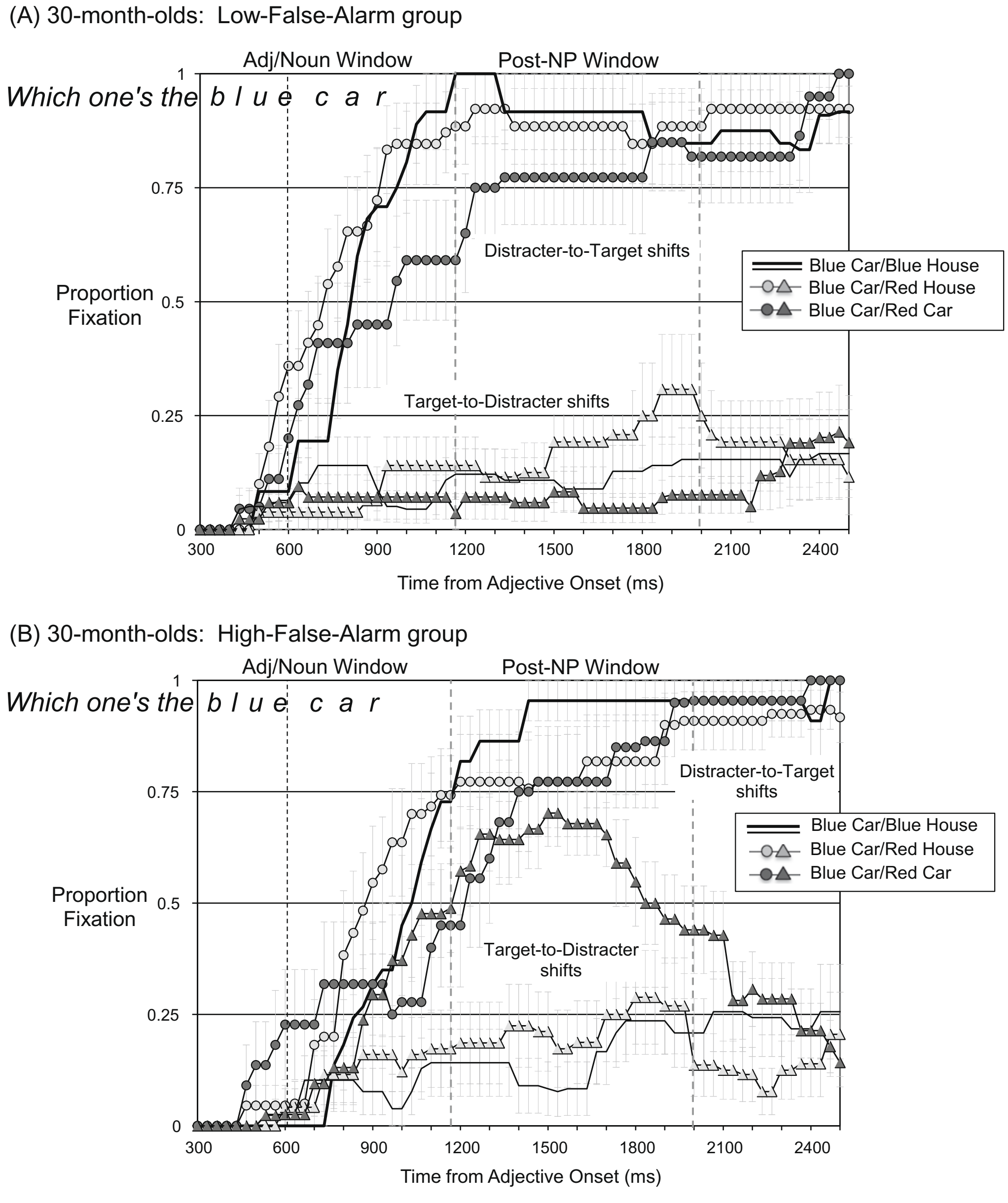

Fig. 3. Onset-contingent plot of distracter- and target-initial trials by 30-month-olds in Experiment 1, divided into two groups by a median split on the mean proportion of off-target looking on Blue Car/Red Car trials: (A) Low-False-Alarm group and (B) High-False-Alarm group. Plots show the mean proportion of trials on which children have shifted away from the picture they started on at adjective onset to fixate the other picture. In each graph, the top three curves show distracter-initial trials and the bottom three curves show target-initial trials in each of the three conditions. These graphs include data from all 32 participants in Experiment 1. Error bars represent standard errors of the mean over participants.

30-month-olds who were more decisive on target-initial trials were also faster to orient to the target object on distracter-initial trials suggests that some children had response patterns that were consistently more efficient than others.

2.3. Discussion of Experiment 1

Although the 30-month-olds in Experiment 1 were all reported to correctly use the adjectives *red* and *blue* in their own speech, we found that many had difficulty in the online interpretation of familiar adjectives in combination with nouns, as shown by three main findings: First, on trials when *red* or *blue* was sufficient to discriminate which picture was the target, 30-month-olds as a group did not take advantage of the potentially informative adjective before hearing the noun to establish reference more quickly, although they were ultimately able to use the color word to identify the correct referent. Second, 30-month-olds had particular difficulty on Blue Car/Red Car trials, when both objects were of the same kind and correct interpretation of the adjective was essential for establishing reference. Third, a closer look at the data revealed substantial variability among the 30-month-olds in their patterns of response on informative-adjective trials. Although Blue Car/Red Car trials were more challenging overall, some 30-month-old children showed evidence of extreme disruption in this condition, frequently false-alarming by shifting away from the target picture, and also shifting slowly to the target only after hearing the noun. These less efficient responders appeared to be momentarily unsure as to whether the car they were looking at was in fact the correct referent, initially rejecting the target before finally settling on the correct picture. Thus, while some children at 30 months showed a more mature pattern of incremental integration of the adjective with the following noun, others showed a less mature pattern of delayed responding, integrating the color word with the object name only after the end of the noun phrase.

3. Experiment 2

To explore further how online processing skills develop over the third year, in Experiment 2 we tested children 6 months older than the 30-month-olds in Experiment 1. There are various reasons to expect that the ability to rapidly integrate adjective–noun phrases would be more advanced at 36 than at 30 months. First, 36-month-olds have more adjectives in their productive vocabularies, and presumably have more well-developed lexical representations for these words. Second, while 2-year-olds show signs of difficulty in integrating adjectives and nouns when tested in offline tasks, by 36 months these difficulties have lessened. Thorpe et al. (2006) found that English-learning 30-month-olds were less accurate in identifying a target picture when they had to integrate the adjective with the noun than when they could rely on an adjective alone to distinguish the target; however, older children did not perform differently on these two tasks. Finally, while 2-year-olds have difficulty learning novel property terms when asked to extend them across different object types (e.g., Hall et al., 1993; Waxman & Kosowski, 1990), between 24 and 36 months children improve in their ability to extend novel adjectives correctly (Mintz, 2005). These findings all support the prediction that the 36-month-olds in Experiment 2 will have greater success in online interpretation of adjective–noun phrases than did the 30-month-olds in Experiment 1.

3.1. Method

3.1.1. Participants

Participants were 32 36-month-olds ($M = 36.3$; range = 35–39); 16 were female. All were typically-developing children from English-speaking families. Four additional participants were excluded due to inattentiveness throughout the test session. Parents reported that all children correctly produced the words *red* and *blue* in their spontaneous speech.

3.1.2. Stimuli, apparatus, procedure, coding and measures

Stimuli, apparatus, procedure, coding methods and measures were identical to those of Experiment 1. The reliability measure of proportion of frames over each trial on which two coders agreed yielded 98% agreement; for the more conservative reliability measure focusing only on shifts in gaze, ignoring steady-state fixations in each trial, coders agreed within one frame on 95% of all shifts.

3.2. Results of Experiment 2

3.2.1. Accuracy

Fig. 1B tracks changes over time in the mean proportion of trials on which 36-month-olds fixated the correct referent, averaged over participants for each 33-ms interval as the adjective–noun phrase unfolded. Mean accuracy scores for each child in each condition in the Adjective/Noun and Post-NP windows were compared in a 2 (window) $\times$ 3 (condition) repeated measures ANOVA. As in Experiment 1, the effect of window was significant, $F(1, 31) = 30.6$, $p < .0001$, with no main effect of condition, $F(2, 30) = 0.2$, $p = .84$. However, the window $\times$ condition interaction was significant, $F(2, 30) = 6.1$, $p < .006$. Pairwise comparisons of condition differences revealed that in the Adjective/Noun window, mean proportion of looking to target was significantly greater on Blue Car/Red House ($M = .71$, $SD = .22$) and Blue Car/Red Car trials ($M = .75$, $SD = .20$) than on control trials ($M = .65$, $SD = .16$), all $p < .03$, although in the Post-NP window there were no reliable differences among conditions. This finding shows that on those trials where the color word was potentially informative, 36-month-olds responded differentially as soon as they heard the adjective. In contrast, the 30-month-olds in Experiment 1 were no more accurate on informative-adjective trials than on control trials in the Adjective/Noun window. Not only were the younger children unable to make use of an informative adjective as they heard it, but many of them experienced disruption in the Post-NP window, as shown in Fig. 3B.

3.2.2. Onset-contingent plots

As in Experiment 1, we examined the data more closely by looking at responses on distracter- and target-initial trials plotted separately, shown in Fig. 2B. This graph looks different from the comparable plot for 30-month-olds in Fig. 2A, especially with respect to responses in the Blue Car/Red Car condition. The overlap in the curves for the Blue Car/Red House and Blue Car/Red Car trials in Fig. 2B suggests that 36-month-olds were more similar across the two informative-adjective conditions in their efficiency of shifting to the target, in contrast to the 30-month-olds who were substantially slower overall to shift to the target on Blue Car/Red Car trials. On target-initial trials as well, the responses of 30- and 36-month-olds appeared to differ: The tendency of 36-month-olds to false alarm by shifting incorrectly away from the target remained relatively low in all three conditions in Experiment 2, as compared to the higher false-alarm rate of 30-month-olds overall in the Blue Car/Red Car condition in Experiment 1.

3.2.3. Reaction time

We next examined the latencies of 36-month-olds to orient to the correct referent on distracter-initial trials. For each participant, a mean RT was calculated over trials in each condition.[4] A one-way repeated measures ANOVA over condition was not significant, $F(2, 17) = 1.9$, $p = .19$; however, planned comparisons showed that 36-month-olds were significantly faster to shift to the target object in the Blue Car/Red House condition ($M = 434$ ms, $SD = 322$) than in the control condition ($M = 578$ ms, $SD = 250$), $t(23) = -2.1$, $p < .05$, and the mean RTs in the two informative-adjective conditions were not reliably different, $t(20) = .01$, $p = .99$. However, the comparison between the Blue Car/Red Car condition ($M = 457$ ms, $SD = 345$) and the control did not reach significance, $t(22) = 1.3$, $p = .21$. Next, we re-computed mean RTs for each child, collapsing across the two informative-adjective conditions. A paired-samples t-test comparing performance on informative-adjective vs. control trials indicated that children were reliably faster to shift from the distracter to the target picture when the adjective was informative ($M = 449$ ms, $SD = 247$) than in the control condition, $t(27) = 2.1$, $p < .05$. In sum, 36-month-olds were quicker overall to orient to the target picture when they heard an adjective that uniquely identified the correct referent than when they had to wait for the noun label, although this effect was weaker in the Blue Car/Red Car condition than in the Blue Car/Red House condition.

[4] In Experiment 2, the analysis of mean RTs included the 19 of 32 participants who made distracter-to-target shifts in all three conditions. Mean RTs for each child in each condition were based on an average of 1.7 trials.

3.2.4. Off-target responses

The next analysis focused on target-initial trials, when 36-month-olds were already looking at the correct referent. To assess whether children demonstrated confusion by shifting away from the correct picture, we calculated the mean proportion of looking to the distracter during the Adjective/Noun and Post-NP windows in each condition.[5] As shown in Table 2, 36-month-olds rarely shifted incorrectly to the distracter picture on target-initial trials. The mean proportion off-target looking was higher in the Post-NP window ($M = .19$) than the Adjective/Noun window ($M = .07$), $Z = -3.5$, $p < .003$, although there were no differences among conditions.

3.2.5. Variability in efficiency of processing at 36 months

In this analysis, we asked whether the tendency to false alarm on Blue Car/Red Car trials was related to overall processing speed for 36-month-olds, as it was for the 30-month-olds. As in Experiment 1, we divided the 36-month-olds into groups based on a median split of false-alarm rates in the Post-NP window, where disruption was greatest for the younger children. Children in the Low-FA group almost never shifted off the correct picture on target-initial trials ($M = .01$, $SD = .02$), while children in the High-FA group shifted more often ($M = .42$, $SD = .19$), although substantially less than the 30-month-olds who were more prone to false alarms. Mean RTs for 36-month-olds in the Low-FA and High-FA groups in the three adjective conditions were compared in a 3 (condition) $\times$ 2 (FA group) mixed ANOVA, with FA group as the between-subjects factor. There were no significant effects of FA group or condition, and no interaction. Although children who were more likely to false alarm on target-initial trials were 150 ms slower to initiate a correct shift ($M = 601$, $SD = 202$) than children who false alarmed less often ($M = 450$, $SD = 202$), this difference was not reliable. Thus, there were fewer differences between the children in the two FA groups at 36 months than at 30 months.

3.3. Discussion of Experiment 2

The main finding in Experiment 2 was that when the prenominal adjective was potentially informative, 36-month-olds as a group could identify the referent of a phrase such as *blue car* as soon as they heard the word *blue*, without waiting to hear the entire noun phrase. And on trials when they started out looking at the correct referent, children at this age did not tend to shift away from the target incorrectly. This result may seem unsurprising, given that most 3-year-olds will choose correctly when asked to identify a blue car in an offline task, given unlimited time to respond. However, our results using online methods revealed a different aspect of this competence: by 36 months, many children can make immediate use of potentially informative adjectives in establishing reference. Although the color words were only 300 ms in duration, children were remarkably fast in integrating the discriminating information provided by these prenominal words. While 3-year-olds were a few hundred milliseconds slower to respond on average than adults in a similar task (e.g., Sedivy et al., 1999), they already showed evidence of mature efficiency, making incremental use of linguistic information as they interpreted speech in relation to the visual scene.

3.4. Comparison of results from Experiments 1 and 2

Comparing response patterns of the 30-month-olds in Experiment 1 with those of the 36-month-olds in Experiment 2, the older children were more efficient overall in identifying the target picture as soon as they heard the prenominal adjective, when the color word enabled them to identify the correct referent without waiting for the noun. The most striking difference between the two age groups was in their performance on Blue Car/Red Car trials. When asked to *Find the blue car* while looking at a blue car paired with a red car, 36 month-old children tended to maintain fixation on the correct picture on target-initial trials, while many of the 30-month-olds were highly likely to look away to the

[5] In Experiment 2, the analysis of off-target responding was based on the 27 of 32 participants who had target-initial trials in all three conditions. Means for each child were based on an average of 2.0 trials with target-initial onsets in each of the three conditions.

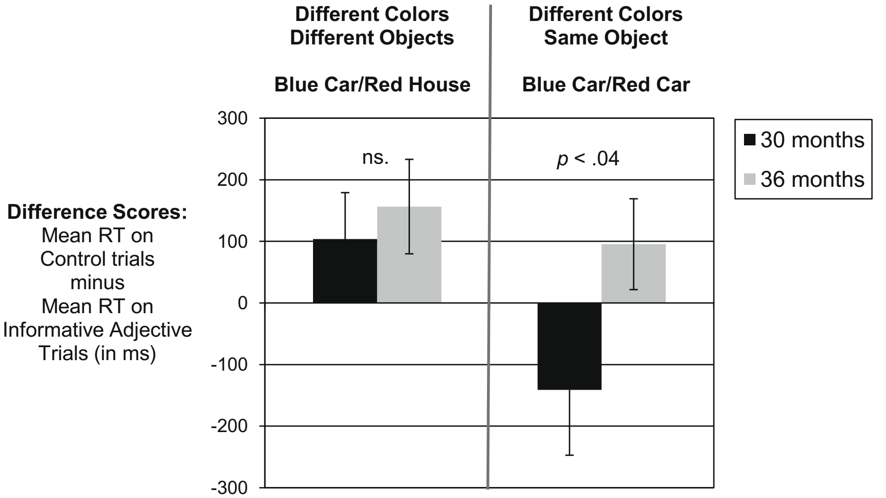

Fig. 4. Difference scores on Informative-Adjective trials as compared to Control trials for 30- and 36-month-olds in Experiments 1 and 2. These scores were calculated for each child by subtracting the mean RTs on Blue Car/Red House trials and Blue Car/Red Car trials from the mean RT on Blue Car/Blue House trials. The data in these graphs represent 20–22 participants at each age who contributed RTs in the relevant conditions. Error bars represent standard errors of the mean over participants.

distracter before returning to the correct picture. Thus the pattern of disrupted *post-nominal* responding observed in Experiment 1 was most evident in a subset of the younger listeners.

The next analysis focused on age differences in efficiency on distracter-initial trials, examining the *relative* differences in response speed on informative-adjective trials as compared to control trials by the 30- and 36-month-olds in Experiments 1 and 2. Independent of any differences in absolute processing speed between these age groups, this analysis compared the extent to which the younger and older children were able to make use of a prenominal adjective by shifting relatively faster to the target on trials when the color word was potentially informative. Difference scores were calculated for each child by subtracting their mean RT in each of the two informative-adjective conditions from their mean RT on control trials, as shown in Fig. 4. Thus a positive difference score indicates a faster response on trials when the adjective was potentially informative than on trials when the child had to wait for the noun to identify the referent. On Blue Car/Red House trials, children at both ages had positive difference scores, with no significant difference between age groups ($p = .63$). In contrast, on Blue Car/Red Car trials there was a clear disparity between the difference scores at 30- and 36-months, $t(41) = 1.9$, $p < .04$, one-tailed. While the mean difference score at 36 months was again positive, at 30 months, the mean difference score was below zero, suggesting that children in this group were actually faster on control trials than on Blue Car/Red Car trials. Thus on trials when the two objects differed in color but were of the same kind, 30-month-olds were reliably *less* efficient in identifying the referent when the color word was potentially informative than when it was not.

4. General discussion

Previous studies of children's emerging ability to make sense of adjectives have relied extensively on offline measures that aim to assess categorically whether or not a child "knows" a particular adjective, based on end-point responses made after the child has heard the word of interest. The experiments reported here take a new approach by investigating children's developing skill in interpreting adjectives in fluent speech, highlighting the importance of understanding in real time. In many situations, adults make use of contextual information to interpret an adjective as it is heard to narrow in on a referent, without waiting for the object to be named (Sedivy et al., 1999). Here we examined young children's ability to interpret adjective–noun phrases incrementally, using their gaze patterns in relation to potential referents in the visual array as an index of understanding. In Experiment 1, when 30-month-old English-learning children heard an adjective–noun phrase such as *blue car*, they failed to take advantage of the prenominal color word to identify the named object more rapidly, and also had difficulty integrating the adjective and noun in combination. The 36-month-olds in Experiment 2 were more successful in this task. When the adjective was potentially informative, they were more likely to orient to the blue car as soon as they heard the word *blue*, without waiting to hear the entire noun phrase. These findings show that children make impressive gains over the third year in understanding sentences in real time. Yet they also reveal that interpreting an apparently simple noun phrase containing a familiar color word can be a challenging processing task for the young language learner, and that becoming more fluent in understanding adjective–noun phrases involves developing efficiency in integrating words in combination.

4.1. The challenge of integrating adjectives with nouns

Studies exploring how adults integrate word meanings in adjective–noun phrases have focused on the process of interpreting a property term in relation to the particular noun it modifies, when a single conceptual unit is created from two words; for example, *blue* and *car* are represented jointly as referring to a unified blue car (e.g., Begg, 1972). Research on how adjective–noun integration affects memory capacity provides one kind of evidence. When presented with lists containing either isolated adjectives and nouns, or two-word adjective–noun phrases, adults recalled as many adjective–noun pairs as either nouns or adjectives alone, indicating they had semantically integrated the adjective–noun pairs to represent them as single units (Paivio, Khan, & Begg, 2000). Other studies have examined the time course of modifier-noun integration, asking whether the process of conceptual combination proceeds sequentially. When interpreting a phrase such as "peeled apple", do adults access the meaning of each word before combining the two words into a new concept? Experiments by Springer and Murphy (1992) suggest that the component words are not necessarily analyzed prior to interpreting the adjective–noun combination. More recent findings challenge this conclusion, showing that adult listeners in a cross-modal priming task activate features of the head noun before activating properties related to the adjective–noun combination (Swinney, Love, Walenski, & Smith, 2007).

Although the question of whether adults access the meanings of individual words sequentially in modifier-noun phrases before interpreting the words in combination is still under discussion, in developmental research this discussion has just begun (e.g. Huang & Snedeker, 2008; Thorpe & Fernald, 2006). Ninio (2004) first proposed that young children struggle to interpret adjective–noun phrases correctly because they have difficulty integrating the meanings of two words to identify a single referent, but this research did not address the issue of sequential processing. Here we extend these results using high-resolution measures of gaze patterns time-locked to the unfolding speech stimulus, to examine more closely how children develop skill in real-time processing of adjective–noun phrases between 30 and 36 months. We discovered that while some of the younger children responded quickly and accurately, others had particular difficulty on trials in one condition. When asked to choose between two objects that differed in color but not in kind, more than half of the 30-month-olds made frequent incorrect shifts to the distracter on trials when they started out looking at the target picture. And those 30-month-olds who were more likely to false alarm were also relatively slower to shift from the distracter to the target object, waiting until after the end of the adjective–noun phrase to make their move. Although these less efficient 30-month-olds still found their way to the correct referent just 2 s after hearing the adjective, the high-resolution time-course measures used in this research revealed an unexpected 'boggle' in processing that was not observed in the older children, and that would not have become apparent using offline measures.

4.2. Developmental changes in processing efficiency

The 36-month-olds in this study were more successful overall than children six months younger in interpreting familiar color words in combination with familiar object names, but how should we characterize this improvement in the fluency of their understanding? Here we consider explanations at three different levels, asking how these results might be related to differences in children's color word knowledge, processing speed, and/or facility in integration.

First, there is ample evidence that children in this age range vary substantially in their lexical and syntactic knowledge. If 30-month-olds are less advanced in their learning of color words, could this account for differences in online comprehension? In the developmental literature, where parent-report vocabulary inventories and forced-choice measures are often interpreted categorically as evidence that a child either "knows" or "does not know" particular words, there is a tendency to characterize lexical knowledge in an all-or-none fashion. However, as Carey (1978) pointed out many years ago, other evidence suggests that lexical representations are built up gradually and become steadily more robust through experience. In a longitudinal study from 25 to 30 months, Sandhofer and Smith (1999) found that color word knowledge involves multiple aspects of understanding that develop slowly over the third year. Children first understand that a word such as *blue* is an appropriate answer to the question *What color is this*? – without yet mapping *blue* to a particular hue. They next show increasingly reliable evidence of color word comprehension, followed by production, and finally by the ability to group objects by color in a sorting task. Across different studies, there is some variation in the order in which these various aspects of color term knowledge become apparent (e.g., Kowalski & Zimiles, 2006; Rice, 1979; Soja, 1994). But the fact that these are all gradual processes that can be dissociated in an individual language learner suggests that the children in our study may also have varied in the robustness of their lexical representations of these words. Although all the 30- and 36-month-olds in this study were reported to produce *red* and *blue*, and thus met traditional criteria for demonstrating "knowledge" of these words, the older children had 6 months more experience interpreting and producing color words. Thus differences in color word familiarity could have contributed to the differences in processing efficiency both between and within age groups at 30 and 36 months. However, the fact that the 30-month-olds had less difficulty interpreting *blue* and *red* on Blue Car/Red House trials than on Blue Car/Red House trials suggests that differences in the strength of color word representations cannot fully account for the differences observed in performance.

A second possibility is that age-related differences in processing speed could contribute to the greater success of the 36-month-olds on informative-adjective trials. Several studies using the looking-while-listening paradigm have found gradual improvement in processing speed with age in children learning English (Fernald et al., 1998, 2006; Zangl & Fernald, 2007; Zangl et al., 2005) and Spanish (Hurtado et al., 2007). Between 15 and 36 months, children's mean response latency to orient to target pictures labeled with familiar object names decreases by around 200–300 ms. However, in the present study, 30-month-olds were not slower overall than 36-month-olds on control trials, when the adjective was irrelevant and all they had to do was identify the familiar noun. Nor did they differ in response speed on Blue Car/Red House trials, when either the adjective or the noun could be used to identify the referent. It was only on the more challenging Blue Car/Red Car trials, when the adjective was both necessary and sufficient to identify the correct referent, that the 30-month-olds were more than 200 ms slower to respond than the 36-month-olds.

Although this pattern of results cannot be explained by a gradual increase in processing speed between 30 and 36 months, processing speed differences *within* the younger age group were associated with differential success in online interpretation of adjective–noun phrases. In evaluating potential advantages of faster processing speed, consider first children's performance on Blue Car/Red House trials, when the child is asked to *Find the blue car*. Here the color word is sufficient to identify the correct referent, but the noun is also sufficient. On distracter-initial trials when the child happens to start out looking at the red house and is able to interpret the color word correctly, then hearing the word *blue* should motivate a shift to the other picture. A rapid shift initiated during the word *blue* would land the child on the blue car as the word *car* is heard, at which point the child only has to verify that the currently fixated object is indeed a blue car. If the child responds more slowly and is still looking at the red house as the word *car* is heard, the mismatch between the incoming noun and the currently fixated object should also be enough to motivate a shift to the blue car. Thus on Blue Car/Red House trials, where the child can rely on either the adjective or the noun to identify the referent, differences in mean RT were not linked to success in referent identification, since children who responded more slowly could use the noun as well as the adjective. But this does not explain the substantial difference in mean RT between age groups on Blue Car/Red Car trials. When asked to *Find the blue car* given two cars differing in color, 30-month-olds were much slower overall than 36-month-olds to shift correctly to the target. Our examination of variability in responses revealed that those 30-month-olds who were slower overall to shift correctly from distracter to target at this age were also much more likely to make incorrect shifts from target to distracter, although this link between mean RT and incorrect shifting was no longer evident at 36 months. While differences in response speed may contribute to the overall pattern of these results, the tendency of some of the younger children to shift more slowly across conditions and also to false-alarm more frequently on Blue Car/Red Car trials cannot be explained in terms of developmental gains in processing speed from 30 to 36 months.

A third possibility is that performance differences between younger and older children on Blue Car/Red Car trials can be framed in terms of a problem with semantic integration. This argument is supported by the distinctive pattern of responses shown by the less efficient 30-month-olds, a combination of substantially delayed shifts to the correct referent on distracter-initial trials and more off-target looking to the incorrect picture on target-initial trials. On Blue Car/Red Car trials, when the child is asked to *Find the blue car*, the color word is necessary to identify the referent – the child cannot rely solely on the noun as in the other two conditions. If the child starts out looking at the red car, then hearing *blue* should motivate a shift to the other picture, just as in the example of the Blue Car/Red House trial discussed above. If the shift is initiated during the adjective, the child will be looking at the blue car by the end of the phrase *blue car*. However, a child who responds to *blue* more slowly may still be looking at the red car as the word *car* is heard, a potentially problematic situation. The mismatch between the word *blue* and the color of the red car may motivate a shift to the other picture; but because the next word *car* is actually compatible with the object the child is still looking at, even if the color word the child just heard was not, this conflict may lead to momentary competition and inhibition of the tendency to shift away from the red car. If the less efficient 30-month-olds are processing the two words sequentially – first *blue*, followed by *car* – and only slowly binding them into a unified conceptual combination, they might first match the noun in the phrase *blue car* with the red car they are currently looking at, resulting in a significant delay in shifting to the correct referent.

Difficulty in combinatorial processing could account not only for the delayed responses of the 30-month-olds on distracter-initial trials, but also for their higher false-alarm rate on target-initial trials. When these children started out looking at the blue car and heard *blue car*, they frequently abandoned the correct picture, checking out the red car in the distracter picture as if they were uncertain which of the two cars was being referred to. If combining the meanings of the modifier and noun is a slow and difficult process for these children, and the word *car* is not immediately integrated with the word *blue*, then hearing *car* could motivate consideration of the red car as another potential referent. In contrast, the 36-month-olds tended to maintain fixation on the correct picture on target-initial trials, efficiently integrating the noun with the adjective in the phrase *blue car* and recognizing that the two-word combination could refer only to the car they were already looking at.

Although developmental differences in vocabulary knowledge, processing speed, and skill in semantic integration could all be contributing factors, none explains why the difference in efficiency between older and younger children would be so much greater on Blue Car/Red Car trials than in the other informative-adjective condition. Those 30-month-olds who were faster in mean RT were also less likely to shift incorrectly on target-initial trials, but this tantalizing link between processing speed and success in integration is complex. A slower response to the adjective may invite interference from the noun, especially on trials when the noun on its own could refer to either of the two objects present. However, these data cannot tell us whether faster processing speed actually enables greater success in integrating the adjective with the subsequent noun, or whether other factors account more directly for the development of skill in integration.

4.3. Sources of variability in children's efficiency in interpreting adjective–noun phrases

Our findings not only revealed developmental gains in processing efficiency from 30 to 36 months but also interesting variability within the group of 30-month-olds. Experiment 2 showed that by 3 years, English-speaking children from the middle-class population represented in our sample had developed some facility in integrating a prenominal color word with the noun that followed. Experiment 1 showed that children 6 months younger were less efficient in this task – not because they were simply slower overall, but because many of the children at this age showed a distinctive pattern of disruption in processing that indicated difficulty in interpreting a familiar color word in combination with a familiar noun. Does this processing 'boggle' reflect a developmental transition that English-learning children typically go through around this age? A longitudinal replication of these cross-sectional findings suggests this may be the case. Using stimuli and procedures comparable to the present study, Thorpe (2007) tested the same 32 children at 24, 30, and 36 months. On Blue Car/Red Car trials, almost all the 24-month-olds showed both delayed shifts to target on distracter-initial trials, and substantial looking away from the correct picture on target-initial trials. Although many of the 24-month-olds could produce *red* and *blue*, few were able to integrate a color word with a familiar object name in real-time interpretation of an adjective–noun phrase. By 30 months, the pattern of results observed was similar to that in the present study, with about half the children processing adjective–noun phrases fluently and the other half still showing less efficient processing. However, this difficulty in semantic integration had resolved to a large extent by 36 months, when only a few of the children showed disruption in processing in this task. Thus these longitudinal results confirmed both the overall pattern of developmental change and the distinctive patterns of variability in processing efficiency at 30 months observed in the current cross-sectional study.

What factors could contribute to such variability among children at 30 months? We consider two possibilities: first, that differential success in adjective–noun integration at 30 months could be related to individual differences in lexical development that are linked to differences in processing efficiency, and second, that differences in working memory may also play a role. We suggested earlier that differences in color word knowledge might account for the difficulty some children experienced in using *red* and *blue* to identify the correct referent. However, it could also be that vocabulary knowledge is more broadly influential, interacting dynamically with processing efficiency. Several studies using online measures with young children have shown that greater processing efficiency is associated with more advanced lexical development. For example, the Fernald et al. (2006) longitudinal study found that faster processing speed at 25 months was associated with more rapid vocabulary learning across the second year. One explanation for these relations is that infants who are faster to encode the speech signal in relation to the visual scene can attend more efficiently to subsequent information in the speech stream and thus develop a more robust network of lexical-semantic representations that can be accessed more reliably during comprehension. Because data on vocabulary size were not available for the children in this study, we were unable to determine whether individual differences in lexical development at 30 months accounted for any of the variability in adjective–noun integration in Experiment 1.

A second possibility is that differences in children's skill in interpreting adjective–noun phrases are related to developmental differences in core cognitive capacities such as working memory. Children's success in identifying familiar words and in learning new words requires the ability to store and manipulate information over short periods. Studies with both children and adults show robust links between processing speed, vocabulary size, and performance on working memory tasks (Gathercole & Baddeley, 1993; Kail & Salthouse, 1994; Montgomery, 2002). Moreover, a recent follow-up of children in the Fernald et al. (2006) longitudinal study revealed that processing speed and vocabulary knowledge at 25 months predicted performance on tests of language and cognition at 8 years of age (Marchman & Fernald, 2008). These results were the first to show that early efficiency in understanding has long-term predictive validity, and that processing speed and vocabulary in infancy are strongly related to measures of working memory in later childhood. These factors may operate in a developmental cascade in which faster speech processing enables more efficient learning of new lexical forms, together facilitating further development in working memory.

Semantic short-term memory capacity, a sub-type of verbal working memory, may be particularly relevant to the kinds of processing challenges children faced in the present study. Research exploring semantic memory problems in patients with damage to left inferior frontal and parietal areas of the brain show that both adult and 10-year-old patients with semantic short-term memory deficits have difficulty comprehending and producing adjective–noun phrases (Hanten & Martin, 2000; Martin & He, 2004). For example, these patients performed poorly when asked to judge whether sentences containing multiple prenominal adjectives were anomalous in meaning (e.g., *She had a rusty old red bathing suit*), suggesting that it was challenging for them to maintain more than one or two property terms in working memory and integrate them with the upcoming noun (Martin & Romani, 1994). In contrast, when adjectives *followed* the noun in control sentences (e.g., *The bathing suit was old, red, and rusty*), patients no longer had difficulty detecting the semantic anomaly. When the head noun came first, listeners could immediately integrate the adjectives into the sentence as they were heard without overtaxing semantic working memory.

If the processing load is greater for adults when an adjective precedes rather than follows the noun, then young children may also be challenged more by an adjective in prenominal position than by the same adjective heard after the noun, and the extent of this challenge may be related to working memory. The present study does not address these hypotheses, but ongoing research in our laboratory explores several possibilities. One series of studies investigates whether individual differences in the non-verbal working memory skills are associated with differential success in interpreting adjective–noun phrases. Other ongoing studies explore the question of memory load using a cross-linguistic design. If it is easier for children to identify a noun first and *then* integrate the adjective as a modifier, as compared to processing a prenominal adjective followed by a noun, then English-learning children have to contend with a less than optimal arrangement. By testing children learning Spanish, a language in which color adjectives are always heard postnominally, we are investigating whether word order plays a role in how fluently adjective–noun combinations are understood. It could be that the order of adjective and noun is irrelevant, and that children initially have difficulty with serial integration of any two content words within a noun phrase, regardless of which one comes first. Or it could be that adjective–noun ordered phrases, as in English, are fundamentally harder to interpret incrementally than are noun-adjective ordered phrases, as in Spanish and Hebrew, given inherent conceptual differences between object labels and property terms that make concrete nouns easier to process than adjectives (e.g. Gentner, 1982).

4.4. Getting to know a color word is a gradual process

An adult who claims to know an unusual color word such as *celadon* would not only be able to identify this color and define the word in relation to other hues (e.g., pale bluish green), but would also be able to produce and interpret this word appropriately in many different contexts. For example, it would be intuitive to the English-speaking adult that an adjective in prenominal position serves to restrict the reference of the following noun. Thus the phrase *celadon vase* could only be used to refer to a vase of that particular color, even if other celadon objects – or vases of other colors – are also present. For young children, such proficiency in understanding and using color words emerges slowly over several years. Although parents often proudly assert that toddlers "know" the words for primary colors like red and blue by the age of two, this categorical conclusion may be premature, given that crucial aspects of color categorization and color-word proficiency are typically not evident until much later (Carey, 1978; Sandhofer & Smith, 1999). Moreover, the development of children's skill in interpreting adjectives more broadly is a prolonged process, and 3- and 4-year-olds continue to have difficulty learning novel property terms and extending them appropriately in some contexts (e.g. Diesendruck, Hall, & Graham, 2006).

By focusing on the time course of young children's interpretation and integration of a prenominal color word with the following noun, our results illuminate an aspect of the development of fluency in understanding adjectives that has not been explored in such fine-grained detail in previous research. One important finding was that when presented with a blue car and a red car and asked *Where's the blue car*, 36-month-olds tended to orient to the correct referent after hearing the word *blue*, while 30-month-olds did not. On these trials, the 36-month-olds seemed to appreciate that the prenominal adjective restricted the reference of the following noun, decisively choosing the blue car without being distracted by the presence of another car that was not blue, comparable to adult performance in a similar task using eye-tracking measures (Sedivy et al., 1999).

However, the more surprising and revealing finding in our research was how difficult this task turned out to be for some of the younger children. The 30-month-olds in Experiment 2 all produced *red* and *blue* in their spontaneous speech, and most of them also identified the correct referent in our online processing task – eventually. But for some children at this age, we observed a brief moment of apparent uncertainty as they encountered the adjective–noun phrase in real time, on those trials when they saw two objects of the same kind that differed in color. The online measures revealed a post-nominal boggle by these younger children, as they struggled to integrate the color word with the noun that followed. That is, they responded as if they were processing the two words in the phrase *blue car* separately rather than binding them into a conceptual unit, and thus thought the word *car* might also refer to the red car that was also present. Moreover, this disruption was part of a pattern of inefficiency in their responses, more likely to occur in those younger listeners who were not as fast to make use of the potentially informative color words. Thus these seemingly simple adjective–noun phrases were actually a formidable language processing challenge for many of the 30-month-olds, who experienced temporary difficulty linking the color word with the following noun to identify the referent that was both blue *and* a car. The striking differences we found in the processing efficiency of children mid-way through their third year could not have been characterized in such detail by the offline measures of lexical knowledge and sentence comprehension used extensively in previous developmental studies. These results provide further evidence that knowing a color word involves much more than being able to match that word correctly to a color sample. Knowing the word *blue* also involves the ability to interpret that word fluently in combination with other words in continuous speech. For the young language learner, efficiency in rapid incremental interpretation of adjective–noun phrases emerges gradually, as the child makes progress over the third year towards more mature skill in real-time language processing.

Acknowledgments

This research was supported by grants from the National Institutes of Health to Anne Fernald (HD 42235 and DC 008838). Our thanks to Dr. Renate Zangl, Dr. Nereyda Hurtado, Grace Budde, Heidi Baumgartner, Casey Lew-Williams, Ana Luz Portillo, and the research assistants at the Center for Infant Studies at Stanford for their many contributions, and to the parents and infants who participated in this study.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17*, 1265–1282.

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language, 32*, 159–173.

Barner, D., & Snedeker, J. (2008). Compositionality and statistics in adjective acquisition: 4-Year-olds interpret *tall* and *short* based on size distributions of novel noun referents. *Child Development, 79*, 594–608.

Bartlett, E. J. (1978). The acquisition of the meaning of colour terms: A study of lexical development. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language: Language development and mother child interaction* (pp. 89–108). New York: Plenum.

Begg, I. (1972). Recall of meaningful phrases. *Journal of Verbal Learning & Verbal Behavior, 11*, 431–439.

Booth, A. E., & Waxman, A. (2009). A horse of a different color: Specifying with precision infants' mappings of novel nouns and adjectives. *Child Development, 80*, 15–22.

Bornstein, M. H. (1985). Colour-name versus shape-name learning in young children. *Journal of Child Language, 12*, 387–393.

Canfield, R. L., Smith, E. G., Brezsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life: A longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development, 62*, 1–45.

Carey, S. (1978). The child as word learner. In M. Halle, J. J. Bresnan & A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Clark, E. V. (1972). On the child's acquisition of antonyms in two semantic fields. *Journal of Verbal Learning & Verbal Behavior, 11*, 750–758.

Clifton, C., & Ferreira, F. (1989). Ambiguity in context. *Journal Language & Cognitive Processes, 4*, 77–103.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28*, 125–127.

Diesendruck, G., Hall, D. G., & Graham, S. A. (2006). Children's use of syntactic and pragmatic knowledge in the interpretation of novel adjectives. *Child Development, 77*, 16–30.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*, 409–436.

Fernald, A., Perfors, A. F., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the second year. *Developmental Psychology, 42*, 98–116.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*, 228–231.
Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72*, 1003–1015.
Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. Sekerina, E. Fernández, & H. Clahsen (Eds.), *Language processing in children*. Benjamins: Amsterdam.
Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 31–45). New York: Springer-Verlag.
Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.
Gelman, S. A., & Markman, E. M. (1984). Implicit contrast in adjectives and nouns: Implications for word learning in preschoolers. *Journal of Child Language, 12*, 125–143.
Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development volume 2: Language, thought, and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge, UK: Cambridge University Press.
Haith, M. M., Wentworth, N., & Canfield, R. L. (1993). The formation of expectations in early infancy. *Advances in Infancy Research, 8*, 251–297.
Hall, D. G., & Graham, S. A. (1999). Lexical form class information guides word-to-object mapping in preschoolers. *Child Development, 70*, 78–91.
Hall, D. G., Waxman, S. R., & Hurwitz, W. M. (1993). How two- and four-year-old children interpret adjectives and count nouns. *Child Development, 64*, 1651–1661.
Hanten, G., & Martin, R. C. (2000). Contributions of phonological and semantic short-term memory to sentence processing: Evidence from two cases of closed head injury in children. *Journal of Memory and Language, 43*, 335–361.
Henderson, J. M., & Ferreira, F. (Eds.). (2004). *The interface of language, vision, and action: Eye movements in the visual world*. New York: Psychology Press.
Huang, Y., & Snedeker, J. (2008). The use of referential contrast in children's on-line interpretation of scalar adjectives. In *Paper presented at the 32nd Boston University Conference on Language Development*. Boston, MA.
Hurtado, N., Marchman, V. A., & Fernald, A. (2007). Spoken word recognition by Latino children learning Spanish as their first language. *Journal of Child Language, 34*, 227–249.
Johnson, E. G. (1977). The development of color knowledge in preschool children. *Child Development, 48*, 308–311.
Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition, 57*, 129–191.
Klibanoff, R. S., & Waxman, S. R. (2000). Basic level object categories support the acquisition of novel adjectives: Evidence from preschool-aged children. *Child Development, 71*, 649–659.
Kowalski, K., & Zimiles, H. (2006). The relation between children's conceptual functioning with color and color term acquisition. *Journal of Experimental Child Psychology, 94*, 301–321.
Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*, F9–F16.
Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language, 89*, 76–82.
Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropshychology, 8*, 506–523.
Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20*, 158–190.
Mintz, T. H. (2005). Linguistic and conceptual influences on adjective acquisition in 24- and 36-month-olds. *Developmental Psychology, 41*, 17–29.
Mintz, T. H., & Gleitman, L. R. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition, 84*, 267–293.
Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science, 12*, 529–562.
Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language, 29*, 259–288.
Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*, 1–135.
Ninio, A. (2004). Young children's difficulty with adjectives modifying nouns. *Journal of Child Language, 31*, 255–285.
Paivio, A., Khan, M., & Begg, I. (2000). Concreteness of relational effects on recall of adjective–noun pairs. *Canadian Journal of Experimental Psychology, 54*, 149–160.
Partee, B. (1995). Lexical semantics and compositionality. In D. Osherson (Ed.), *Invitation to cognitive science* (2nd ed.. In L. Gleitman & M. Liberman (Eds.). *Part I: Language*, pp. 311–360). Cambridge: MIT Press.
Pitchford, N., & Mullen, K. T. (2005). The role of perception, language, and preference in the developmental acquisition of basic color terms. *Journal of Experimental Child Psychology, 90*, 275–302.
Rice, M. L. (1979). The effect of children's prior nonverbal color concepts on the learning of color words. (Doctoral dissertation, University of Kansas, 1978). *Dissertation Abstracts International, 39*(8-A), 4915.
Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology, 35*, 668–679.
Sandhofer, C. M., Smith, L. B., & Luo, J. (2000). Counting nouns and verbs in the input: Differential frequencies, different kinds of learning? *Journal of Child Language, 27*, 561–585.
Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*, 109–149.
Smith, E. E., & Osherson, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8*, 337–361.
Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12*, 485–527.

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

Soja, N. N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Development, 65*, 918–937.

Springer, K., & Murphy, G. L. (1992). Feature availability in conceptual combination. *Psychological Science, 3*, 111–117.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representations in very young children. *Cognition, 76*, 147–166.

Swingley, D., Pinto, J., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*, 73–108.

Swinney, D., Love, T., Walenski, M., & Smith, E. E. (2007). Conceptual combination during sentence comprehension. *Psychological Science, 18*, 397–400.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*, 557–580.

Taylor, M., & Gelman, S. A. (1988). Adjectives and nouns: Children's strategies for learning new words. *Child Development, 59*, 411–419.

Thorpe, K. (2007). Young children's developing skill in real-time interpretation of adjective–noun phrases [Doctoral dissertation, Stanford University, 2007].

Thorpe, K., Baumgartner, H., & Fernald, A. (2006). Children's developing ability to interpret adjective–noun combinations. In *Proceedings of the 30th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Trueswell, J. C., & Tanenhaus, M. K. (Eds.). (2005). *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.

Thorpe, K., & Fernald, A. (2006). Knowing what a novel word is not: Two-year-olds "listen through" ambiguous adjectives in fluent speech. *Cognition, 100*, 389–433.

Vendler, Z. (1963). The grammar of goodness. *Philosophical Review, 72*, 446–465.

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology, 36*, 571–581.

Waxman, S. R., & Kosowski, T. D. (1990). Nouns mark category relations: Toddlers' and preschoolers' word-learning biases. *Child Development, 61*, 1461–1473.

Waxman, S. R., & Markow, D. B. (1998). Object properties and object kind: Twenty-one-month-old infants' extension of novel adjectives. *Child Development, 69*, 1313–1329.

Zangl, R., & Fernald, A. (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language Learning & Development, 3*, 199–231.

Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development, 6*, 179–208.