Knowing a lot for one’s age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults

Arielle Borovsky a,b,*, Jeffrey L. Elman a,c, Anne Fernald b

a Center for Research in Language, University of California, San Diego, La Jolla, CA 92093, USA
b Department of Psychology, Stanford University, Stanford, CA 94305, USA
c Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093, USA

ABSTRACT

Adults can incrementally combine information from speech with astonishing speed to anticipate future words. Concurrently, a growing body of work suggests that vocabulary ability is crucially related to lexical processing skills in children. However, little is known about this relationship with predictive sentence processing in children or adults. We explore this question by comparing the degree to which an upcoming sentential theme is anticipated by combining information from a prior agent and action. 48 children, aged of 3 to 10, and 48 college-aged adults’ eye-movements were recorded as they heard a sentence (e.g., The pirate hides the treasure) in which the object referred to one of four images that included an agent-related, action-related and unrelated distractor image. Pictures were rotated so that, across all versions of the study, each picture appeared in all conditions, yielding a completely balanced within-subjects design. Adults and children quickly made use of combinatorial information available at the action to generate anticipatory looks to the target object. Speed of anticipatory fixations did not vary with age. When controlling for age, individuals with higher vocabularies were faster to look to the target than those with lower vocabulary scores. Together, these results support and extend current views of incremental processing in which adults and children make use of linguistic information to continuously update their mental representation of ongoing language.

© 2012 Elsevier Inc. All rights reserved.
Introduction

One of the challenging aspects of real-time spoken language comprehension is that language must be processed incrementally and at a relatively fast speed. The meaning of a sentence evolves as it unfolds, and sentential meaning cannot generally be inferred from any single word alone. In light of these challenges, it has been hypothesized that listeners use a strategy of continually generating expectancies about upcoming referents. Although this hypothesis remains controversial, over the past two decades a growing body of computational evidence suggests that prediction is a powerful mechanism for learning (Elman, 1990; Misyak, Christiansen, & Bruce Tomblin, 2010; Rodriguez, Wiles, & Elman, 1999), and a large number of empirical studies using behavioral and neurophysiological techniques have provided support for expectancy generation in sentence comprehension in adults.

Much less is known about the role of predictive processing during childhood. There are strong suggestions that prediction may underlie children’s ability to segment continuous streams of auditory input into word-sized chunks (Estes, Evans, Alibali, & Saffran, 2007; Saffran, Aslin, & Newport, 1996). Nor is it known what factors might influence individual children’s abilities to engage in expectancy generation, although one likely candidate is vocabulary size. For the purposes of this study, we were particularly interested in the relationship between vocabulary and predictive processing in sentences. Vocabulary size is highly associated with speed of comprehension in looking tasks (Fernald, Perfors, & Marchman, 2006; Marchman & Fernald, 2008). Moreover, there are a number of proposals in the language development literature that highlight the relationship between lexical and grammatical development. These suggest that vocabulary knowledge may serve an important role in language tasks that require meaning and structure to be parsed and interpreted across multiple words, as in online sentence comprehension. For example, early growth of the lexicon has been argued to be the critical foundation from which basic grammar emerges (Bates & Goodman, 1997; Marchman & Bates, 1994). This proposal is supported by many observations that development of the lexicon and early morphosyntactic competence is strongly correlated, with lexical development preceding grammatical development. This relationship has been noted in cross-sectional and longitudinal studies (Dale, Dionne, Eley, & Plomin, 2000; Fenson, Dale, Reznick, Bates, & Thal, 1994) and across many languages, including Icelandic (Thordardottir, Ellis Weismer, & Evans, 2002), Italian (Caselli, Casadio, & Bates, 1999), and Hebrew (Maital, Dromi, Abraham, & Bornstein, 2000), in bilingual language acquisition (Marchman, Martínez-Sussmann, & Dale, 2004), and in atypical language development (Moyle, Ellis Weismer, Evans, & Lindstrom, 2007). The idea that sentence processing may be linked to lexical knowledge is also supported by a number of grammatical theories that firmly root syntactic competence in the lexicon, including Combinatory Categorical Grammar (Steedman & Baldridge, 2006), Head-driven Phrase Structure Grammar (Pollard & Sag, 1994), and Lexical-Functional Grammar (Bresnan, 2001). Therefore, a central goal of this study was to extend our understanding of the development of anticipatory sentence processing mechanisms by directly comparing the online processing of transitive sentence comprehension in 3- to 10-year-old children with that in adults and by testing the degree to which sentence interpretation and prediction skills are associated with differences in vocabulary level.

Much of the evidence for predictive processing in sentential comprehension in adults involves inferences made from measurement of eye movements in response to language while viewing a visual scene. As objects are mentioned, visual attention is directed toward the spatial location of the object in the scene (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), often before the word is even complete (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001). More dramatically, eye gaze may be directed toward objects even before they are mentioned (Altmann & Kamide, 1999; Altmann & Mirkovic, 2009; Kamide, Altmann, & Haywood, 2003). Thus, eye gaze appears to capture moment-to-moment changes in the comprehension of language. Results from numerous studies using this technique (known as the visual world paradigm [VWP] in the adult literature), and others suggest that adults continuously update their mental representation of ongoing events in the sentence and use cues from many different sources. These include semantic features (Federmeier & Kutas, 1999), event-level expectations (Bic-
knell, Elman, Hare, McRae, & Kutas, 2010; Hald, Steenbeek-Planting, & Hagoort, 2007; Kamide et al., 2003; Metusalem, Kutas, Hare, McRae, & Elman, 2010), prosodic cues (Salverda, Dahan, & McQueen, 2003), phonological information (DeLong, Urbach, & Kutas, 2005; VanBerkm, Brown, Zwitserlood, Kooijman, & Hagoort, 2005), verb tense markers (Altmann & Kamide, 2007), grammatical and biological gender (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Lew-Williams & Fernald, 2010; Tanenhaus, Magnuson, Dahan, & Chambers, 2000; Daha, Moreno, & Kutas, 2004), verb selectional restrictions (Altmann & Kamide, 1999), pronominal adjectives (Sedivy, Tanenhaus, Chambers, & Carlson, 1999), verb structural biases (Trueswell, Tanenhaus, & Kello, 1993), and referential restrictions (Tanenhaus et al., 1995). Crucially, this updating mechanism is proactive. It not only “reacts” and “integrates” to information it receives but also actively generates expectations for plausible sentence continuations.

A concrete example of this process is illustrated in a series of studies reported by Kamide et al. (2003), which yielded some of the earliest and most convincing support for active prediction during adult sentential comprehension. In their Study 2, participants viewed simple visual scenes containing multiple agents and objects with similar affordances, including drinkable objects such as BEER and MILK. Participants then heard sentences such as “The man will drink the beer” while their eye movements to this scene were simultaneously recorded. After hearing the verb drink but before beer, participants fixated on the image of the BEER more than the image of the MILK, indicating that their eye movements reflected more than a simple lexical association between the verb drink and the upcoming object; rather, it reflected predictions motivated by a combination of the prior agent and verb. Appropriate controls also established that the effect was not the result of lexical associations between man and beer or due to visual saliency of BEER over MILK.

This result demonstrates that expectancy generation may reflect integration from multiple cues. In this case, those cues were the agent and the verb in combination. Moreover, the integration draws on real-world knowledge regarding which objects are likely patients of a verb given the agent carrying the action denoted by the verb (for related findings, see also Bicknell et al., 2010; Hare, Jones, Thomson, Kelly, & McRae, 2009; Matsuki et al., 2011).

Children also have been found to be able to integrate information from multiple sources to comprehend sentences in real time. However, children are not always sensitive to the same cues as adults. For example, children under 5 or 6 years of age seem to show difficulty in interpreting referential or extralinguistic cues to meaning in a visual scene (Kidd & Bavin, 2005; Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill, & Logrip, 1999). Despite these limitations, children still do use many sources of information with great skill and speed. Both young children and adults can call on their understanding of the syntactic and semantic meanings of verbs (Altmann & Kamide, 1999; Fernald, Zangl, Portillo, & Marchman, 2008; Nation, Marshall, & Altmann, 2003), adjectives (Fernald, Thorpe, & Marchman, 2010), and article grammatical gender (Lew-Williams & Fernald, 2010) to correctly predict an upcoming referent before it is spoken.

For example, Fernald and colleagues (2008) have shown that children as young as 26 months are able to generate anticipatory looks to an object (e.g., cookie) on hearing a verb “Eat the...” much like adults (Altmann & Kamide, 1999). Moreover, electrophysiological and neuroimaging data suggest that the neural mechanisms underlying predictive skills such as noting incongruities in label–object pairings are present by the second year of life (Friedrich & Friederici, 2004, 2005; Mills, Coffey-Corina, & Neville, 1997; Travis et al., 2011). Taken together, these discoveries suggest that the skills adults use to swiftly and efficiently understand rapidly spoken sentences are in place from the earliest moments of language acquisition.

As was noted above, it seems that such integration and prediction skills are both a marker and predictor of language skill and growth, especially in relation to vocabulary. Vocabulary skill has an important relation to processing speed, as children with larger vocabularies are also quicker to comprehend spoken words. Accuracy and speed in online lexical comprehension at 25 months of age correlates with earlier measures of vocabulary growth between 12 and 24 months (Fernald et al., 2006). Even more striking, these early differences in the speed of lexical processing at 25 months have long-term consequences for language acquisition and are correlated with language outcomes 6 years later (Marchman & Fernald, 2008). Differences in language ability are also associated with more complex measures of predictive ability. For example, the speed with which 26-month-olds correctly predict
the upcoming object of a verb in a sentence such as “Eat the cookie” is also associated with concurrent vocabulary size (Fernald et al., 2008).

It is well established that dramatic differences in vocabulary size between individuals can be observed from across the earliest stages of learning during infancy through adulthood (Fenson et al., 1994; Verhaeghen, 2003). Children typically begin to produce their first words at around the end of their first year, and vocabulary expands rapidly throughout childhood. However, the timeline and trajectory of this growth varies considerably (Fenson et al., 1994). Whereas some children may speak their first words at 12 months and may know as many as 550 words by 24 months, others might not begin to speak for another half year and might know only 50 words by 2 years (Fenson et al., 1994). It should be emphasized that children at both ends of this lexical learning spectrum are still considered within the “normal” range despite this tremendous initial variability.

These differences become even more sizable by the time children begin school. For instance, some preliterate 5- or 6-year-olds in first grade will have an expressive vocabulary of as many as 5000 words, whereas others may produce only half as many words (Beck & McKeown, 1991). Across the school years, the average student will learn thousands of new words per year, amounting to several new words a day (Anglin, 1993; Graves, 1986). However, vocabulary growth in school is considerably slower for children who begin with smaller vocabularies (White, Graves, & Slater, 1990).

Finally, vocabulary growth does not stop after childhood, and vocabulary size differences continue throughout the lifespan (Verhaeghen, 2003). For instance, differences in vocabulary between individuals who have attended college and those who have not can be significant. One study estimated the vocabulary knowledge between adults with a high school education and those with a college degree to differ by 5000 word families, which is a measurement of word knowledge that counts a root word plus its inflected forms and derivations as a single family (Zechmeister, Chronis, Cull, D’Anna, & Healy, 1995). Variability in vocabulary size is notable even in college students—a population that has been at least partially selected by performance on tests of linguistic proficiency in entrance exams (Martino & Hoffman, 2002).

Differences in early vocabulary size have important impacts on later language outcomes and school achievement. Early vocabulary growth not only is a building block for acquiring grammatical skills (Bates & Goodman, 1999; Marchman & Bates, 1994) but also is at least partially related to language and reaching achievement outcomes many years later (Cunningham & Stanovich, 1997; Marchman & Fernald, 2008). Vocabulary size has also been shown to associate with a number of other linguistic and cognitive abilities in school-age children, including phonological working memory (Gathercole & Baddeley, 1989), phonological awareness (Metsala, 1999), and reading comprehension (Cunningham & Stanovich, 1997; Stahl & Nagy, 2006). A particularly strong relationship between vocabulary level and reading comprehension has been noted since the 1920s (Whipple, 1925; see Nation, 2009, for a review). However, early vocabulary size alone is not enough to reliably detect which individuals will be most at risk for later language and learning disabilities, as noted by numerous investigators who have tried, and largely failed, to accurately identify children who will receive later diagnoses of language impairments solely from infant measures of vocabulary (Ellis Weismer, 2007; Paul, 1996; Rescorla, 2000; Whitehurst & Fischel, 1994). In sum, vocabulary and linguistic processing speed seem to have an important relationship during early childhood, although it is likely that the relationship between the two and later outcomes may be more complex.

Taken together, these findings motivate a need for a deeper and more detailed understanding of children’s predictive abilities in language comprehension and for a better understanding of the relationship of this ability to other markers of language skill such as vocabulary. The goal of the current study was to address these gaps by asking three specific questions. First, we asked whether the results of prior studies that find evidence for linguistic prediction across one or two words scale up to more complex language processing tasks that use more challenging stimuli. Advancing our understanding of how words are understood in sentences that require children to calculate more complex multiword contingencies is necessary if we are to fully understand how humans can swiftly and efficiently understand complex and novel multiword utterances, which is arguably one of the defining characteristics of language (Hockett, 1960). Second, to better understand the developmental trajectory of this skill,
we examined the development of this skill across a wide range of ages. In the current study, we included participants from 3–10 years of age as well as adults. Finally, because vocabulary appears to be an important marker of processing speed in young children, we asked whether this relationship also exists in performance on this experimental task.

To answer these questions, we built on the task used by Kamide and colleagues (2003) described above. Several modifications were made to this task, and extensive norming was conducted to ensure that both adults and children could comprehend the images and sentences. One important adaptation involved the layout of the visual display. Rather than provide visual scenes that included the sentential agents as was done in the original study by Kamide and colleagues, we used a four-alternative forced-choice display in which participants were asked to indicate which image corresponded to the sentence-final target. The purpose of this modification was twofold. First, it served to reduce the visual complexity of the visual scenes, thereby increasing the likelihood that we would get clean and motivated looks to the target items from the children. Second, it allowed us to compare the relative activation and anticipation of candidate target meanings in response to the agent and action as the sentence unfolds by controlling the relationship of the distractor items to words in the sentence.

Each visual scene consisted of the target and three types of distractors: an object that was associated with the agent (agent-related), an object that was associated with the action (action-related), and an object that was unassociated with either the agent or action (unrelated). Because the target was by definition associated with both the agent and action, this meant that two objects were associated with the action (target and agent-related pictures) and two objects were associated with the agent (target and action-related pictures). Only the target was appropriate given both the agent and action.

This design was chosen because it also allowed us to study in some detail the potential comprehension strategies that children and adults might employ as they interpreted the sentence in real time. One type of possible strategy would involve a staged elimination of potential sentential targets as the sentence unfolds. Consider the sample sentence “The pirate hides the treasure,” with a visual display consisting of TREASURE (target), a (PIRATE) SHIP (agent-related distractor), BONE (action-related distractor), and CAT (unrelated distractor). After hearing pirate, we might expect looks to the two objects that are associated with the agent (SHIP and TREASURE). When the action (hides) is heard, this makes one member of the initial cohort (SHIP) unlikely, and one might expect subsequent looks only to the target (TREASURE). This approach is analogous to the strategy proposed by the Cohort model of word recognition (Marslen-Wilson, 1987), in which the beginning of a word activates all words that are consistent with that beginning; as the rest of the word is spoken, subsequent cues eliminate inconsistent members of the cohort until only one word—the correct word—remains active.

An alternative strategy is suggested by the TRACE and Merge/Shortlist models of speech perception (McClelland & Elman, 1986; Norris, 1994). In those models, subsequent cues may both eliminate members of the initial cohort and activate new words that are consistent with later cues—even if those new words are inconsistent with the initial sound of the word. Thus, as the beginning of a target word such as beaker is heard, listeners might initially activate not only beaker but also words with similar onsets such as beetle and bee. When the second syllable is heard, this might eliminate beetle and bee. But TRACE also predicts that new words that are consistent with that second syllable, such as speaker, will become active. The strategy makes it possible to recover from initial errors in perception or production and, in fact, seems to provide a better fit for the empirical data (Allopenna et al., 1998). In the case of transitive sentence processing, after hearing the agent and action in “The pirate hides…” example, we might then expect fixations not only to the target object (TREASURE) but also to the item that is locally consistent with the action (BONE). There is increasing evidence that adults’ sentential comprehension may be at least partially influenced by local coherence effects analogous to those predicted by TRACE for word recognition (Kukona, Fang, Aicher, Chen, & Magnuson, 2011; Kuperberg, 2007; Tabor, Galantucci, & Richardson, 2004). The extent to which such local strategies are applicable sentence processing is debated, and whether children use different strategies than adults is unknown.
Method

Adult participants

A total of 48 native English-speaking college students (30 women and 18 men) between 18 and 28 years of age ($M = 21.4$ years) took part in this study in return for course credit. An additional 12 participants took part and received credit but were excluded from analysis: 11 for significant exposure to other languages during childhood and 1 for receiving prior speech therapy. Participants had normal or corrected-to-normal vision and normal hearing, and they reported no history of diagnosis or treatment for cognitive, attentional, speech, or language issues.

Child participants

A total of 48 monolingual English-learning children (23 girls and 25 boys) between 3 and 10 years of age ($M = 6.2$ years) were recruited from families in the surrounding metropolitan region of San Diego, California, in the southwestern United States. Children either had previously participated in child language research or had answered flyers and ads posted in the community. An additional 5 children also participated but were excluded from analysis: 2 for inattentiveness during testing (failing to look or respond on a majority of trials), 2 for receiving speech therapy, and 1 for being not typically developing (diagnosis of PDD–NOS [pervasive developmental disorder–not otherwise specified]). Children received a toy in return for their participation, and their families also received $10$ compensation for time and travel. For children who were included in analysis, their parents reported them to have normal hearing and vision and to be primarily hearing English at home. Parental report also indicated that all children were typically developing, without significant birth histories and with no recent or chronic ear infections and no diagnosis or treatment for other language, speech, motor, or cognitive issues.

Stimuli

Eight sets of image/sentence quartets were selected for the study. Sentence quartets were constructed by crossing two agents with two actions to create four sentences following the design principles described in the Introduction. Images corresponded to the sentential themes in the quartet (see Fig. 1 for an illustration). For example, one quartet crossed the agents (pirate and dog) with actions (hides and chases) to result in the following four sentences:

1. The pirate hides the treasure.
2. The pirate chases the ship.
3. The dog hides the bone.
4. The dog chases the cat.

The target object images for these four sentences would be TREASURE, SHIP, BONE, and CAT, respectively, and participants would see all four images concurrently. For any individual sentence, each image would correspond to one of the four image conditions (target, agent-related, action-related, or unrelated). In a sentence such as “The pirate hides the treasure,” TREASURE would be the target image, SHIP would be the agent-related image, BONE would be the action-related image, and CAT would be the unrelated distractor. Across each quartet, each image would appear in each condition once, thereby yielding a completely balanced design (see Fig. 1 for further illustration). Thus, each word and each image serves as its own control across multiple lists in which the role of the word and object changes. In this way, intrinsic differences between salience or attractiveness of words and objects were exactly balanced across lists. Visual images were selected to be typical exemplars of target items and were photo-realistic images. Images were then edited to fit within a $400 \times 400$-pixel square, and objects were placed on a light background.
Prior to the study, we conducted two norming experiments to evaluate whether our stimuli would be familiar to young children. First, to ensure that the images and their labels were known and recognizable, we asked 3- and 4-year-olds (who did not participate in the final experiment) at preschools to select a named image when it appeared along with the other images in each quartet. The images were recognized with a very high level of accuracy (>95%) on this task. Second, we normed these sentence/image pairs to make sure that children would have the requisite world knowledge to pick the correct target if given only the agent and action in an offline picture selection task. Preschoolers (3–4 years of age) were asked to judge which would be the likely target picture in each image quartet when given only the agent and action such as “Which one would a pirate chase?” or “Which one does a dog hide?” Performance on this task was also very high (90%). Quartets were not selected for the eye-tracking study if any single image in the quartet was not appropriately selected by at least 70% of preschoolers tested. We began the norming process with 12 sentence/image quartets, but 4 did not meet our norming criteria for final selection in the study, yielding 8 quartets.

The auditory stimuli were recorded by a female native English speaker (A.B.) in a child-directed voice and sampled at 44,100 Hz on a single channel. Each sentence was edited using Praat audio

Fig. 1. Illustration of stimuli and conditions in the experimental task.
editing software (Boersma, 2001). First, the onsets of each word (article 1 [Art1], agent, action, article 2 [Art2], and target) were marked and the word duration was normalized to the mean duration of all words in the same sentential position. In cases where actions included a particle such as jump into, the action duration was marked to include the duration of the verb and particle. The mean intensity of all sentences was then normalized to 70 dB. This arrangement yielded 32 sentences of equal length, in which the onset of each word was standardized across all sentences. This was done to ensure that participants were given the same amount of time in each sentence to use the agent and action information to anticipate the target. The sentential word durations were as follows: Art1, 134 ms; agent, 768 ms; action, 626 ms; Art2, 141 ms; target, 630 ms. The final experimental sentences were then judged by several listeners who were unfamiliar with the goals of the study to sound natural. The sound files were presented to participants via headphones so that the same stimulus was presented to each ear. The experimental sentences are listed in the Appendix.

In any one version of the study, participants heard 16 of 32 sentences. Each quartet of four objects was seen twice, with two of the four possible sentences for each set being presented. Across all versions the position of each object was presented with equal frequency in each quadrant, and in each version the target image appeared in each quadrant an equal number of times.

Procedure

Experimental task

Participants were seated in a comfortable chair in front of a 17-inch LCD display, and the eye-tracker was focused and calibrated. Stimuli were presented using a PC computer running EyeLink Experiment Builder software (SR Research, Mississauga, Ontario, Canada). Participants were told that they would be seeing pictures and listening to sentences, and they were told to either point to (children) or click on using the mouse (adults) the picture that “goes with the sentence.” It was expected that participants would point to or click on the sentence-final object, which was the target of the action, and indeed this is what participants did. They were given one practice trial before starting the study. Before the start of the experimental trials, a manual 5-point calibration and validation routine was performed using a standard black-and-white 20-point bull’s-eye image. Before each trial, participants were shown the same centrally located bull’s-eye image, which they were instructed to fixate on before starting the trial. This dot also served as a drift correction dot before each trial. Once participants had fixated on this location, the experimenter began the trial.

Participants were shown the set of four images for 2000 ms before sentence onset, and the images remained on the screen after sentence offset until participants had selected an image from the array with the mouse or by pointing. Recalibration of the eye-tracker was performed between trials if necessary, although this was rarely needed. Participants were given a break halfway through the study. This portion of the experiment lasted 5–10 min.

Eye movement recording

Eye movements were recorded using an EyeLink 2000 remote eye-tracker with remote arm configuration (SR Research) at 500 Hz. Because the physical size of our adult and child participants varied considerably, we individually adjusted the position of the display using a remote arm configuration of the eye-tracker such that the display and eye-tracking camera were placed 580–620 mm from each participant’s face. Head and eye movements were automatically tracked by the eye-tracker system with the use of a target sticker affixed to the participant’s forehead. This arrangement allowed for stable tracking even during some movement and shifting of the participant’s position relative to the camera and display.

Fixations were recorded in each trial from the onset of the images until the participant or experimenter clicked on or pointed to the selected picture. The recorded eye movements were automatically classified as saccades, fixations, and blinks using the eye-tracker’s default threshold setting. Offline, the data were binned into 10-ms intervals, over which subsequent analyses were performed.

Offline measurements

After the completion of the eye-tracking task, participants were administered two offline language measures: the Peabody Picture Vocabulary Test–Version 4 (PPVT-4; Dunn & Dunn, 2007) and the
Sentence Completion subtest of the Comprehensive Assessment of Spoken Language (CASL–SC; Carrow-Woolfolk, 1999). The PPVT-4 is a norm-referenced test designed to estimate receptive vocabulary for standard American English in participants between 2 and 90 years of age. The test contains 220 items arranged in order of increasing difficulty with age-appropriate starting points. Each item has four colorful illustrations arranged in a multiple-choice format. The participant’s task is to select the picture considered to correspond to the best meaning of a stimulus word presented orally by the examiner. The CASL–SC test contains 60 items of increasing difficulty with various age-appropriate levels. In each item, the examiner reads a sentence, leaving off the final word. The participant’s task is to provide any semantically and syntactically appropriate word to fit the context.

Results

Behavioral accuracy

To ensure that adults and children understood the sentences and task, we first examined their accuracy to select the correct target picture. Accuracy was very high on the task in both groups, and very few errors were made in selecting the correct target picture. There were 4 incorrect responses across adults (99.5% correct), and there were 20 incorrect responses across children (97.4% correct). All analyses below were conducted, and all results are reported, with these incorrect responses removed.

Time course measurements

To examine the time course of incremental sentence processing in adult and child groups, we calculated the mean proportion of time spent fixating to the target, agent-related, action-related, and unrelated images at each 10-ms time window across all adult participants (Fig. 2A) and children (Fig. 2B). Furthermore, to inspect the data for potential developmental and vocabulary-related differences in processing, we split the child participants into older (6 years 5 months [6;5]–10;0) and younger (3;0–6;3) age groups (Fig. 3) and divided adults and children into high and low vocabulary groups according to a median split of age-normalized PPVT (Fig. 4) and age-normalized CASL–SC scores (Fig. 5). Relevant demographics and characteristics of children in each median split group are reported in Table 1.

These time course plots illustrate several interesting fixation patterns. Most prominently visible is the robust acceleration of (anticipatory) fixations to the target object that begins as the verb is spoken. In addition, looks to the agent-related and action-related objects also increase after the agent and action onset, respectively. Notably, the timing at which fixations to the target object diverge from looks to other items, especially the agent-related item, visibly varies between age and vocabulary groups.

![Fig. 2. Time course of fixating to target and competitor’s interest areas (with SE bars) during the sentence in all adults (A) and all children (B) plotted in 10-ms time bins.](image-url)
Time course of looks to target

We initially characterized the timing of these apparent differences in fixation timing between groups by conducting point-by-point \( t \) tests at each 10-ms bin between mean fixation proportions to the target object and fixations to the agent-related item because this was the second most highly fixated object in each group. To minimize the possibility that differences measured by these multiple \( t \) test comparisons might have arisen by chance, we report the earliest time at which a minimum of
five subsequent and consecutive one-tailed $t$ tests with an alpha level of $p < .05$ indicate a significant positive difference between fixations to the target and other items.

According to these analyses, the time at which fixations to the target diverge varies considerably between adults (1220 ms) and children (1420 ms). Although these differences are smaller between older and younger children (1410 and 1460 ms, respectively). Adults with higher and lower (age-normalized) vocabulary scores show some difference in target divergence times (1210 and 1300 ms, respectively), and these values are very similar for high and low sentence completion groups (1220 and 1300 ms, respectively). The target divergence times are quite marked between children with higher
(age-normalized) vocabularies (1220 ms) and those with lower (age-normalized) vocabularies (1520 ms). Relatively smaller but still sizable differences in target divergence were also found on the sentence completion measure in children (high: 1320 ms; low: 1510 ms).

This preliminary inspection and description of the time course between groups indicated that there were differences in the timing of anticipatory looking to the target. Because these differences are revealed by a measure of the overall proportion of time spent fixating on the object, there are a number of ways in which these timing differences might have arisen. For one, it is possible that they are due to a difference in the speed in initial looking to the target object. Alternatively, the mean duration or number of individual fixations might vary. In the next set of analyses, we investigated these various possibilities.

**Timing of initial looks**

To examine whether the observed timing differences arose from differences in the speed at which each group initially looked to the target, we measured the latency of the initial saccades that landed on the target object after the action onset. Children initially fixated to the target more rapidly than adults, \( t(94) = 2.057, p = .04 \). Older children were significantly faster than younger children to fixate on the target, \( t(46) = 2.28, p = .03 \), as were individuals with higher age-normalized vocabularies, \( t(94) = 2.24, p = .03 \), whereas individuals with higher sentence completion scores were marginally faster to fixate on the target, \( t(94) = 1.80, p = .07 \).

**Duration of looking**

Next, we asked whether group differences in looking to the target might be driven by variation in the average individual fixation time. This was measured by calculating the total time spent fixating on the target during the period from action word onset to target word offset and dividing this by the total number of fixations generated to the target image during this period. There were no group differences in any age, vocabulary, or sentence completion comparison examined with this measure.

**Number of fixations**

Finally, we asked whether the total number of fixations to the target varied between groups by measuring the overall number of fixations to the target from action word onset to target offset. Children with higher vocabulary scores, \( t(46) = 3.298, p = .0019 \), and sentence completion scores, \( t(46) = 3.350, p = .0016 \), made a greater number of fixations to the target object than their lower scoring peers. There were no other group effects for adults or across age groups.

**Associations among anticipatory fixations, age, and measures of linguistic ability**

The above analyses explore group differences across measures that capture looking behaviors across time windows that precede and follow the onset of the target word, that is, using eye movement measures that are not solely anticipatory. Our primary interest, however, was to understand the interrelationships among patterns of anticipatory fixations, age, and linguistic ability. To investigate this question, we operationally defined anticipatory looking as the mean proportion of time spent fixating on the target object from the onset of the action word to the target word, that is, over the portion of the sentence during which the necessary information to generate anticipatory looks was available. Because we are only interested in fixations that were generated after the onset of the action, fixations were excluded from analysis if their preceding saccade was initiated prior to action onset.

With these measurements, we first sought to determine whether participants’ fixations to the target region significantly differed from fixations to other regions during the same period. Consequently, we conducted several planned comparisons to compare the difference in fixation proportion between the target region and (a) the agent-related distractor, (b) the action-related distractor, and (c) the unrelated distractor. If eye movements were driven by combinatorial integration of meanings from both the sentential agent and action, then we should expect to see that differences between looks to the target and any other region should be significantly greater than zero. That is, if participants are indeed anticipating the upcoming target word before it is spoken, then we would expect participants to fixate to the target image with greater magnitude than other objects before the target itself is
spoken. This is in fact what we found in all but two groups: children with lower sentence completion scores and children with lower vocabulary scores. These comparisons are summarized in Table 2.

Table 3 shows associations between online measures of mean proportion of anticipatory fixations to the target object, age, and offline measures of language ability. Even though vocabulary and sentence completion were highly intercorrelated, \( r(96) = .83, p < .0001 \), only vocabulary was significantly associated with the online measure of anticipatory fixation, \( r(96) = .27, p = .0086 \).

To further explore the relationship between the experimental and offline measures, we entered age, vocabulary, and sentence completion measures into a multiple regression model. Again, the only variable to explain significant variance in this model was vocabulary, \( r(96) = .33, p < .001 \). Neither age nor sentence completion scores explained significant variance in our experimental measure when they were entered into the model singly or in combination with other factors. Because the sentence completion measure was highly intercorrelated with the vocabulary measure and it failed to account for significant variance in task performance, we dropped this factor from subsequent analyses.

We further investigated the possibility that the relationship between vocabulary and sentence processing might potentially interact with age by carrying out an additional (2 \times 2) analysis of variance (ANOVA) with factors of age (adult or child) and a median split of age-normalized vocabulary scores (high or low). This analysis found a significant effect of vocabulary, \( F(1,95) = 11.23, p = .0012 \), but not of age or Age \times Vocabulary (Fs < 1). To rule out that the general lack of age effects might be driven by the inclusion of the adult participants in our analyses, we then conducted an identical ANOVA with only the children included. In this case, two age groups were determined by a median split of older and younger children (younger range = 3:0–6;3; older range = 6:5–10:0). Again, there was a main

### Table 2

T-scores for difference between fixation proportions in anticipatory time window to target image minus distractors.

<table>
<thead>
<tr>
<th></th>
<th>Target – agent-related</th>
<th>Target – action-related</th>
<th>Target – unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td>All adults</td>
<td>7.14***</td>
<td>8.34***</td>
<td>8.74***</td>
</tr>
<tr>
<td>High vocabulary adults</td>
<td>6.44***</td>
<td>7.38***</td>
<td>6.66***</td>
</tr>
<tr>
<td>Low vocabulary adults</td>
<td>3.81**</td>
<td>4.68***</td>
<td>6.00***</td>
</tr>
<tr>
<td>High sentence completion adults</td>
<td>5.64***</td>
<td>6.53***</td>
<td>5.87***</td>
</tr>
<tr>
<td>Low sentence completion adults</td>
<td>4.55***</td>
<td>5.26***</td>
<td>7.05***</td>
</tr>
<tr>
<td>All children</td>
<td>2.50**</td>
<td>7.5***</td>
<td>8.3***</td>
</tr>
<tr>
<td>Older children</td>
<td>1.90**</td>
<td>6.72***</td>
<td>7.05***</td>
</tr>
<tr>
<td>Younger children</td>
<td>1.59*</td>
<td>4.24*</td>
<td>4.85***</td>
</tr>
<tr>
<td>High vocabulary children</td>
<td>5.00**</td>
<td>9.80***</td>
<td>6.90***</td>
</tr>
<tr>
<td>Low vocabulary children</td>
<td>–.552</td>
<td>3.06*</td>
<td>5.05***</td>
</tr>
<tr>
<td>High sentence completion children</td>
<td>4.05**</td>
<td>9.99***</td>
<td>6.81***</td>
</tr>
<tr>
<td>Low sentence completion children</td>
<td>.190</td>
<td>3.09**</td>
<td>5.01***</td>
</tr>
</tbody>
</table>

* \( p < .01 \).
** \( p < .001 \).
*** \( p < .0001 \).
# \( p = .062 \).

### Table 3

Pearson correlation coefficients (\( r \)) between online measures of mean proportion of anticipatory fixations to target object, age, and offline measures of language ability.

<table>
<thead>
<tr>
<th></th>
<th>Anticipatory fixations</th>
<th>Vocabulary</th>
<th>Sentence completion</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anticipatory fixations</td>
<td>–</td>
<td>.27*</td>
<td>.17</td>
<td>.012</td>
</tr>
<tr>
<td>Vocabulary</td>
<td></td>
<td>–</td>
<td>.83**</td>
<td>.004</td>
</tr>
<tr>
<td>Sentence completion</td>
<td>.27*</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Age</td>
<td>.17</td>
<td>.83**</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Note. \( N = 96 \) (48 children and 48 adults).

* \( p < .01 \).
** \( p < .0001 \).
effect of vocabulary, \( F(1,47) = 6.84, p < .01 \), but not of age or Age \( \times \) Vocabulary (\( Fs < 1 \)). This same pattern also holds when age and vocabulary are entered as continuous factors into a multiple regression model with children only: vocabulary, \( F(1,47) = 9.35, p = .0038 \), age and Age \( \times \) Vocabulary (\( Fs < 1 \)). Similarly, a median split of adults alone into higher and lower scoring vocabulary groups yields significant effects of vocabulary both for median splits of age-normalized vocabulary scores, \( F(1,47) = 3.99, p = .05 \), and for raw PPVT, \( F(1,47) = 5.99, p = .02 \). As continuous variables, adult raw PPVT scores significantly associate with anticipatory fixations, \( r(47) = .30, p = .04 \), but age-normalized PPVT scores do not.

### Time course of fixations to action-related item

The time course in Figs. 2–5 suggests that participants in all groups generated additional looks to the action-related distractor at varying time points after hearing the verb. To quantify the timing of when these fixations diverged from that of looks to the distractor items, we conducted point-by-point \( t \) tests to determine the time window where the mean fixation proportions to the action-related distractor exceeded those to the unrelated distractor. As above, only time windows at which a minimum of five consecutive time points reached significance on this \( t \) test measure are reported. This analysis revealed that children and adults in all examined groups fixated on the action-related item in greater magnitude than on the unrelated item, as reported in Table 4.

### Discussion

In this investigation, we measured the relationships among anticipatory fixations during a simple spoken sentence comprehension task, age, and linguistic ability. To our knowledge, this study is the first to show an association between vocabulary knowledge and incremental sentence interpretation in both adults and children. The study also replicates and extends a number of prior findings in the incremental sentence processing literature. As reported in previous studies (Altmann & Kamide, 1999; Kamide et al., 2003), adults in this study launched anticipatory fixations to the target item starting at the action prior to the target. The timing with which these eye movements diverged from looks to other object regions was rapid in adults (1210 ms)—approximately 300 ms after the onset of the action. Given that it is generally considered to take approximately 200 ms to launch a saccade (cf. Altmann, 2011; Haith, 1993), the speed at which these fixations diverged indicates that adults were integrating information from the combined agent and action with remarkable speed.

We began by asking three questions. The first was whether the predictive behaviors found in children in simple looking tasks scale up to more complex language processing in which listeners must integrate cues over multiple words. The answer is an unequivocal yes. Curiously, the anticipatory looking behaviors seem to be relatively larger than those seen by Kamide and colleagues (2003). We believe that much of this difference simply reflects methodological differences in the visual display.

### Table 4

Summary of fixation divergence times for target and action-related items.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Target vs. agent-related (ms)</th>
<th>Action-related vs. unrelated (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All adults</td>
<td>1220</td>
<td>1430–1490, 1520–2240</td>
</tr>
<tr>
<td>All children</td>
<td>1420</td>
<td>1480–2370</td>
</tr>
<tr>
<td>Older children</td>
<td>1410</td>
<td>1530–2400</td>
</tr>
<tr>
<td>Younger children</td>
<td>1460</td>
<td>1440–2280</td>
</tr>
<tr>
<td>High vocabulary adults</td>
<td>1210</td>
<td>1650–1990</td>
</tr>
<tr>
<td>Low vocabulary adults</td>
<td>1310</td>
<td>1310–1470, 1510–2250</td>
</tr>
<tr>
<td>High vocabulary children</td>
<td>1220</td>
<td>1510–2350</td>
</tr>
<tr>
<td>Low vocabulary children</td>
<td>1520</td>
<td>1570–2290</td>
</tr>
<tr>
<td>High sentence completion adults</td>
<td>1220</td>
<td>1510–2050, 2150–2230</td>
</tr>
<tr>
<td>Low sentence completion adults</td>
<td>1300</td>
<td>1620–2170</td>
</tr>
<tr>
<td>High sentence completion children</td>
<td>1320</td>
<td>1490–2340</td>
</tr>
<tr>
<td>Low sentence completion children</td>
<td>1520</td>
<td>1570–2290</td>
</tr>
</tbody>
</table>
and response demands between these two studies. Kamide and colleagues used a more naturalistic visual scene that included representations of both potential agents, whereas we presented a simpler four-choice array in order to reduce the distractibility of the scene for our younger participants. In addition, whereas we asked our participants to select the item that represented the sentential object, Kamide and colleagues simply asked participants to attend to the visual display and sentences. As a result, we measured numerically larger fixation proportions to all items in the experiment, including the distractor. Even with these differences, the pattern of findings remains similar across the two studies. It is clear that these methodological adaptations were successful in our younger participants. The children understood and easily performed the task given that they made few response errors and generated anticipatory eye movements with speeds that rivaled those of the adult participants. Yet, it should be noted that some minor differences between children and adults did exist. The time course of looking to the agent-related and target items early in the sentence did appear to be slightly different between the child and adult groups. Because every item in our experimental design was completely balanced across conditions, it is unlikely that this was due to visual properties of our stimulus. However, it may represent a potential age-related strategy to look back at the agent-related item before selecting the target.

Our second question was how and whether such abilities vary across childhood. The (nonanticipatory) looking behaviors noted in Fernald, Pinto, Swingley, Weinberg, and McRoberts (1998) showed clear increases in speed between 15 and 24 months of age. We expected similar increases in this study. Although the children were older (3–10 years), the task was considerably more challenging. However, here we were surprised to find that age did not predict performance considering children either alone or in comparison with adults. Strikingly, the children with the highest vocabulary levels performed nearly as quickly as adults with high vocabulary scores. This pattern is notable because high vocabulary children were younger (average age = 6;5, range = 3;0–9;2) and had lower raw PPVT scores on this task (mean PPVT score = 160, range = 113–206) than their college student counterparts (mean PPVT score = 213, range = 208–219), who are at least partially selected by having the requisite language ability to gain admission to the university. One potential caveat of this finding, however, concerns the nature of our stimuli. Because these sentences were spoken using child-directed prosody, there is some possibility that adults, who are less familiar with this type of speech than children, may have been slowed to some extent on this task. Nonetheless, even when considering only child performance, the effects of age were not as notable as those of vocabulary skill (corrected for age).

These results address our third question, which was what role one measure of linguistic ability—vocabulary skill—might play in predicting anticipatory looking skills. Vocabulary has been shown to strongly predict nonanticipatory looking in other tasks with very young children (Hurtado, Marchman, & Fernald, 2008). Similarly, other measures of linguistic ability, such as reading comprehension in school-aged children (Nation et al., 2003) and self-rating of second language proficiency in college-aged adults (Chambers & Cooke, 2009), also seem to correlate with some aspects of predictive processing. In our tasks, we observed a complementary finding, where the participant’s age-normalized score on a test of receptive vocabulary is a better predictor of anticipatory looking than the age of the participant. That is, having a large vocabulary for one’s age is important in this task. This result complements the earlier findings regarding vocabulary size per se and contributes to the appreciation that the various relationships among age, vocabulary, and speed of linguistic processing are somewhat complex, with the potential for multiple mediating factors. Because we failed to find age-related changes in processing on our task, our result is consistent with the interpretation that processing speed itself may contribute independently to vocabulary learning rather than the other way around (Hurtado et al., 2008). The relationship between vocabulary and online sentence comprehension is a topic that merits further study. It is also noteworthy that another measure of language ability, sentence completion, was not associated with performance on this task even though sentence completion is itself highly correlated with vocabulary. This finding emphasizes the need to more finely determine what processing subskills and specific linguistic knowledge support predictive processing in sentence comprehension.

Successful anticipatory performance on our task required not only an awareness of the semantic fit or association between the verb and the correct upcoming target but also a more complex calculation
of higher order contingencies among the agent, action, and patient. For instance, in the sentence “The pirate chases the ship,” the associated visual display contains two objects that are potentially chaseable (SHIP and CAT) but only one that is likely to be chased by the pirate. The definition of likelihood here specifically invokes knowledge of real-world events and situations. The current data indicate that, as seems true for adults (Bicknell et al., 2010; Hald et al., 2007; Kamide et al., 2003; Matsuki, McRae, Hare, & Elman, 2008; Matsuki et al., 2011; Metusalem et al., 2012), 3- to 10-year-olds are able to rapidly integrate multiple cues during sentence processing and use real-world knowledge of events and situations to guide their expectancies regarding upcoming referents.

Finally, we comment on the strategy that listeners appeared to be using during this integrative process. The sentences in our design began with a cue (the agent) that potentially activated two visual items: the target and the agent-related distractor. As we noted above, one strategy, consistent with the Cohort model (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978), might be for listeners to use the subsequent cue (the action) to eliminate the agent-related distractor from this initial cohort, leaving active only the cohort. In this case, looks to the target should increase at mention of the action (which they did), but looks to all other distractors should decrease, including the initially activated agent distractor.

In fact, what appeared to happen is that listeners used a strategy more consistent with the TRACE model (McClelland & Elman, 1986). In TRACE, early cues activate an initial cohort of items that may be pruned by subsequent cues that are inconsistent with it. However, TRACE also allows subsequent cues to activate new items that are locally consistent with the new cues even if they are inconsistent with the earlier cues. This might seem like a nonoptimal strategy—why entertain hypotheses that one knows are not likely to be correct?—but it does allow for recovery under imperfect listening conditions in which the initial word onset was not processed correctly.

Indeed, this second strategy appears to best correspond to what listeners actually do when they identify words (Allopenna et al., 1998) and seems to resemble what our participants did when processing sentences. Rather than only narrowing the range of targets they considered, listeners showed a willingness to predict targets that were locally consistent with the action even though these targets were not expected given the already mentioned agent. Thus, looks to the action-related target (e.g., CAT) increased at the verb when participants heard “The pirate chased...” In fact, there is increasing evidence in the adult sentence processing literature that locally coherent syntactic and semantic interpretations are temporarily activated (Kukona et al., 2011; Tabor et al., 2004). We suggest that this strategy may be advantageous when encountering unexpected outcomes during language comprehension. Rather than entirely discounting less expected possibilities, adults and children allow for the possibility that the unexpected may happen—although these less probable outcomes are weighted accordingly.

Acknowledgments

We express our sincere gratitude to the families and children throughout the San Diego region who participated in this research. We also thank Sarah Creel and Heather Pelton for assisting with two prior norming studies leading up to the research and thank Kim Sweeney and Alex Huynh for assisting with data collection. In addition, we thank the editor, an anonymous reviewer, and Gerry Altmann for their helpful comments and suggestions. This work was funded by grants received from the National Institutes of Health (HD053136 to J.E. and DC010106 to A.B.).

Appendix

List. of 32 sentences designed for the study

The fireman rides the truck.
The girl rides the bike.
The fireman tastes the hamburger.
The girl tastes the candy.
The shark swims in the ocean.
The boy swims in the pool.
The shark catches the fish.
The boy catches the ball.

The pirate hides the treasure.
The dog hides the bone.
The pirate chases the ship.
The dog chases the cat.

The baby drinks the milk.
The woman drinks the water.
The baby wears the diaper.
The woman wears the dress.

The cat catches the bird.
The frog catches the fly.
The cat jumps onto the couch.
The frog jumps into the pond.

The monkey jumps through the trees.
The dolphin jumps through the waves.
The monkey eats the banana.
The dolphin eats the fish.

The baby eats the cookie.
The cow eats the grass.
The baby sleeps in the crib.
The cow sleeps in the barn.

The boy flies the kite.
The pilot flies the airplane.
The boy wears the t-shirt.
The pilot wears the helmet.

References


