30-month-olds use the distribution and meaning of adverbs to interpret novel adjectives

Kristen Syrett\textsuperscript{a} and Jeffrey Lidz\textsuperscript{b}

\textsuperscript{a} Rutgers, The State University of New Jersey

\textsuperscript{b} University of Maryland, College Park

address correspondence to
Kristen Syrett
Rutgers University Center for Cognitive Science
Psychology Building Addition, Busch Campus
152 Frelinghuysen Rd.
Piscataway, NJ 08854
k-syrett@ruccs.rutgers.edu
ABSTRACT

Although in language acquisition the category of adjectives has frequently been treated as a uniform category, there are semantic differences among adjectives that make the word learning process much more complex. We focus on one difference within the class of *Gradable Adjectives* (GAs). We first show that robust patterns of adverbial modification in the exposure language sort GAs according to scalar structure: proportional modifiers (*completely*) tend to modify absolute maximum standard GAs (*full*), while intensifiers (*very*) tend to modify relative GAs (*big*). We then show in a word-learning experiment that 30-month-olds appear to be aware of such information and recruit it in word learning, assigning an interpretation to a novel adjective based on its modifier. We argue that this is a form of syntactic bootstrapping: children track the range of adjectives modified by an adverb and the range of adverbs modifying an adjective, and use this surface-level information to classify new words according to possible semantic representations.
1. Introduction

A central challenge of word learning, classically presented by Quine (1960), is the following: given a word produced in a discourse context, there are simply too many possible interpretations for the young word learner to hone in on one in particular without taking advantage of other sources of information. One such source of information is the syntactic context in which a word appears. (See Woodward (2000) for a review of other possible sources.) The semantic constraints of a syntactic context can help to narrow down the space of possible word meanings by highlighting the syntactic category and distribution of words appearing in that context. The syntactic context has been appealed to in many accounts of noun and verb learning (e.g., Brown, 1957; Fisher, 2002; Gleitman, 1990; Katz, Baker, Macnamara, 1974; Landau & Gleitman, 1985; Naigles, 1990), and in cross-category learning between nouns (or noun phrases) and adjectives (Booth & Waxman 2003; Hall, 1994; Waxman & Booth, 2001). In this paper, we examine the role of the syntactic context in the acquisition of subcategories of gradable adjectives, such as big. To do this, though, we must first identify these subcategories, the linguistic motivation and diagnostics for them, and the way in which syntactic distributional patterns highlight differences among them.

Although in the child language literature, adjectives are typically talked about as referring to ‘object properties’, in contrast to nouns, which are said to refer to ‘object kinds’, this differentiation does not capture the full range of adjective meanings that a child must acquire. Some adjectives can be more or less captured by this type of description; their referent can simply found in the intersection of two sets (Clark, 1970; Kamp, 1975; Montague, 1974; Parsons, 1972). That is, a ‘herbivorous mammal’ stands at the intersection of herbivorous things and things that are mammals. Other intersective adjectives include wooden, square, or six-legged
and (at least under the standard interpretation) colors such as *red*. However, not all adjectives are intersective.

*Subsective* adjectives require reference to a comparison class (cf. Bartsch & Vennemann, 1972a; Klein, 1980; Rusiecki, 1985; Siegel, 1979; Unger, 1975). For example, it is not possible to say whether something is *big* until we know about its set membership (i.e., what counts as *big* for a mouse is not what counts as *big* for a mountain). This has implications for the truth value of the sentence in which such adjectives appear: *X is big* might be *false* if *X* is a mouse, or *true* if *X* is a mountain. Subsective adjectives can appear in certain linguistic environments in which other types of adjectives cannot appear, as illustrated in (1-2).

1. a. How expensive was Boston’s Big Dig?
   b. The Big Dig was more expensive than officials had anticipated.

2. a. #How extinct/Herbivorous is the diplodocus?
   b. #The diplodocus is more extinct/Herbivorous than the mammoth.

In fact, one hallmark of subsective adjectives is their ability to be followed by a *for an X phrase*, as in (3).

3. a. That is expensive for a gallon of gas.
   b. #That dinosaur is extinct for an animal.

However, there are other adjectives (e.g., *full*) that can also appear in comparative environments, as in (4), but which do not depend on a comparison class for interpretation, as shown in (5). For

---

1 This comparison class can be provided by a variety of sources (e.g., the modified noun, the discourse context, etc.), a discussion of which we set aside here. Here, we also leave aside discussion of adjectives such as *alleged*, *former*, and *fake*, which fall into neither of these groups (i.e., an *alleged crook* is not both *alleged* and a *crook*, nor is s/he *alleged for a crook*) (cf. Heim & Kratzer, 1998; Kamp & Partee, 1995; Partee, in press).
example, you can compare degrees of a container’s fullness without reference to other containers of that sort.

4.  a.  **How full** is the pitcher?

   b.  The pitcher was **fuller than** I had anticipated (so I spilled water all over the table).

5.  ??That is full for a pitcher.

This difference between adjectives such as *herbivorous* and *extinct* on the one hand, and *big* and *full* on the other, reflects a difference encoded in their semantic representations, which cuts across subsective and intersective classifications. The former are *non-gradable adjectives*, and the truth value of the sentence in which they appear rests on whether the property can be predicated of the entity. The latter are known as *gradable adjectives*. In contrast to non-gradable adjectives, gradable adjectives (GAs) are said to measure the degree to which an entity possesses a property along a given dimension (e.g., SIZE, HEIGHT, WEIGHT, etc.). One way to think of this difference is that with GAs, the relevant question is not necessarily, **Does the object have the property?** but rather **To what extent does the object have the property?**

Previous work has demonstrated that by three years of age, young children are aware of the role of the comparison class in the interpretation of GAs such as *big* (Barner & Snedeker, 2007; Ebeling & Gelman, 1988, 1994; Gelman & Ebeling, 1989; Smith, Cooney, & McCord, 1986) and of restrictions on the role of the context in the interpretation of different kinds of GAs (Syrett, 2007; Syrett, Kennedy, & Lidz, 2006, under revision). Our focus here is on how children learn about these within-GA distinctions. More specifically, we ask whether child can recruit

---

2 It should be noted that while we refer to differences among ‘adjectives’, we are actually discussing differences among possible interpretations, or readings, of adjectives.
surface-level patterns in the speech stream to correctly classify novel GAs according to the abstract representations described above.

2. The Semantics of Gradable Adjectives

As we stated above, GAs can be easily identified by their appearance in comparative environments. They are able to do so because they are measure functions that take an object as the input and return as the output a degree on a scale, which measures out a specific dimension. Now because reference to degrees is encoded in the meaning of GAs, even in the non-comparative, positive form (i.e., big, vs. bigger, tall vs. taller, and so on), GAs are still implicitly comparative. Taking as an example the GA big and the scale in (6), let us say that we are comparing the size of entities. The standard size for what counts as big in that context is represented by (s), while (x) represents the degree to which x is big. In this scenario, the sentence x is big is true, because (x) exceeds (s). If instead the degree to which x is big were represented by (x'), then the sentence x is big is false, since (x') does not exceed (s).

6. \[ \bullet \quad (x') \quad (s) \quad (x) \]

The difference between GAs such as big and full that was noted earlier arises from differences in this scalar structure. Relative GAs such as big, which depend on a contextually-determined comparison class, map objects onto an open-ended scale. Absolute maximum standard GAs such as full, which do not depend on the context for the standard, map onto a scale

---

3 GAs in the positive form have been argued to have a null positive degree morpheme POS, responsible for encoding a relation between a degree in the representation of the GA and the degree of the standard of comparison such that the degree of an individual x stands in a ‘greater-than-or-equal-to’ relation to the standard degree s (cf. Bartsch & Vennemann, 1972b; Bogusławska, 1975; Fara, 2000; Kennedy, 2007; Kennedy & McNally, 2005; von Stechow, 1984).
that is closed at one or both ends, and it is this endpoint that serves as the standard. With full, the endpoint signals maximality of fullness (or alternatively, zero degree of emptiness).

Thus, GAs differ with respect to their scalar structure on two separate, but intimately related aspects: the presence or absence of scalar endpoint(s) and the role of the endpoint or context in determining the standard of comparison. This difference is captured in (7).\footnote{We concentrate here on the contrast between absolute maximum standard GAs (MAX-GAs), which map objects onto maximally closed scales, and relative GAs (REL-GAs), which map objects onto open scales, leaving aside absolute minimum standard GAs, such as dirty and wet. As we hint at in this discussion, absolute maximum and absolute minimum standard GAs go hand in hand, in referring to the absence or presence, of the relevant property, respectively. See Kennedy & McNally (2005) and Rotstein & Winter (2004) for further discussion.}

7. a. REL-GAs (open-ended scale) \(\rightarrow\) (e.g., big/small, old/young)  
b. MAX-GAs (scale closed on one/both ends) \(\rightarrow\) (e.g., clean/dirty; full/empty)

Returning to the challenge the word learner confronts, we can now ask, How does a child know how to map a newly encountered GA onto the right representation? We propose that one such strategy is for children to appeal to the linguistic environment in which a GA appears and take advantage of the strong correlation between surface-level form and abstract meaning.

3. Form, Meaning, and Adverbial Modification

Relevant to the argument that children could learn how to subcategorize GAs on the basis of their syntactic distributions is the observation that patterns of adverbial modification highlight differences in semantic representations among GAs. As (8) illustrates, proportional modifiers such as half, almost, and completely are able to modify some adjectives, but not others.

8. a. The glass is half/almost/completely full.  
b. #Her brother is half/almost/completely big.
We can contrast this pattern with the one in (9), which illustrates that intensifiers such as very that have a more widespread distribution.

9. a. The glass is very full.

b. Her brother is very big.

These patterns of modification fall out directly from the restrictions these adverbs place on the scalar structure of the adjectives they modify. Proportional modifiers can only modify absolute GAs that are able to supply a maximal endpoint (Cruse, 1980; Paradis, 1997; Kennedy & McNally, 2005; Rotstein & Winter, 2004), whereas intensifiers can modify any GA which allows for a relative interpretation, even if it is not relative by default.\(^5\) In fact, the rampant imprecision in language usage should ensure that intensifiers modify absolute minimum standard GAs more often than would be predicted by a strictly semantic account. Thus, patterns of modification on the surface are indicative of underlying differences in the abstract semantic representation of these lexical items.

This form-meaning correspondence has played a central role in research in language acquisition, most prominently in discussions of syntactic bootstrapping in verb learning (Gleitman, 1990; Landau & Gleitman, 1985). There is by now considerable evidence that infants and young children recruit information about sentence structure (i.e., the number and type – e.g., NP, PP, or S – of syntactic arguments and their position in the sentence with respect to the verb) to form hypotheses about the semantic representation of a verb (Fisher, 2002; Naigles, 1990;

---

\(^5\) Note that although our focus is on adjectives here, the same selectional restrictions will hold for any lexical item being modified, regardless of grammatical category, since the selectional restrictions are semantic, not syntactic. Thus, for example, no matter what almost modifies, it will require there to be an endpoint or boundary (e.g., almost full, almost ran 5 miles, almost passed the bill).
Naigles, Fowler, & Helm, 1992; see also Wagner, 2006). This mapping from syntax to semantics is also well attested in other areas of word learning where young children demonstrate an awareness of the role of lower-level morphosyntactic information (e.g., presence and type of determiner, affixation encoding aspecltual or grammatical category status) when assigning a novel word to a grammatical category (Bernal, Lidz, Millotte & Christophe, 2007; Booth & Waxman, 2003; Brown, 1957; Gelman & Markman, 1985; Hall, 1991, 1994; Hall & Graham, 1999; Hall & Moore, 1997; Hall, Waxman, & Hurwitz, 1993; Katz, Baker, & Macnamara, 1974; Taylor & Gelman, 1988). Such knowledge is language-specific, and there is evidence that the ability to correctly distinguish among these forms in English and recruit them in word extension tasks improves with age (cf. Soja, Carey, & Spelke, 1991; Waxman & Booth, 2001, *inter alia*), suggesting that from the outset, language learners are paying close attention to this distributional information in the input and the correlations between such forms and their referents in the world.

The contribution of the current study is twofold. First, we seek to extend the syntactic bootstrapping approach to the domain of adjectives by offering evidence that young children attend to patterns of adverbial modification to partition the adjectival category. Second, this research weds together two approaches – linguistic corpus analysis and experiments employing the preferential looking paradigm – to demonstrate that not only is distributional information robustly present in the exposure language, but that children appear to be aware of it and recruit it when learning new words.

4. **Corpus Analysis**

   **A. Corpus and Searching Procedure**

   The corpus selected for analysis was the British National Corpus (BNC). The entire BNC has over 100 million words collected from spoken and written texts. We narrowed our search to
the set of transcripts of spoken language (approximately 10,365,000 words), which allowed us to capture patterns that are present in both the input and the ambient speech to arrive at a broader picture of the range of linguistic input to which children are exposed. That is, whereas with a search of transcripts in the Child Language Data Exchange System (CHILDES) database (MacWhinney, 2000), we would need to be somewhat cautious about projecting from a caregiver’s child-directed speech to a larger speech sample, with the BNC, we are already searching over a more realistic speech sample, capturing the exposure language.

Using the Zurich BNCweb Query System, we targeted two sets of lexical items in our search. The first is a set of 10 adverbs, further divided into two sets: five proportional modifiers and five intensifiers. The second is a set of 10 adjectives, further divided into two sets: five absolute maximum standard GAs (MAX-GAs) and five relative GAs (REL-GAs). The entire set of targeted items is presented in Table 1.

| Table 1 | INSERT Table 1 ABOUT HERE. |

On the standard query page, with the corpus restricted to the spoken subset, the lexical item (adverb or adjective) was entered into the search field. Once the search results were displayed, a post-query ‘tag sequence search’ was conducted to gather collocations of adverbs and adjectives. For each adverb, we searched for adjectives in the positive form (tagged as “AJO”) in the “+1” position. For each adjective, we searched for adverbs (tagged as “AVO”) in the “-1” position. This set of results was then downloaded and compiled in a Microsoft Excel workbook, and the lists were then reviewed and coded by hand.

Adverbs were coded as proportional modifiers or not (i.e., whether the adverb selects for a for a maximally closed scale) if the resulting phrase meant that possession of the property was almost achieved, that it applied without exception, or that it picked out proportions of coverage.
Each adjective modified by one of the 10 target adverbs was coded as *maximal* (i.e., *maximally closed*) or not if it could be modified by *100%* or *almost* (cf. Kennedy & McNally, 2005, Cruse, 1980, and Rotstein & Winter, 2004), if the property corresponding to its antonym is absent in the adjective’s meaning (e.g., *healthy, clean vs. sick, dirty*), or if it is by default non-gradable, but has a highly frequent alternative allowing for an imprecise interpretation (e.g., *equal* and *identical*).

Two main sets of analyses were then conducted on these results. In the first, we examined the distribution of adverb-adjective bigrams with a frequency greater than one, evaluating modificational patterns for adverbs and adjectives separately. In the second, we calculated the conditional probability that each set of adverbs modifies an adjective with a maximally closed scale. Each set of analyses is reported separately in the following section.

Given the discussion in § 3, we make two predictions. First, adjectives corresponding to a maximally closed scale (MAX-GAs) will be more likely to be modified by proportional modifiers than those adjectives with an open scale (REL-GAs), since only the former adjectives provide the endpoints that such adverbs require (that is, we expect to see examples such as *completely full* more often than *completely big*). Second, intensifiers, despite being able to modify a wide range of adjectives, will be more likely to appear with REL-GAs, since these adjectives by default provide the open-ended scale these adverbs require (that is, we expect to see examples such as *very big* more often than *very full*). Finally, the same pattern should hold for adjectives, for all of the same reasons: MAX-GAs should be more likely to be modified by proportional modifiers than REL-GAs are, and REL-GAs should be more likely to be modified by intensifiers than MAX-GAs are.
B. Results

Distribution of Adverb-Adjective Bigrams

The distribution of adverb-adjective bigrams demonstrates a clear split in adverbial modification and the scalar structure of adjectives, based on reference to scalar endpoints. An analysis of the adjectives appearing with the 10 target adverbs (see Table 2) reveals two trends. First, proportional modifiers almost twice as likely to modify maximal adjectives than non-maximal ones. Note too that while maximal adjectives are modified by both types of adverb, approximately one third of the cases of adverbial modification (32.7%) involve proportional modifiers (compared to a mere 1.3% of non-maximal adjectives). Second, intensifiers are much more likely to modify non-maximal adjectives than maximal ones (Chi-square: \( \chi^2 = 45,473,928.5866, p < 0.001 \)).

An analysis of the adverbs modifying the 10 target adjectives reveals highly similar results (see Table 3). REL-GAs are much more likely than MAX-GAs to be modified by intensifiers, and MAX-GAs are more likely to be modifier by proportional modifiers. For example, while just over one tenth (10.6%) of MAX-GAs are modified by proportional modifiers, less than one percent (0.2%) of the REL-GAs are (\( \chi^2 = 231.85, p < 0.001 \)).

Conditional Probabilities

In the second analysis, we calculated a conditional probability based on an adverb’s likelihood of modifying an adjective with a closed scale. This analysis allows us to provide an answer to the following question: when a target adverb modifies an adjective, what is the
probability that the adjective maps onto a maximally closed scale? The results are presented in Table 4. Column A indicates the total number of appearances in the corpus. B indicates the number of instances in which members of the target group modify or are modified by members of the other group. C is the number of instances in which the adjective is a MAX-GA. Finally, Column D takes the information from Columns B and C and returns a conditional probability that reflects patterns of adverbial selectional restrictions related to adjectival scalar structure.6

The $\mathcal{P}$ for proportional modifiers ranges from .26 to .63, with an average of .49, while the $\mathcal{P}$ for non-proportional modifiers ranges from .03 to .16, with an average of .06.7 A two-tailed t-test reveals that these results are highly significant ($t(8) = 6.25, p = 0.00025$). The data in the table offer two main conclusions. First, proportional modifiers have a much higher probability of selecting for maximally closed scales than intensifiers do. Thus, a learner positing a maximally closed scale in an adjective’s representation is more likely to be correct when the adjective is modified by a proportional modifier. Second, the extremely low conditional probability for the intensifiers indicates that the appearance of these adverbs is a strong cue to the open-ended scalar structure of the adjectives they modify.

A look at which adverbs appear as modifiers with the two sets of GA is also revealing. Combined, the adverbs *a bit, as, fairly, quite, really, so, too, very* account for 80% of the adverbs modifying the target REL-GAs and only 41% of the adverbs modifying the MAX-GAs (two-6

---

6 See Goldberg, Casenhiser, and Sethuraman (2005) for a similar analysis of distributional information.

7 For reasons of space, we collapse across lexical items in these sets. Probabilities for individual lexical items are reported in Syrett (2007).
tailed t-test: t(4) = 6.4, p = 0.003). REL-GAs are by far more likely to be modified by these intensifiers or comparative adverbs than by proportional modifiers (two-tailed t-test: t(4) = 42.3, p < 0.00001). Thus, while REL-GAs are in general much more frequent than MAX-GAs and are more likely to be modified by adverbs in general (i.e., have more tokens), they are actually modified by a narrower range of adverbs (i.e., have fewer types), which are for the most part intensifiers that highlight their open-ended scalar structure.

C. Discussion

Two sets of analyses of the exposure language, as captured by the spoken subset of the BNC, converge to demonstrate that the two target sets of adverbs (proportional modifiers and intensifiers) and the two target sets of adjectives (MAX-GAs and REL-GAs) are qualitatively different in terms of modificational tendencies, a pattern that can only be driven by abstract semantic differences, namely restrictions related to adjectival scalar structure. The results demonstrate that MAX-GAs such as *full* are more likely to be modified by proportional modifiers such as *completely* than REL-GAs such as *big* are, and that intensifiers such as *very* are more likely to modify adjectives whose scales are not maximally closed. Consequently, the lion’s share of adverbs modifying REL-GAs are intensifiers highlighting their open scalar structure. These differences may be informative to the language learner about GA scalar structure. We now ask whether children track this information and recruit it in word learning. Experiments 1 and 2 were designed to address this question.
5. Experiment 1

A. Method

i. Participants

The participants were 33 children (19 girls, 14 boys; M age = 29 months 9 days; range = 28 months 0 days to 32 months 2 days). An additional 19 children were excluded because of fussiness (n = 2), inattentiveness (n = 9), experimenter error (n = 3), and age and gender balance across experimental conditions (n = 5). Children’s average vocabulary production was 515 words, as measured by parents’ responses on the MacArthur-Bates CDI: Words and Sentences (toddler form). There were also 30 adult controls (Northwestern University undergraduates fulfilling an experimental requirement for a Linguistics course) who completed a pen and paper version of the task. Data from three additional adults were excluded, because the participants indicated that a language other than English was their native language.

ii. Materials

The materials for both Experiments 1 and 2 consisted of five sets of five different objects, which were either photographs or computer-generated images. Each set of objects was labeled by a different novel adjective (*pelgy, wuggin, zaipin, vickel, keetel*). The auditory stimuli were recorded by a female native speaker of American English in a sound-attenuated booth. The speaker read from a script and was instructed to produce the stimuli in a style modeling child-directed speech. Sound files were edited using Praat software (Boersma & Weenink, 2005), controlling for articulation, pitch, amplitude, length, and overall consistency. Once finalized, the...

---

8 Children were recruited from College Park, MD, and the surrounding area. In Experiment 2, children were also recruited from the North Shore and greater Chicagoland area. Only those parents whose children were in the process of acquiring English as their native language and who had reported that less than 20% of a non-English language was spoken in the home environment were contacted.
sound files were synchronized with the video files using Final Cut Pro software by Apple Inc. These files were presented to participants on a computer in the laboratory at a rate of 30 frames per second.

iii. Procedure

Experiments 1 and 2 employed the intermodal preferential-looking paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Hollich, Rocrui, Hirsh-Pasek, & Golinkoff, 1999; Spelke, 1979). The structure of the experiments was modeled after the design used in Waxman and Booth (2001) and Booth & Waxman (2003). Child participants were run individually in our experimental laboratory, in a quiet and dimly lit room. They were seated directly facing a approximately 6' away from the screen on which the images were projected⁹, either in a highchair with their caregiver seated behind them, or on their caregiver’s lap while the caregiver wore a visor. Caregivers were asked to refrain from talking or offering any form of encouragement to their child. During the experimental session, if children solicited their caregiver’s attention, the caregiver was permitted to direct the child’s attention back to the general direction of the screen. The experimenter remained in the control room, out of view of the participants. Children’s looks to the experimental stimuli were recorded with a Sony EVI-D100 Color Video Camera centered inconspicuously above the screen. These videos were captured digitally onto an iMac computer using QuickTime.

９The child participants were run in two different laboratories, both designed and equipped to run the intermodal preferential looking paradigm. There were two differences between the two locations, based on room size and the screen on which the images were presented. One room was 14’ x 7’ and stimuli were presented via computer to a 44” x 24.5” wall-mounted plasma television screen. The other room was 14’ x 10’; visual stimuli were presented via a ceiling-mounted projector onto a 4.5’ by 4.5’ projection screen and audio stimuli were presented from two speakers located directly below the screen.
Each experiment consisted of five experimental trials designed to assess children’s interpretation of five different novel adjectives. Trials were presented in one of two orders, balanced across conditions and counterbalanced across subjects. Participants were randomly assigned to one of the two orders and one of three conditions, depending on whether and how the novel adjective was modified: completely, very, and no adverb. Age, gender, and vocabulary production were balanced within and across conditions. Paired t-tests revealed no significant difference across these three conditions (completely, very, and no adverb).

Prior to each trial, a still black-and-white photograph of a smiling infant appeared at the center of the video screen for four seconds, accompanied by the sound of a baby giggling. This segment was followed by a second four-second pre-trial phase in which participants saw a white screen and heard a voice announce the arrival of the objects. In this pre-trial phase, the speaker uttered the novel adjective without the adverb to help participants segment the adverb and adjective separately from the speech stream in the adverb conditions. Participants then moved on to the actual trial. Each trial consisted of three distinct phases: familiarization, contrast, and test. Participants all saw the same objects: what varied was what they heard. (See Table 5 for a representative set of objects and utterances for one trial.)

During the familiarization phase, participants saw two objects (e.g., containers), presented simultaneously, one on either side of the screen, then one at a time and heard a female voice describe the objects. Both objects could be described by a REL-GA (e.g., tall) and an MAX-GA (e.g., clear). A two-part contrast phase then followed, in which participants were first

---

10 Other object pairs included 2 long, straight sticks (contrast: short, curly); 2 wide smooth balls (contrast: narrow, bumpy); 2 large, patterned blocks (contrast: small, solid); and 2 high, closed windows (contrast: low, open). The corresponding test phase items were selected as described in the text.
shown a distractor object that instantiated the opposite properties of those seen in the familiarization phase and heard the speaker explicitly state that it could not be referred to with the adjective. This exemplar helped to narrow the range of possible referents by providing limits as to what can be described by the novel adjective. Participants were then shown one of the familiar objects from the previous phase. Intonation in this segment was controlled so that contrastive focus was placed on the adverb, when present or on the copula *is* in the no-adverb condition.

Finally, during a *test* phase, two images related to those seen during familiarization were displayed simultaneously on either side of a video screen. One object could be described with the REL-GA from familiarization, but not the MAX-GA (e.g., it was tall, but not clear), while the other could be described with the MAX-GA, but not the REL-GA (e.g., it was clear, but not tall). At the beginning of the test phase, the speaker drew the participants’ attention to the new objects. The screen then went momentarily blank for .33 seconds. The two objects then reappeared, and the speaker asked the participants to turn their attention to the object fitting the description of the novel adjective. The left-right position of the anticipated match was counterbalanced across trials. Audio stimuli were presented from two speakers located along the sides of the screen.

-------------------------------------------------------------

**INSERT Table 5 ABOUT HERE.**

-------------------------------------------------------------

Adults viewed the same video as the children and were run in groups of 1 to 3 in a 9' x 10.5' dimly-lit room in our laboratory. The video was projected from a Sony Digital8 Handycam onto a 60" Sony rear-projection television set. Participants were seated approximately 6' from the TV. Each participant was given a clipboard and a response packet to complete. Participants were instructed to shield their response packets from each other during the experimental session. The
experimenter either waited unobtrusively in a back corner of the room while the video played, or else waited outside the door of the room until the video was finished.

iv. Coding

Adults’ choice of one of the two objects during the test phase of each trial was averaged across the five trials and across subjects. The dependent measure for the adult data is the overall percentage of times adults chose the object that could be described by the REL-GA (e.g., the tall, solid container). Children’s eye gazes during the experimental sessions were recorded and saved as .mov files on a Macintosh computer, which were then coded off-line by an experimenter using the SuperCoder software (Hollich, 2003). Videos were coded frame by frame. The sound was removed to ensure that the coder, who was blind to the experimental condition, was only coding the direction of visual fixation during the test phase. Two same-size windows of the test phase were targeted for analysis. Each window lasted 45 frames, or 1.5 seconds. The first window, baseline, began 10 frames (approximately 333 ms) from the onset of “Look! They’re different” and was designed to assess child participants’ baseline attention. The second window, response, began 10 frames from the onset of the adjective and was designed to assess the interpretation participants assigned to the novel adjective.

Over the length of each window, the relevant measure was the proportion of time spent looking to the object that could be described by the REL-GA out of the total time spent looking at either object. This information was averaged across individual trials for each child, then across children in each condition. Data from trials in which the child was inattentive for 30% or more of the time were discarded. Children’s data were also discarded if they were inattentive (i.e., looked away from the two images on screen) for more than 20% of the time in three or more of the five trials. One experimenter was designated as the primary coder for each of the infant videos. A
second experimenter independently coded five of the videos (the test phase portion) across all
conditions in both experiments. There was 96% agreement between coders.

v. Predictions

Children’s looking pattern during the response window and adults’ choice of object
during the test phase are assumed to correlate with the interpretation they have assigned to the
novel adjective. Since the only difference between the experimental conditions is the presence
and type of adverbial modifier, we predict that any differences in interpretations and
consequently any differences in looking time we observe across experimental conditions must be
attributed to participants’ knowledge of the semantic consequences of the adverb’s selectional
restrictions and/or conditional probabilities of modification.

Thus, if children are able to use different adverbs as a cue to the meaning of the
adjectives they modify, then we predict the following. First, children in the no-adverb condition
will not look longer at either object during either test phase window (baseline or response),
unless their attention is captured by a salient property. Second, children in the two adverb
conditions will display differences in looking time that (a) differ significantly from chance in one
or both of the test phase windows, and (b) differ significantly between the two test phase
windows. During the response window, children in the very condition should look longer at the
object that can be described by the REL-GA property highlighted during familiarization,
hereafter the REL-GA object (e.g., the tall, solid container); whereas children in the completely
condition should look longer at the object that can be described by the MAX-GA property
highlighted during familiarization (e.g., the short, clear container).
B. Results

Children

The results bear out these predictions. (See Figure 1.) Not only does the presence of an adverb matter, but the kind of adverb matters. While children in the no-adverb condition showed no difference in looking time between the baseline and response windows, or from chance in the two windows, children in both of the adverb conditions pulled away from this baseline, in opposite directions, as predicted. Overall, children in the completely condition began by looking at the REL-GA object during the baseline window, and then switched to the MAX-GA object during the response window when asked about the meaning they assigned to the novel word.

By contrast, children in the very condition began by looking at the MAX-GA object during the baseline window, then switched to looking at the REL-GA object during the response window. Children in the no-adverb condition did not pattern consistently and were overall not above chance level in either test phase window. This asymmetry between the two adverbial conditions underscores the role of different types of adverbial modifiers in assigning an interpretation to a novel adjective. While one adverb (very) pulled children’s attention toward the REL-GA object during the response window, the other adverb (completely) pulled their attention away from it.

An 2 x 3 ANOVA was performed over the percentage of time spent looking at the REL-GA object with condition (completely, very, no adverb) as the between-subject factor and test phase window (baseline, response) as the within-subjects factor. The analysis revealed a main effect of test phase window, F(1, 33) = 6.196, p < 0.02, with the two adverb conditions, but not the no-adverb condition, displaying significant differences between test phase windows. There
was also a marginally significant interaction, $F(2, 30) = 2.653, p < 0.08$, driven by a difference in the direction of the percentages between the two adverb conditions, but no main effect of condition, $F(2, 32) = 0.272, p = 0.763$. This difference in looking time between the two adverb conditions held for the majority of children in each condition (completely: 8 out of 11, very: 10 out of 11), regardless of their overall level of vocabulary production.

A second analysis was conducted to determine whether responses in each of the two windows differed from chance (.50). Two tailed t-tests (df = 10) revealed that in the baseline window, only the children in the completely condition behaved significantly different from chance (completely: $t = 3.47, p = 0.006$; very: $t = -0.772, p = 0.458$; no adverb: $t = 1.028, p = 0.3282$), while in the response window, only the children in the very condition behaved significantly different from chance (completely: $t = -0.012, p = 0.991$; very: $t = 2.93, p = 0.015$; no adverb: $t = 0.34, p = 0.741$).

Why should there be more looks to the REL-GA object in the baseline window for children in the completely condition? Children who are aware of the meaning of completely and its restrictions on the scalar structure of its argument should assign the novel adjective a maximally closed scalar structure (e.g., clear in the above example). This pattern of looks appears to reflect an initial preference for a property that is novel with respect to the meaning assigned to the novel adjective, given the selectional restrictions of the adverb (cf. Waxman & Markow, 1995). This initial novelty preference is followed by a switch in attention when participants are asked directly for their interpretation of the adjective, toward the object and property that align with this meaning. Children in the very condition displayed a strikingly different pattern, hovering at chance level during the baseline window, but spending more time looking at the object with the REL-GA property during the response window. This pattern is
easily explained by appealing to the corpus data, which show that this intensifier has the
potential to appear with a wide range of adjectives, but is much more likely to modify those
without a maximal endpoint. Thus, children ultimately look more towards the REL-GA object in
this condition.

**Adults**

Adults patterned similarly. While they generally resisted selecting the REL-GA object,
the presence and type of adverbial modifier influenced their likelihood of choosing the REL-GA
object (completely: 18%, very: 34%, no adverb: 42%). A one-way ANOVA with condition as the
between-subject factor revealed a main effect of condition, $F(2, 29) = 4.661, p < 0.02$. A Tukey’s
HSD post-hoc ($p < 0.01$) analysis showed that adults in the completely, but not the very,
condition were significantly different from the no-adverb condition. A two-tailed t-tests ($df = 9$)
revealed that adults in both of the adverbial conditions, but not the no-adverb condition, were
significantly different from chance (completely: $t = -5.079, p < 0.001$; very: $t = -3.077, p = 0.013$;
no adverb: $t = -1.455, p = 0.180$).

For one trial (long, curly vs. short, straight), however, adults never selected the REL-GA
object (the long, curly one), no matter what condition they were in. Excluding this data point
from the analysis results in a similar trend of percentages (completely: 23%, very: 43%, no
adverb: 53%). A new set of t-tests reveal that while the completely and no-adverb conditions
continue to differ from each other, $F(2, 29) = 4.6016, p < 0.02$, now, only the completely
condition is significantly different than chance (completely: $t = 3.481, p = 0.007$; very: $t = 1.154$,
$p = 0.278$; no adverb: $t = -0.362, p = 0.726$).
C. Discussion

The corpus search showed a robust asymmetry in the distribution of certain adverbs and adjectives that aligned with the selectional restrictions of the adverbs and the scalar structure of the adjectives. In the present experiment, we sought to determine whether children recruit knowledge of this distributional asymmetry and form-meaning correspondence in word learning. The results provide evidence that this may, indeed, be the case. Children who hear a novel adjective that is not modified by an adverb show no systematic preference for one or another object property, whereas children who hear the same novel adjective modified by either completely or very display a very different pattern – the former apparently mapping it onto an MAX-GA meaning and the latter mapping it onto a REL-GA meaning.

However these results raise another question regarding the nature of the information being recruited. While completely and very differ in terms of their selectional restrictions, they also differ with respect to another feature: lexical frequency. In the spoken BNC, very has a frequency of 25,041, or 2421.36 instances per million words, and is highly likely to appear with REL-GAs. By contrast, completely has a frequency of only 822, or 79.48 instances per million words, making very over 30 times more frequent. In fact, only one child in our sample was reported to be producing completely, while 16 were reported to be producing very. If the children in these experiments knew nothing about the meaning of completely, and only recognized it as a novel adverb, distinct from very, it is possible that the mere presence of a novel adverb simply drew their attention away from their default preference and further, that any low-frequency or novel adverb – regardless of its semantics – would have done the same. We explore this possibility in Experiment 2.
6. Experiment 2

The purpose of Experiment 2 was to determine whether the low-level feature of (relative) lexical frequency was responsible for the pattern of results we saw in the previous experiment. In this experiment, we substituted in two new adverbs in place of very and completely to address this question. One (extremely) shares a similar semantic representation and distribution with that of the high-frequency intensifier very but like completely, has a low lexical frequency. The other (penticly) is a novel adverb phonologically similar to completely, which as a novel adverb, has no frequency in the exposure language. A summary of these features is presented in Table 6.

----------------------------------------------------------------------------
INSERT Table 6 ABOUT HERE.
----------------------------------------------------------------------------

If the pattern of responses we observed in the completely condition in the previous experiment can be attributed to a sheer novelty effect, then we should expect to see similarities between the completely condition and the two new conditions. If, however, the pattern was actually due to the meaning of completely, then we should see a difference between the completely condition and these two new conditions. Children in the penticly condition should perform at chance, while children in the extremely condition will either also be at chance if they are unaware of its meaning, or pattern the children in the very condition, if they are aware of it.

A. Method

i. Participants

The participants were 22 children (11 girls, 11 boys; range: 28;2 to 31;3 months, M: 29;8 months). An additional 12 children were excluded because of fussiness (n = 4), inattentiveness (n = 6), or equipment error (n = 2). Children’s average vocabulary production was 556 words. There was no significant difference in age or vocabulary production across conditions. There
were also 20 adult controls (Northwestern University undergraduates fulfilling an experimental requirement for a Linguistics course). All participants were native speakers of English.

ii. Materials and Procedure

The materials and procedure were identical to Experiment 1, with the exception of the adverbs. Two new adverbs, extremely and the novel adverb penticly, were substituted in place of the others. Children were randomly assigned to one of two experimental conditions: penticly (5 girls 6 boys, M: 29.8; average vocabulary: 439); and extremely (6 girls 5 boys, M: 30.0; average vocabulary: 556). Paired t-tests revealed no significant difference in vocabulary production among these three conditions. Coding was conducted as in Experiment 1.

B. Results and Discussion

Children

Response patterns for children across all conditions from Experiments 1 and 2 are presented in Figure 2. Unlike in Experiment 1, where the responses for children in both of the adverb conditions diverged from the no-adverb baseline and from chance, the responses for children in the extremely and penticly conditions were no different from chance-level.

A 2 x 2 ANOVA was performed over the percentage of time spent looking at the REL-GA object with condition (extremely, penticly) as the between-subject factor and test phase window (baseline, response) as the within-subjects factor. This analysis revealed no main effect of test phase window, $F(1, 21) = 0, p = 1.0$, no main effect of adverb, $F(1, 21) = 0.5, p = 0.4877$, and only a marginally-significant interaction $F(1, 21) = 3, p = 0.0986$. As in Experiment 1, a two-tailed t-test analysis was conducted to determine whether these conditions differed significantly from chance. In neither window and for neither condition was the response pattern
different from chance (baseline: extremely: \( t = 0.871, p = 0.404 \); penticly: \( t = -0.329, p = 0.749 \); response: extremely: \( t = 1.13, p = 0.285 \); penticly: \( t = -1.386, p = 0.196 \)).

Three sets of one-way ANOVAs were run collapsing over Experiments 1 and 2. The first compared looking time during the baseline window and revealed no main effects, \( F(4, 54) = 1.599, p = 0.189 \). The second compared looking time during the response window and revealed a marginally significant main effect, \( F(4, 54) = 2.212, p = 0.081 \). An ANOVA run on the difference in looking time between the baseline and response windows with condition as the between-subject factor, revealed a significant main effect, \( F(4, 54) = 3.449, p = 0.014 \). A post hoc Tukey’s HSD analysis (\( p < .01 \)) revealed that the significance arises from a difference in the very and completely conditions.

Thus, these analyses show that what matters is not necessarily a comparison among the absolute numbers in the two separate test phase windows for the different conditions, but rather a comparison between the two windows for each condition. That is, did the question about the meaning of the novel word draw the child’s attention toward one object or another and away from an initial looking preference? The results show that it is only in the completely and very conditions that this happened, although (importantly) in opposite directions. This claim is supported graphically in Figure 3, where the difference between the baseline and response windows among the conditions is made particularly clear.

We would like to take some space to consider what might be happening for children in the extremely condition. In this condition, children did not pattern differently between the baseline and response windows and were no different from chance in either window. Thus, the semantic representation and distribution of this adverb appear not to have played a role for them
as the semantics of very did. Recall that extremely is about half as frequent as completely. It may be that although extremely is much more likely to modify a non-maximal adjective than a maximal one (in our analysis, 390 vs. 10 adjective-modifying occurrences), it is simply too infrequent for children to have formed a hypothesis about its selectional restrictions. However, it may be the case that children have a budding hypothesis about extremely’s selectional restrictions, but the low frequency of this lexical item coupled with the high frequency of a lexical item with overlapping meaning (very) results in a blocking effect where children are reluctant to map the adjective onto a REL-GA interpretation, since they expect that if that corresponding property were to be singled out, that the adjective would have been modified by very instead. We leave this possibility open for future research.

**Adults**

Adults in Experiment 2 were more likely to select the REL-GA object when asked for the meaning of the novel word that had been modified by either penticly or extremely (completely: 18%, very: 34%, no adverb: 42%, penticly: 54%, extremely: 56%). A one-way ANOVA collapsing over the results from both experiments with condition as the between-subject factor revealed a main effect, $F(4, 49) = 8.134, p < 0.0001$. A Tukey’s HSD post-hoc analysis showed revealed significant differences between completely and no adverb ($p < 0.05$), penticly ($p < 0.01$), and extremely ($p < 0.01$).

As in Experiment 1, adult participants were again reluctant to select the REL-GA object for the long, curly vs. short, straight trial. Percentages excluding this trial are in the same direction (completely: 23%, very: 43%, no adverb: 53%, penticly: 65%, extremely: 70%). An ANOVA excluding this data point supports the main effect of condition, $F(4, 49) = 7.656, p < 0.0001$, with a Tukey’s HSD post-hoc analysis revealing the same significant differences. Two-
tailed t-tests (df = 9) were also run to determine if the conditions differed significantly from chance level. For an analysis including all five trials, neither condition differs significantly from chance (extremely: t = 0.667, p = 0.523; penticly: t = 0.667, p = 0.522). However, an analysis excluding the pelgy trial shows that the extremely condition differs significantly from chance (t = -4, p = 0.003), and the difference between the penticly condition chance is marginally significant (t = -1.974, p = 0.080).

For adults, then, the combined results from Experiments 1 and 2 tell us the following. When a novel adjective is not modified by an adverb, and adults are not given any guidance regarding its semantics, they are not more likely to map it to a MAX- or REL-GA interpretation. When the adjective is modified by completely, however, adults are highly likely to assign it an absolute GA interpretation. By contrast, when the adjective is modified by extremely, adults are likely to assign it a REL-GA interpretation. Interestingly, extremely was chosen as a modifier in Experiment 2 because it has semantics similar to that of very; however, when a novel adjective is modified by very, adults are more likely to assign the adjective an MAX-GA interpretation than a REL-GA interpretation.

This pattern makes sense when we consider two things. First, the function of very is to indicate a sizable divergence from the standard, whereas extremely indicates an even larger divergence, which may result in extremely having a more limited distribution and to be a better indicator of the REL-GA status of its argument. Second, it is possible that adults are simply being more economical. Although very is more likely to modify adjectives with a REL-GA interpretation by default, it can also modify those adjectives that have a default absolute GA or non-gradable interpretation, but which also lend themselves to a REL-GA interpretation. REL-GAs by definition encode the context in their semantic representation. Adults may prefer to
assign a novel adjective an interpretation that is not linked to the context unless they have good reason to do so (cf. Kennedy, 2007). Thus, when given a choice between a REL-GA and absolute GA interpretation, they opt for the latter, because its standard does not depend on the context.

7. General Discussion

The current research contributes to our understanding of how children learn to classify novel adjectives according to their abstract semantic representations by offering evidence concerning the informativity of the input and the nature of the information children recruit in word learning. First, the results of a corpus analysis demonstrate that robust patterns of adverb-adjective bigrams in the exposure language allow children to sort adverbs according to their selectional restrictions, and adjectives according to their scalar structure. Specifically, GAs such as *full* that map objects onto a maximally closed scale and proportional modifiers such as *completely* that select for a scale with endpoints are drawn to each other, while intensifiers such as *very* that are more widespread in their distribution, but which indicate a relative interpretation, are drawn to GAs such as *big* that map onto an open scale and have a contextually-based standard of comparison. These patterns converge to highlight the overwhelming evidence for the language learner that differences arising from scalar structure in the semantic representations of these lexical items are signaled in the exposure language.

Second, the results from a set of word-learning experiments demonstrate that 30-month-olds appear to be aware of such information and recruit it when assigning interpretations to novel words. In these experiments, children were familiarized to two objects that had a property best described by REL-GA and another best described by a MAX-GA and heard the objects described by a novel adjective such as *pelgy*. A test, when the object properties were teased apart and the children were asked about the interpretation they assigned to the adjective, they were more likely
to turn their attention toward the REL-GA object if the adjective had been modified by *very*, but away from this object and toward the MAX-GA object if the adjective had been modified by *completely*. Because the presence and type of adverbial modifier was the only variable among the conditions, children’s decision to associate the meaning of the adjective with one object property or another must have been guided by the information provided by the adverbs. Thus, even infants who are not yet producing these adverbs seem to have performed a probabilistic analysis of their frequency in the exposure language and deduced something about their meaning, and consequently the likely meaning of their arguments.

Adverbs are thus one of the syntactic cues that are doing for adjectives what syntactic frames (or clausal argument structures) do for verbs. In both cases, distributional differences of surface-level cues are informative about abstract differences that partition the lexical items according to their semantic representations. Just as language learners might track the relative frequency with which a given verb appears across a range of frames, and the relative frequency with which frames appear across a range of verbs (Alishahi & Stevenson, 2005a, b; Naigles & Hoff-Ginsberg, 1995), so they track the relative frequency with which adjectives are modified by a range of adverbs and the relative frequency with which adverbs modify a range of adjectives. In both cases, language learners use what they have deduced about meanings of these items when making deductions about the meanings of new items. We propose that this is a *reciprocal* word learning process, whereby children use what they have learned about adverbs to deduce something about the meaning of the adjectives they modify, and at the same time use what they have learned about adjectives to deduce something about the selectional restrictions of the adverbs that modify them.
Importantly, learners must have some prior knowledge about what gives rise to these diverging patterns of modification, since mere surface-level differences in distribution and conditional probabilities will not be directly informative about what underlies differences in relative frequency. It seems likely that children must have some \textit{a priori} expectation that lexical items will vary along the lines of scalar structure and are waiting to encounter evidence that will provide them with information about how to classify a new lexical item – that is, whether or not it maps onto a scale (with continuous or discrete intervals), and if so, whether that scale is open or closed. When children are learning about adjectives, adverbs are one cue that can provide them with just this sort of evidence.
References


British National Corpus (BNC). Available from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights reserved. (http://www.natcorp.ox.ac.uk/)


Goldberg, A. E., Casenhiser, D., & Sethuraman, N. (2005). The role of prediction in

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it:
Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*,
23-45.

Hall, D. G. (1991). Acquiring proper names for familiar and unfamiliar animate objects: Two-

Development, 65*, 1299-1317.

Hall, D. G., & Graham, S. A. (1999). Lexical form class information guides word-to-object

understanding of the semantics of adjectives and count nouns. *Journal of Experimental

adjectives and count nouns. *Child Development, 64*, 1651-1664.


comprehension in infants: Introducing the split-screen preferential-looking paradigm.
Poster presented at the Biennial Meeting of the Society for Research on Child
Development, April 15-18, Albuquerque, NM.

Natural Languages* (pp. 123-155). Cambridge: Cambridge University Press.


Tables

Table 1: Targeted lexical items (adverbs and adjectives)

<table>
<thead>
<tr>
<th>adverbs</th>
<th>adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>proportional modifiers</td>
<td>absolute maximum standard GAs (MAX-GAs)</td>
</tr>
<tr>
<td><em>almost, completely, entirely, half, totally</em></td>
<td><em>clean, dry, empty, full, straight</em></td>
</tr>
<tr>
<td>intensifiers</td>
<td>relative GAs (REL-GAs)</td>
</tr>
<tr>
<td><em>extremely, really, relatively, too, very</em></td>
<td><em>big, high, long, tall, wide</em></td>
</tr>
</tbody>
</table>

Table 2: Adjectives appearing with the 10 target adverbs

<table>
<thead>
<tr>
<th></th>
<th>maximal adjectives</th>
<th>non-maximal adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>proportional modifiers</td>
<td>490</td>
<td>257</td>
</tr>
<tr>
<td>intensifiers</td>
<td>1010</td>
<td>18957</td>
</tr>
</tbody>
</table>

Table 3: Adverbs appearing with the 10 target GAs

<table>
<thead>
<tr>
<th></th>
<th>MAX-GAs</th>
<th>REL-GAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>proportional modifiers</td>
<td>39</td>
<td>4</td>
</tr>
<tr>
<td>intensifiers</td>
<td>330</td>
<td>2453</td>
</tr>
</tbody>
</table>
Table 4: Probability of modifying an adjective with a maximal scalar endpoint

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td># instances</td>
<td>9359</td>
<td>1126</td>
<td>585</td>
<td>0.49</td>
</tr>
<tr>
<td># modifying</td>
<td>proportional modifier</td>
<td>50167</td>
<td>21281</td>
<td>1078</td>
</tr>
</tbody>
</table>

\[ \mathcal{P}(\text{modifies a maximal adjective} | \text{modifies an adjective}) = \frac{C}{B} \]
Table 5: Representative set of objects and speaker utterances used in Experiment 1

<table>
<thead>
<tr>
<th>pre-trial familiarization</th>
<th>contrast (distractor, target)</th>
<th>test</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 sec</td>
<td>18 sec (6 s, 6 s, 6 s)</td>
<td>7 sec</td>
</tr>
<tr>
<td></td>
<td>7 sec</td>
<td>12 sec (4 s, 8 s)</td>
</tr>
<tr>
<td>(blank screen)</td>
<td>2 tall, clear jars</td>
<td>short, opaque</td>
</tr>
<tr>
<td></td>
<td>jar</td>
<td>short, transparent jar</td>
</tr>
</tbody>
</table>

Let’s look at some things that are pelgy!

Look! These are both {adverb / ∅} pelgy!

This one is not pelgy!

Uh oh!

Yay! This one is completely / is very / is ∅ pelgy!

Look! They’re different!

Which one is pelgy?

Uh oh!

This one is pelgy!

Table 6: Adverbs used in Experiments 1 and 2

<table>
<thead>
<tr>
<th>adverb</th>
<th>frequency</th>
<th>instances per million</th>
<th>semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>very</td>
<td>25,041</td>
<td>2421.36</td>
<td>intensifier</td>
</tr>
<tr>
<td>completely</td>
<td>822</td>
<td>79.48</td>
<td>maximally-closed scale selecting</td>
</tr>
<tr>
<td>extremely</td>
<td>480</td>
<td>46.41</td>
<td>intensifier</td>
</tr>
<tr>
<td>penticly</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Figures

Figure 1: Children’s responses as measured by percentage of time spent looking at the REL-GA object

Figure 2: Children’s responses as measured by percentage of time spent looking at the REL-GA object
Figure 3: Children’s responses as measured by difference between two test phase windows