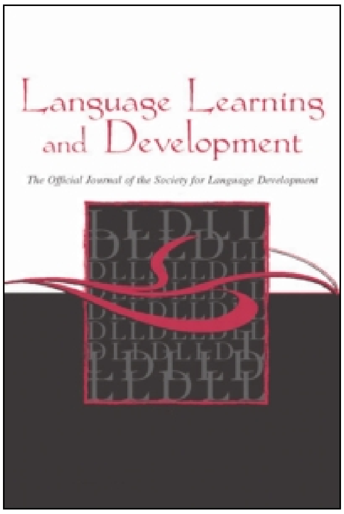


Language Learning and Development
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653671

30-Month-Olds Use the Distribution and Meaning of Adverbs to Interpret Novel Adjectives

Kristen Syrett[a]; Jeffrey Lidz[b]
[a] Center for Cognitive Science, Rutgers, The State University of New Jersey, [b] Linguistics Department, University of Maryland, College Park

Online publication date: 12 October 2010

To cite this Article Syrett, Kristen and Lidz, Jeffrey(2010) '30-Month-Olds Use the Distribution and Meaning of Adverbs to Interpret Novel Adjectives', Language Learning and Development, 6: 4, 258 — 282
To link to this Article: DOI: 10.1080/15475440903507905
URL: http://dx.doi.org/10.1080/15475440903507905

Language Learning and Development, 6: 258–282, 2010

ISSN: 1547-5441 print / 1547-3341 online
DOI: 10.1080/15475440903507905

Psychology Press
Taylor & Francis Group

30-Month-Olds Use the Distribution and Meaning of Adverbs to Interpret Novel Adjectives

Kristen Syrett
Center for Cognitive Science, Rutgers, The State University of New Jersey

Jeffrey Lidz
Linguistics Department, University of Maryland, College Park

Word learners are able to use the syntactic context of a word as one source of information to narrow down the space of possible meanings. We examine this bootstrapping process in the domain of adjectives, focusing on the acquisition of subcategories of *Gradable Adjectives* (GAs). We first show that robust patterns of adverbial modification in natural language sort GAs according to scalar structure: proportional modifiers (e.g., *completely*) tend to modify absolute maximum standard GAs (e.g., *full*), while intensifiers (e.g., *very*) tend to modify relative GAs (e.g., *big*). We then show in a word-learning experiment that 30-month-olds appear to be aware of such distributional differences and recruit them in word learning, assigning an interpretation to a novel adjective based on its modifier. We argue that children track both the range of adjectives modified by a given adverb and the range of adverbs modifying a given adjective, and use such surface-level information to classify new words according to possible pre-existing semantic representations.

INTRODUCTION: ADJECTIVES IN WORD LEARNING

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 available for young word learners to enable them to home in on one in particular. Thus, other sources of information must be available to guide word learning. One such information source resides in a word's syntactic context (see Woodward, 2000, for a review of other possible sources). Because different syntactic contexts are associated with different semantic features, learners can narrow down the space of possible word meanings by restricting the meaning to those that fit with the syntactic context. The usefulness of syntactic context has been appealed to in many accounts of noun, adjective, and verb learning (e.g., Brown, 1957; Fisher, 2002; Gleitman, 1990; Katz, Baker, & Macnamara, 1974; Landau & Gleitman, 1985; Naigles, 1990; Booth & Waxman, 2003; Hall, 1994; Waxman & Booth, 2001; see Waxman and Lidz, 2006, for a

Correspondence should be addressed to Kristen Syrett, Center for Cognitive Science, Rutgers University, Psychology Building Addition, Busch Campus, 152 Frelinghuysen Rd., Piscataway, NJ 08854. E-mail: k-syrett@ruccs.rutgers.edu

review). In this paper, we examine the role of the syntactic context in the acquisition of gradable adjectives, such as *big* and *full,* and, more specifically, subcategories of these adjectives. To do this, we must first identify the relevant subcategories, the linguistic motivation and diagnostics for them, and the way in which distributional patterns in the syntax highlight differences among these subcategories.

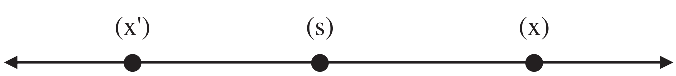
Although in the child language literature, adjectives are typically talked about as referring to "object properties" in contrast to nouns, which are said to refer to "object kinds," this differentiation does not capture the full range of adjective meanings that a child must acquire. Some adjectives, such as *herbivorous*, *wooden*, and *six-legged*, can be treated this way. When combined with a noun the referent of the resulting phrase is found in the intersection of the set of things bearing the adjectival property with the set of things bearing the nominal property (Clark, 1970; Kamp, 1975; Montague, 1974; Parsons, 1972). For example, a "herbivorous mammal" stands at the intersection of things that are herbivorous and things that are mammals. However, not all adjectives denote properties that are all-or-none. *Gradable adjectives*, for example, allow for objects to possess varying degrees of the relevant property.

Gradable adjectives (GAs) are distinguished from nongradable adjectives by their ability to appear in comparative constructions, 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.

This contrast follows from the difference in meaning associated with gradable versus nongradable adjectives. Nongradable adjectives such as *extinct* or *herbivorous* predicate a property of an individual. In contrast, GAs such as *big* or *expensive* measure the degree to which an entity possesses a property (e.g., SIZE, PRICE). 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?*

Because reference to degrees is encoded in the meaning of GAs, even in the noncomparative, or *positive*, form (i.e., *big* vs. *bigger*, *tall* vs. *taller*, and so on), GAs are always implicitly comparative. Taking as an example the GA *big* and the scale in (3), 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* would be false, since (x′) does not exceed (s).

3. 

THE SEMANTICS OF GRADABLE ADJECTIVES

Our interest is not so much on the nongradable versus gradable distinction but rather on a distinction within gradable adjectives. GAs fall into two subcategories. One subcategory, *relative*

GAs, requires 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 an elephant). This comparison class can be provided by a variety of sources (e.g., the modified noun, the discourse context). See Kamp and Partee (1995) for further discussion. This has implications for the truth value of the sentence in which such adjectives appear: holding Mickey's size constant, the sentence *Mickey is big* might be FALSE if Mickey is an elephant, or TRUE if Mickey is a mouse. In fact, one distributional feature of relative GAs is their ability to be followed by a *for an X phrase*, as in (4).

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

The other subcategory of gradable adjectives, *absolute GAs*, can also appear in comparative environments, as in (5), but does not depend on a comparison class for interpretation, as shown in (6). For example, you can compare degrees of a container's fullness without reference to other containers.

5. a. How full is the pitcher?
   b. The pitcher was fuller than I had anticipated (so I spilled water all over the table).
6. ??That is full for a pitcher.

This difference between GAs such as *big* and *full* arises from differences in their scalar structure. *Relative GAs* such as *big*, which depend on a contextually determined comparison class and standard of comparison, map objects onto an open-ended scale, such as the one in (3) above. In contrast, *absolute maximum standard GAs* (or *maximal GAs*) such as *full*, which do not depend on the context for the standard, map objects onto a scale that is closed at one or both ends, and the endpoint(s) serve(s) as the standard. For example, with *full*, the endpoint signals maximal fullness (or alternatively, zero degree of emptiness) (cf. Kennedy & McNally, 2005; Rotstein & Winter, 2004).

Thus, GAs differ with respect to their scalar structure on two separate but intimately related aspects: the presence or absence of scalar endpoints and the role of the endpoint or context in determining the *standard of comparison*. This difference is captured in (7).[1]

7. a. relative GAs (open-ended scale) (e.g., *big/small*, *old/young*)

   b maximal GAs (scale closed on one/both ends) (e.g., *full/empty*, *clean/dirty*)

[1]As we mentioned, GAs come in two varieties: *relative* and *absolute GAs*. We concentrate here on the contrast between *relative GAs*, which map objects onto open scales, and *absolute maximum standard GAs*, which map objects onto maximally closed scales, leaving aside *absolute minimum standard GAs*, such as *dirty* and *wet*. As we hint at in this discussion, *maximal* and *minimal GAs* go hand in hand, in referring to the absence or presence, of the relevant property, respectively (see Kennedy & McNally, 2005; Rotstein & Winter, 2004, for further discussion).

We note that while we refer to differences among GAs, we are actually discussing differences among possible *interpretations*, or *readings*, of these adjectives. Thus it is possible for a GA to have multiple interpretations, with a default or primary interpretation relative interpretation and a secondary absolute interpretation, or vice versa. We note, too, that in this discussion we leave aside mention of *absolute minimum standard GAs*, such as *dirty* and *wet*, which go hand in hand with *absolute maximum standard GAs*.

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 relative GAs such as *big* (Barner & Snedeker, 2008; Ebeling & Gelman, 1988, 1994; Gelman & Ebeling, 1989; Smith, Cooney, & McCord, 1986) as well as the restrictions on the variable role of the context in the interpretation of different subcategories of GAs (Syrett, 2007; Syrett, Kennedy, & Lidz, 2010). In the current paper, we ask how preschoolers could have arrived at this knowledge and how children could know how to map a newly encountered GA onto the right representation. We propose that one strategy is for children to appeal to the linguistic environment in which a GA appears and to take advantage of a strong correlation between surface-level distributional patterns and abstract meaning.

FORM, MEANING, AND ADVERBIAL MODIFICATION

The distributional feature that we examine here is different patterns of adverbial modification. Different GAs are modified by different adverbs. As (8) illustrates, *proportional modifiers* such as *almost*, *completely*, and *half* are able to modify some GAs, but not others.

8. a. The glass is almost/completely/half full.
   b. #Her brother is almost/completely/half big.

This pattern can be contrasted with the one in (9), which illustrates that *intensifiers* such as *very* 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 that these adverbs place on the scalar structure of the lexical items they modify, whether they be adjectives or phrases (e.g., *almost ran 5 miles*). *Proportional modifiers* such as *half*, *almost*, and *completely* can only modify GAs that are able to supply a maximal scalar endpoint (Cruse, 1980; Paradis, 1997; Kennedy & McNally, 2005; Rotstein & Winter, 2004), whereas *intensifiers* such as *very* can modify any GA that allows for a relative interpretation, even if it is not relative by default. Thus, surface patterns of modification are informative about underlying differences in the abstract semantic representation of these lexical items.

As noted in the introduction, discussions of similar form-meaning correspondences have 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 distributional 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 inform hypotheses about the semantic representation of a verb (Fisher, 2002; Naigles, 1990; Naigles, Fowler, & Helm, 1992; see also Wagner, 2006). The inferences from distributional features to semantic categories are licensed in this kind of learning mechanism only because of the child's prior knowledge of the set of possible semantic representations and the rules for linking semantic representations to syntactic representations. Of course, the particular surface realization of these syntactic representations will vary from language to language, and so identifying the link between surface form and syntactic representation is a prerequisite for engaging this learning mechanism.

Children's ability to use the syntactic distribution to guide hypotheses about semantic interpretation 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 aspectual 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). Knowledge of particular surface distributional properties (e.g., that *the* is a determiner) is language-specific, and the ability to correctly distinguish among these forms in English and recruit them in word extension tasks improve with age (cf. Soja, Carey, & Spelke, 1991; Waxman & Booth, 2001, *inter alia*). This suggests that from the outset, language learners are paying close attention to the distributional features of the input and the correlation between the distributional properties of a word and its meaning.

The contribution of the current study is twofold. First, we offer evidence that young children attend to patterns of adverbial modification (and the concomitant adverbial meanings) in order to partition the adjectival category into distinct subcategories. Second, this research weds together two approaches—linguistic corpus analysis and word learning experimentation—to demonstrate (a) that robust distributional differences between categories are evident in natural speech and (b) that children are sensitive to such distributional information at a very young age when learning new adjectives.

CORPUS ANALYSIS

Corpus and Searching Procedure

The purpose of the corpus search is to demonstrate that patterns of adverbial modification in natural language partition gradable adjectives (GAs) into two distinct subcategories—those with an open scale (which are more likely to be modified by intensifiers such as *very*) and those with a maximally closed scale (which are more likely to be modified by proportional modifiers such as *completely*). Having illustrated this pattern, we can then proceed with a set of language learning experiments relying upon such differences. The corpus of natural language we selected for analysis was the British National Corpus (BNC). The entire BNC has more than 100 million

TABLE 1
Targeted Lexical Items (Adverbs and Adjectives)

Adverbs	*Adjectives*
proportional modifiers	absolute maximum standard (maximal) GAs
almost, completely, entirely, half, totally	*clean, dry, empty, full, straight*
intensifiers	relative GAs
extremely, really, relatively, too, very	*big, high, long, tall, wide*

words collected from spoken and written texts. We narrowed our search to the set of transcripts of spoken language (approximately 10,365,000 words).[2]

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 maximal GAs and five relative GAs. The entire set of targeted items is presented in Table 1. In this search, we targeted canonical exemplars from each set of lexical items that are frequently cited in discussions of GAs in the semantics literature.

On the standard query page, with the corpus restricted to the spoken subset, we entered the lexical item (adverb or adjective) into the search field. Once the search results were displayed, we conducted a postquery "tag sequence search" to gather collocations of adverbs and adjectives. For each adverb, we searched for adjectives in the positive (i.e., noncomparative) form (tagged as "AJO") in the "+1" position (i.e., immediately after the adverb). For each adjective, we searched for adverbs (tagged as "AVO") in the "−1" position (i.e., immediately before the adjective). This set of results was then downloaded and compiled in a Microsoft Excel workbook, and the lists were then reviewed and coded by hand by the first author and reviewed by at least two other linguists with a background in semantics.

Adverbs were coded as *proportional modifiers* or not (i.e., whether the adverb selects 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. Examples of such modifiers included *nearly*, *half*, and *spotlessly* to be contrasted with intensifiers such as *quite* and *really*. Each adjective modified by one of the 10 target adverbs was coded as *maximal* (i.e., *maximally closed*) if it could be modified by *100%* or *almost* (cf. Kennedy & McNally, 2005; Cruse, 1980; Rotstein & Winter, 2004), if the property corresponding to its antonym is absent in the adjective's meaning (e.g., *healthy* vs. *sick*, *clean* vs. *dirty*), or if the adjective is by default nongradable but has a highly frequent alternative allowing for an imprecise

[2]This approach to our corpus analysis allowed us to present baseline evidence for the targeted distinctions and perhaps at the same time capture distinctions that may be present in the exposure language. That is, whereas with a search of transcripts in the Child Language Data Exchange System (CHILDES) database (MacWhinney, 2000), we would need to be 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. An additional consideration is that the CHILDES corpus is relatively small and contains relatively few instances of the adjectives and adverbs in question. To ensure that our corpus was a realistic estimate, we compared a portion of the CHILDES corpus with a comparably sized subset of the BNC. We found parallels in those corpora with respect to adjective and adverb distribution which warrants treating the BNC as a realistic sample of both speech in general and speech to children (see Syrett, 2007, for details).

interpretation. Examples of such adjectives included *clear*, *identical true*, and *universal* to be contrasted with nonmaximal adjectives such as *busy*, *good*, and *horrible*.

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.

Given the discussion in §3, we make two predictions. First, proportional modifiers, which require a closed scale, will be more likely to appear with maximal GAs, which supply a scalar endpoint these 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 relative 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 GAs, for all of the same reasons: maximal GAs should be more likely to be modified by proportional modifiers than relative GAs are, and relative GAs should be more likely to be modified by intensifiers than maximal GAs.

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 are almost twice as likely to modify maximal GAs as nonmaximal ones. While maximal GAs are modified by both types of adverbs, approximately one-third of the cases of adverbial modification of these adjectives (32.7%) involve proportional modifiers (compared to a mere 1.3% of nonmaximal adjectives). Second, intensifiers are 19 times more likely to modify nonmaximal adjectives as maximal GAs ($\chi^2 = 4547.66$, $p < 0.001$).[3]

TABLE 2
Adjectives Appearing With the 10 Target Adverbs

	Maximal GAs	*Relative GAs*
proportional modifiers	490	257
intensifiers	1,010	18,957

[3]In fact, removing the intensifier that is by far the most frequent, *very*, the picture is even more dramatic, given the ability of *very* to modify a wider range of adjectives as discussed earlier (an ability that is arguably frequency-driven): approximately 73% of the cases of adverbial modification of maximal GAs is by proportional modifiers (490 vs. 178 compared to 490 vs. 257 in Table 2), and the remaining four intensifiers together are approximately 30 times more likely to modify nonmaximal adjectives (5270 vs. 178 compared to 18957 vs. 1010 in Table 2).

TABLE 3
Adverbs Appearing With the 10 Target GAs

	Maximal GAs	*Relative GAs*
proportional modifiers	39	4
intensifiers	330	2453

An analysis of the adverbs modifying the 10 target adjectives reveals similar results (see Table 3). Relative GAs are much more likely than maximal GAs to be modified by intensifiers, and maximal GAs are more likely to be modified by proportional modifiers than relative GAs (about seven and nearly 10 times, respectively). While just over one tenth (10.6%) of maximal GAs are modified by proportional modifiers, less than 1% (0.2%) of the relative GAs are ($\chi^2 = 231.85$, $p < 0.001$).

Conditional Probabilities

In the second analysis, we looked at conditional probability, which allowed us to 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 (i.e., is a maximal GA)? The results are presented in Table 4. Column A indicates the total number of appearances in the corpus. Column B indicates the number of instances in which these adverbs modify an adjective. Column C is the number of instances in which the adjective is a maximal 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 (see Goldberg, Casenhiser, & Sethuraman, 2005, for a similar analysis of distributional information in verb learning).

The $\wp$ for proportional modifiers ranges from 0.26 to 0.63, with an average of 0.49, while the $\wp$ for nonproportional modifiers ranges from 0.03 to 0.16, with an average of 0.06. For reasons of space, we collapse across lexical items in these sets. Probabilities for individual lexical items are reported in Syrett (2007).

A two-tailed t-test reveals that these distributions are significantly different ($t(8) = 6.25$, $p < 0.0003$). The data in the table offer two main conclusions. First, proportional modifiers have a much higher probability of selecting for maximally closed scales than do intensifiers. Thus,

TABLE 4
Probability of Modifying an Adjective With a Maximal Scalar Endpoint

	A	*B*	*C*	*D*
	# instances	# modifying adjective	# modifying a maximal adjective	$\wp$(modifies a maximal GA | modifies an adjective) (C/B)
proportional modifier	9359	1126	585	0.49
intensifier	50167	21281	1078	0.06

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. As an illustration of this difference, we note that among the most frequent adjectives modified by proportional modifiers are adjectives such as *certain*, *different*, *new*, *separate*, and *wrong*. By contrast, the most frequent adjectives modified by intensifiers include canonical relative GAs such as *good*, *bad*, *difficult*, *important*, *late*, and *nice*.

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*, and *very* account for 80% of the adverbs modifying the target relative GAs and only 41% of the adverbs modifying the maximal GAs (two-tailed t-test: $t(4) = 6.4$, $p = 0.003$). Relative GAs are significantly 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 relative GAs are in general much more frequent than maximal 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.

DISCUSSION

Two sets of analyses of natural language, as captured by the spoken subset of the BNC, converge to demonstrate two things. First, maximal GAs such as *full* are more likely to be modified by proportional modifiers such as *completely* than are relative GAs such as *big*. Second, intensifiers such as *very* are more likely to modify nonmaximal GAs. Consequently, the lion's share of adverbs modifying relative GAs consists of intensifiers highlighting the adjectives' open scalar structure. This pattern is clearly driven by differences in the adjectives' scalar structures. Thus, these surface-level differences observed in natural language may be informative to the language learner about the semantic representation of these lexical items. We now ask whether children appear to track such information and recruit it in word learning. The following set of experiments was designed to address this question.

EXPERIMENT 1

The purpose of this experiment was to determine whether information that is available to children in the form of adverbial modification could be used to assign an interpretation to novel adjectives consistent with the corresponding semantic restrictions and distributional tendencies discussed above.

Method

Participants. The participants were 33 children (19 girls, 14 boys; M = 29 months 9 days; range = 28 months 0 days to 32 months 2 days). An additional 14 children were excluded because of fussiness ending the session (n = 2), inattentiveness[4] (n = 9), or technical difficulties/experimenter error (n = 3). Children were recruited from College Park, Maryland, and the surrounding area. In Experiment 2, children were also recruited from the North Shore and greater Chicago 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.

Children's average vocabulary production was 515 words as measured by parents' responses on the MacArthur-Bates CDI: Words and Sentences (toddler form). Age, gender, and vocabulary production scores were balanced within and across three experimental conditions: *completely* (6 girls 5 boys, M = 30 months; average vocabulary: 483), *very* (6 girls 5 boys, M: 29 months; average vocabulary: 522), no adverb (7 girls 4 boys, M = 30 months; average vocabulary: 539). Paired t-tests revealed no significant difference in overall vocabulary production among these three conditions. A set of adult controls also participated; their results are reported along with Experiment 2.

Materials

Materials for both Experiments 1 and 2 consisted of five videos (each constituting one trial), each presenting a set of five objects, images of which were either photographs or computer-generated. Each set of five 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 recording booth. The speaker read from a script and was instructed to produce the stimuli in a style modeling the prosody of 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 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.

Procedure

Experiments 1 and 2 employed the intermodal preferential-looking paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Hollich, Rocroi, Hirsh-Pasek, & Golinkoff, 1999; Spelke, 1979). The structure of the experimental trials and analysis was loosely modeled after the design used in Booth and Waxman (2003, 2009), Waxman and Booth (2001), and Waxman et al. (2009). Children were run in one of two different laboratories, both designed and equipped to run the

[4]Data were excluded from children who were inattentive for more than 30% of the time for three or more of the five trials. While this may seem like too stringent a measure, when we look at all of the individual children whose data were excluded over Experiments 1 and 2 for reasons of inattentiveness, it appears less so. There was only one child whose average 'attentive time' over all five trials approached 70% (69%); only three others were above 60%, and all others – with the exception of one child whose average attention was 33% – were between 50 and 60%. In addition, for those children whose two remaining trials averaged over 70% attentive time, their excluded trials averaged less than 50%, which the exception of one child, whose average was only 60% for the excluded trials.

intermodal preferential looking paradigm. In each laboratory, children were tested individually in a quiet and dimly lit room, and images were projected on a screen.

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′ × 7′ and stimuli were presented via computer to a 44″ × 24.5″ wall-mounted plasma television screen with speakers to the sides of the screen. The other room was 14′ × 10′; visual stimuli were presented via a ceiling-mounted projector onto a 4.5′ × 4.5′ projection screen, and audio stimuli were presented from two speakers located directly below the screen.

In either location, children were seated directly facing the screen approximately 6′ away, either in highchairs with their caregivers seated behind them or on their caregivers' laps while the caregivers wore visors. Caregivers were asked to refrain from talking or offering any form of encouragement to the children. During the experimental session, if children solicited their caregivers' 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.

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, counterbalanced within conditions and 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. 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 the actual trial.

Table 5 captures a representative trial sequence. The trial began with a four-second introduction in which participants were shown a white screen and heard the female speaker invite them to look at some objects. At this point, 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. Following the introduction, each trial consisted of three distinct phases in which the participants were shown a series of objects labeled by the novel adjective: *familiarization*, *contrast*, and *test*. Participants in all experimental conditions saw the same objects. What varied was what the participants heard (i.e., how the objects were labeled).

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 relative GA (e.g., *tall*) and a maximal GA (e.g., *clear*). This method of presentation was designed to facilitate comparison of the objects and abstraction of the common properties (cf. Booth & Waxman, 2009; Kovack-Lesh & Oakes, 2007; Namy & Gentner, 2002). A two-part *contrast* phase then followed. Here, participants were first shown a distractor object that instantiated the opposite properties of those seen in the familiarization phase (e.g., short and opaque vs. tall and clear) 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 target objects from the familiarization phase. Intonation in this segment of the contrast phase was controlled so that contrastive focus was placed on the adverb, when present, or on the copula *is* in the 'no adverb' condition.

TABLE 5
Representative Trial Sequence Used in Experiment 1

Introduction	*Familiarization Phase*	*Contrast Phase*		*Test Phase*
4 sec	18 sec (6 s, 6 s, 6 s)	7 sec (distractor)	7 sec (target)	12 sec (4 s, 8 s)
(blank screen)	2 tall, clear jars 	short, opaque jar 	tall, clear jar 	tall, opaque jar short, transparent jar 
Let's look at some things that are *pelgy*!	Look! These are both {adverb / ø} *pelgy*! This one is {adverb / ø} *pelgy*, and this one is {adverb / ø} *pelgy*!	Uh oh! This one is not *pelgy*!	Yay! This one {is completely / is very / is ø} *pelgy*!	Look! They're different! Which one is *pelgy*?

Finally, during the *test* phase, two images related to those seen during the familiarization and contrast phases were displayed simultaneously on either side of a video screen. One object could be described with the relative GA from familiarization but not the maximal GA (e.g., it was tall but not clear), while the other could be described with the maximal GA but not the relative 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 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. 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).

Coding

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 frame by frame by a trained experimenter using the SuperCoder software (Hollich, 2003). The sound was removed to ensure that the coder, who was blind to the experimental condition, only coded the direction of visual fixation during the test phase. Following previous studies (cf. Booth & Waxman, 2009; Bunger, 2006; Waxman et al., 2009), we targeted two same-size windows of the test phase (*baseline* and *response*) for analysis. Each window lasted 60 frames, or 2 seconds.[5] The first window, *baseline*, began 10 frames (approximately 333 ms) after the onset of "*Look! They're different*" and was designed to assess child participants' baseline attention. The second window, *response*, began 10 frames after the onset of the novel adjective and was designed to assess the interpretation participants assigned to the adjective.[6]

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 relative GA (the relative GA object) out of the total time spent looking at either object. This information was averaged across individual trials for each child (excluding trials in which children were inattentive for more than 30% of the time),

[5]The windows lasted long enough to get a snapshot of children's decisions—as indicated by their gazes—during the relevant sections of the test phase. Shortly after making their choice, children generally looked elsewhere, so these looking patterns were not informative to the study. Other researchers employing the intermodal preferential looking paradigm have also restricted their analysis of eye gaze to much smaller timeframes within the test phase. Halberda (2006) targeted two-second windows following the onset of the target word. Booth and Waxman (2009) targeted one-second windows within the test phase to track the divergence of experimental conditions and the time course of infants' mapping of novel words. Having divided their test phase into four two-second intervals, Gertner et al. (2006) observed that two-year-olds exhibited a preference for the video screen matching the target utterance within the first two seconds. Finally, using the "looking while listening" paradigm, Thorpe and Fernald (2006) also restricted the range of time for interpretation, under the assumption that reactions after a certain period of time could be considered noise.

[6]This short lag time between the beginning of the lexical information and the beginning of the window allows time for the participants to process the relevant lexical information. Previous psycholinguistic research on spoken word recognition has demonstrated that both adults (cf. Marslen-Wilson, 1987; Marslen-Wilson & Zwitserloood, 1989) and children as young as 18 months (cf. Fernald et al., 1998; Swingley, Pinto, & Fernald, 1999; Fernald, Swingley, & Pinto, 2001) recognize and respond to speech stimuli before the offset of a target word, approximately 300 ms into the speech signal. The timeframe of 333 ms after the onset of the lexical item for online processing of speech has also been used by, for example, Thorpe and Fernald (2006) with 24-month-olds.

then across children in each condition.[7] One experimenter was designated as the primary coder for each of the infant videos. A second experimenter independently coded the test phase portion of five randomly-selected videos in each of the experimental conditions of both experiments. There was 96% agreement between coders.

Predictions

Since the only difference between the experimental conditions is the presence and type of adverbial modifier, any differences observed across experimental conditions should be attributed to the interpretation participants assigned to the novel adjective, driven by their knowledge of the semantic consequences of the adverb's selectional restrictions and/or conditional probabilities of adverbial modification. Thus, if children are able to use different adverbs as a cue to the meaning of the novel adjectives they modify, then we predict the following.

First, children in the "no adverb" condition will show no difference between the baseline and response windows of the test phase, since there is no adverb present to constrain the interpretation of the novel adjective. Second, children in the two adverb conditions will display differences in looking time that differ significantly between the two test phase windows and also that differ from each other. Now, because it is not clear what factors control attention during the initial portion of the test phase (the baseline window), we are abstracting away from any variability across conditions by only analyzing deflections from the baseline window, measured as a difference in looking time between the baseline and response windows. Measuring looks to the object that can be described by the relative GA property highlighted during familiarization (e.g., the tall, solid container), hereafter the "relative GA object," children in the *very* condition should show a significant *increase* in looks from the baseline to the response window, while children in the *completely* condition should show a significant *decrease* in looks to this object from baseline to the response window.

RESULTS

The results, captured in Figure 1, bear out our predictions. Not only does the *presence* of an adverb matter, but the *kind* of adverb matters as well. (Note that this figure incorporates results from all conditions across both Experiments 1 and 2. Here we concentrate on the three leftmost pairs of bars.)

Children in the "no adverb" condition showed no evidence of learning a meaning for the novel adjective; they exhibited no difference in looking time between the *baseline* and *response* windows (Wilcoxon signed-rank test: $W = 22$, $p = 0.17$, one-tailed, $r = .20$). Children in both of the adverb conditions, however, pulled away from this baseline in the response window in opposite directions, suggesting that the meaning assigned to the novel adjective differed in the two adverb conditions (*very*: $W = -52$, $p = 0.01$, $r = -.49$; *completely*: $W = 44$, $p = 0.03$, $p = .41$). This difference in looking time between the two windows held for the majority of children (9 out of 11) in both of the two adverb conditions, regardless of their overall level of vocabulary production, an indication that the effect was not driven by overall vocabulary level

[7] See Bunger (2006) for a similar approach for data inclusion/exclusion.

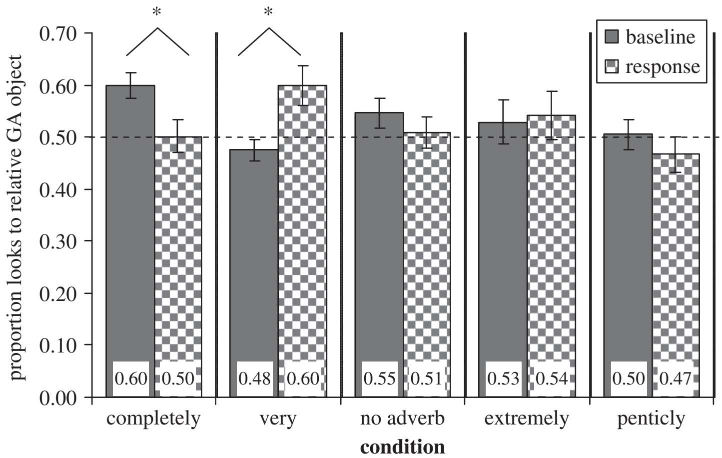


FIGURE 1 Children's responses as measured by comparing the average time spent looking at the relative GA object during key sections of the baseline and response windows of the test phase (Experiments 1 and 2).

but rather by children's knowledge of these specific lexical items. The asymmetry between the two adverb 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 relative GA object during the *response* window, the other adverb (*completely*) pulled their attention *away* from it.

Children in the *very* condition devoted equal attention to the two objects during the *baseline* window but spent more time looking at the relative GA object during the *response* window. This significant increase in looking time to the relative GA object (from baseline to response) suggests that the children assigned a relative meaning to the adjective when it was modified by *very*. This behavior fits naturally with the results of the corpus analysis. While this intensifier has the potential to appear with a wide range of adjectives, it is much more likely to modify those without a maximal endpoint (i.e., nonmaximal GAs and, more specifically, relative GAs). Hence, children recruited the distributional features of the adverb in assigning a meaning to the novel adjective.

Children in the *completely* condition showed a very different pattern. These children showed a significant decrease from baseline to response in their looking time to the relative GA object. This suggests that they mapped the adjective to an absolute maximal (not relative) meaning. This behavior also fits naturally with the results of the corpus analysis. The adverb *completely* occurs predominantly with maximal GAs. Hence, children recruited this distributional property in assigning a meaning to the novel adjective.

A 2 × 3 repeated-measures analysis of variance was performed over the percentage of time spent looking at the relative GA object with condition (*completely*, *very*, no adverb) as the between-subject factor and test phase window (*baseline*, *response*) as the within-subjects factor. While the analysis revealed no overall main effects of condition ($F(2, 32) = 1$, $p = 0.38$, $\eta^2 = .43$) or window ($F(1, 33) = 0$, $p = 1.0$), there was a significant condition × window interaction

($F(2, 30)= 7$, $p = 0.003$), driven by the two different patterns in the adverb conditions. Planned comparisons of the baseline and response windows for each condition reveal significant main effects for both the *very* condition ($F(1, 10) = 7.13$, $p < 0.03$) and the *completely* condition ($F(1, 10) = 5.81$, $p < 0.04$), but no significant main effect for the no adverb condition ($F(1, 10) = 1.33$, $p = 0.28$).

DISCUSSION

In this experiment, we sought to determine whether young word learners recruit adverbial modification as a cue to the semantic structure of a novel adjective. The results suggest they do. Children who hear a novel adjective that is not modified by an adverb show no systematic preference for one or another object property, suggesting that they were unable to assign a meaning to the novel adjective in the absence of adverbial modification. In contrast, children who hear the same novel adjective modified either by the proportional modifier *completely* or by the intensifier *very* display a very different pattern—the former apparently mapping it onto a maximal GA meaning and the latter mapping it onto a relative GA meaning.

However, there may be an alternative explanation for the set of results we obtained in Experiment 1. While *completely* and *very* differ in terms of their selectional restrictions, they also differ with respect to their lexical frequency. In the spoken BNC, *very* has a frequency of 25,041, or 2421.36 instances per million words. By contrast, *completely* has a frequency of only 822, or 79.48 instances per million words, making *very* more than 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*. However, there was no statistical difference in the performance between children who were reported to be producing *very* and those who were reported to be producing no adverbs (two-tailed t-test ($t(30) = 1.33$, $p = 0.19$). It is possible that this frequency asymmetry between *completely* and *very* is somehow responsible for the asymmetry in response patterns.

Imagine, for example, that the children knew nothing about the meaning of *completely* and only recognized it as a novel adverb. It is possible, then, that their response pattern in the *completely* condition merely reflects the pattern that derives from hearing a novel adverb. Children might have responded appropriately in the *very* condition because they knew the semantic consequences for adjectives modified by that adverb. At the same time children might have responded appropriately in the *completely* condition just because they did the opposite of what they would have done with a known adverb like *very*. Now, not only did children in the *completely* condition show a significant decrease from baseline to response, but also their looks to the relative GA object actually began above chance in the baseline window. (We reflect on this pattern of looks in the Discussion section of Experiment 2.) If children in this condition were simply driven by the novelty of the adverb *completely*, we would expect that they would pattern similarly given any low-frequency or novel adverb, regardless of its semantics. In Experiment 2, we put this hypothesis to the test by examining what children do with an additional low-frequency adverb and a novel adverb.

EXPERIMENT 2

The purpose of Experiment 2 was to determine whether a low-level feature, namely (relative) lexical frequency, was responsible for the pattern of results we saw in Experiment 1. Here we introduced two new adverbial modifiers to address this question. One—*extremely*—shares similar selectional restrictions and distributional tendencies with that of the high-frequency intensifier *very*, but like *completely* has a low lexical frequency (and in fact, has an even lower frequency). The other—the novel adverb *penticly*—is phonologically similar to *completely* but, as a novel adverb, has no lexical frequency. A summary of these features is presented in Table 6.

If the pattern of responses we observed in the *completely* condition in Experiment 1 can be attributed to a sheer novelty effect, driven by its simply being distinct from the high-frequency intensifier *very*, then we expect to see that the results of the two new adverb conditions in Experiment 2 parallel those of the *completely* condition in Experiment 1. If, however, the pattern we observed in Experiment 1 was actually due to the *meaning* of *completely*, then we should see a difference between the *completely* condition and these two new conditions.

Under this hypothesis, children in the *penticly* condition should pattern no differently between the two test phase windows, as did children in the no adverb condition. The pattern of behavior in the *extremely* condition, however, will depend on whether they know the meaning of this adverb. If they *do*, they should pattern with the children in the *very* condition of Experiment 1. If they *do not,* they should pattern with the children in the *penticly* condition in this experiment and the no adverb condition of Experiment 1. To be clear, a lack of effect in Experiment 2 would be a welcome result, because it would indicate that children in the *completely* condition in Experiment 1 actually appealed to information specific to this adverb when assigning a meaning to the novel adjectives it modified. That is, the argument depends not simply on the lack of effect in Experiment 2 but also on the comparison of response patterns between Experiments 1 and 2.

METHOD

Participants

The participants were 22 children (11 girls, 11 boys; M: 29 months 8 days; range: 28;2 to 31;3 months). An additional 12 children were excluded because of fussiness (n = 4), inattentiveness (as described in Experiment 1, n = 6), or equipment error (n = 2). Children's average

TABLE 6
Adverbs Used in Experiments 1 and 2

Adverb	*Frequency*	*Instances per Million*	*Semantics*
very	25,041	2421.36	intensifier
completely	822	79.48	maximally closed scale selecting
extremely	480	46.41	intensifier
penticly	n/a	n/a	n/a

vocabulary production was 556 words. As in Experiment 1, age, gender, and vocabulary production scores were balanced within and across conditions: *penticly* (5 girls 6 boys, M: 30 months; avg. vocabulary: 439), *extremely* (6 girls 5 boys, M: 30 months; avg. vocabulary: 556). Paired t-tests revealed no significant difference in vocabulary production between the participants in the two experiments. Fifty adult controls (Northwestern University undergraduates fulfilling an experimental requirement for a Linguistics course) also participated, completing 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.

Materials and Procedure

The materials and procedure were identical to Experiment 1, with the exception of the two new adverbs: the low-frequency intensifier *extremely* and the novel adverb *penticly*. Children were randomly assigned to one of these two experimental conditions, as in Experiment 1. Coding of children's eye gazes was also conducted as in Experiment 1.

Adult participants were also randomly assigned to one of five experimental conditions from Experiments 1 and 2. Adults were told they were participating in an experiment designed for young children. They viewed the same videos as the children did and were run in groups of one to three in a $9' \times 10.5'$ 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 feet from the TV. They were each given a clipboard and a response packet and 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. 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 was the overall percentage of times adults chose the relative GA object.

RESULTS AND DISCUSSION

Children

Recall that the response patterns from Experiments 1 and 2 are presented in Figure 1. Unlike in Experiment 1, where the responses for children in both of the adverb conditions diverged from the no adverb baseline (the two leftmost sets of bars), the responses for children in the *extremely* and *penticly* conditions (the two rightmost sets of bars) patterned no differently from the 'no adverb' condition in Experiment 1 (the middle set of bars). There was no difference between the baseline and response window for either condition (*extremely*: $W = -2$, $p = 0.47$; *penticly*: $W = 26$, $p = 0.13$, one-tailed), and neither condition resembled the pattern observed in the *completely* condition in Experiment 1.

A 2×2 analysis of variance was performed over the percentage of time spent looking at the relative 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 condition ($F(1, 21) = 1.5$, $p = 0.23$, $\eta^2 = .65$), no main effect of window ($F(1, 22) = 0$, $p = 1.0$), and no interaction $F = 1$, $p = 0.33$. A one-way analysis of variance run on the difference in looking time between the *baseline* and *response* windows in the five conditions in Experiments 1 and 2 revealed a significant main effect ($F(4, 50) = 3.75$, $p < 0.01$). A post hoc Tukey's HSD analysis ($p < .01$) revealed that this effect was driven by a difference between the *very* and *completely* conditions.

These analyses show that the appropriate interpretation of these data requires a comparison between the two windows for each condition. What matters is whether the experimenter's question about the meaning of the novel word drew children'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 in opposite directions. This claim is supported graphically in Figure 2, which complements in illustrating the difference between the *baseline* and *response* windows among the conditions.

We would like to take some space here to reflect on the eye-gaze patterns in two different conditions. First, consider the baseline window in the *completely* condition of Experiment 1. Unlike every other condition, children in this condition began by looking to the relative GA object more often than predicted by chance ($t(10) = 3.47$, $p < 0.01$), then switched their gaze away from it in the response window. What could account for this pattern? We think the most likely possibility is that it was driven by 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

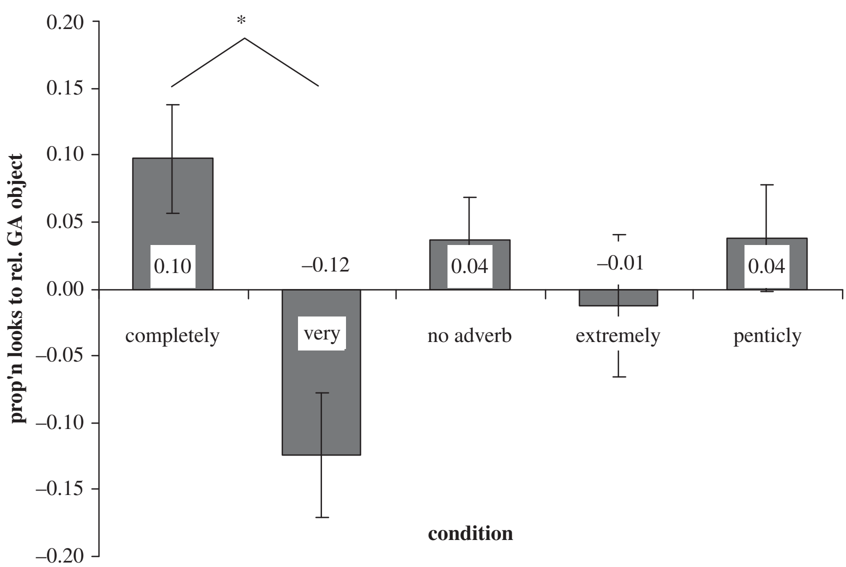


FIGURE 2 Children's responses as measured by the difference between key sections of the baseline and response test phase windows of the test phase.

completely. That it is not simply a novelty preference based on the objects alone is reflected in the lack of such a pattern in the 'no adverb' and *penticly* conditions.[8]

Given that *completely* selects for a maximal endpoint, at the beginning of the test phase children are drawn to inspect an object that lacks the corresponding property. Upon hearing the experimenter's question containing the novel adjective—whose meaning they assigned earlier in the trial based on the adverb—they switch their gaze away from the relative GA object to the other object. Thus, it is precisely *because* of the meaning of this adverb that children respond this way, and do not do so in any of the other conditions. While one might expect children in the *very* condition to pattern in the opposite direction, their looking pattern is easily explained by appealing to the corpus data: because this intensifier has the potential to appear with a wide range of adjectives, children await hearing the novel adjective before directing their gaze to one object or the other.

Next, consider the comparison between the baseline and response windows in the *extremely* condition of Experiment 2. Unlike the adverb conditions in Experiment 1, in the *extremely* condition, children did not pattern differently between the *baseline* and *response* windows. Thus, neither the semantic representation nor the distribution of this adverb guided children's responses. 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 corpus data, 390 vs. 10, respectively), it is simply too infrequent for children to have formed a hypothesis about its selectional restrictions. Given this pattern, it follows that the asymmetry between *very* and *completely* in Experiment 1 was not due to the relative difference in frequency between them, but rather to children having acquired appropriate representations for both of these adverbs.

Adults

Adults patterned in the predicted direction, choosing the relative GA object the least often when the novel adjective was modified by *completely* (see Table 7). Interestingly, for one of the five trials adults never selected the relative GA object (the long and curly object rather than the short and straight one), regardless of their condition. Excluding this trial from the analysis results in

TABLE 7
Percentage of Time Adults Chose the Relative GA Object

	All Trials	*4 of 5 Trials*
completely	18	23
very	34	43
no adverb	42	53
extremely	56	70
penticly	54	65

[8]See Waxman and Markow (1995) for discussion of the effects of a novel word on novelty preferences and Matsuo et al. (2008) for a potentially similar meaning-driven above-chance looking pattern during baseline.

a similar trend of percentages. Two-tailed t-tests ($df = 9$) on these data revealed that only the *completely* and *extremely* conditions are significantly different than chance (though in opposite directions from each other), and the difference between the *penticly* condition and chance is marginally significant (*completely*: $t = 3.48$, $p = 0.007$; *very*: $t = 1.15$, $p = 0.28$; no adverb: $t = -0.36$, $p = 0.73$; *extremely*: $t = -4$, $p = 0.003$; *penticly*: $t = -1.97$, $p = 0.08$).

The fact that adults resisted selecting the relative GA object when the novel adjective was modified by the intensifier *very* may reflect the fact that as a result of its widespread distribution and high frequency, this intensifier has a diluted meaning for adults (see Note 3). By contrast, modification of an adjective by *completely* has a clear implication for adults: the adjective most likely corresponds to a maximal GA interpretation. Modification by the lower frequency intensifier *extremely,* which may also indicate a more sizable divergence from the standard of comparison than *very* does, signals a relative GA interpretation. That is, although *very* is more likely to modify adjectives with a relative 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 relative GA interpretation. Adults may also have been economical in their decision. Relative GAs by definition encode the context in their semantic representation. Adults may in general hold off on assigning a novel adjective an interpretation that is not linked to the context unless they have good reason to do so (Kennedy, 2007). Thus, when given a choice between a relative GA and absolute GA interpretation, they might opt for the latter, because its standard is not context-dependent.

GENERAL DISCUSSION

The current research contributes to our understanding of how children classify novel adjectives according to their abstract semantic representations. First, the results of a corpus analysis demonstrate that robust patterns of adverb-adjective bigrams exist in natural language, and may provide children with evidence as to how to sort adverbs according to their selectional restrictions, and adjectives according to their scalar structure. Specifically, proportional modifiers such as *completely*, which select for a scale with endpoints, are likely to be paired with maximal GAs, which map objects onto a maximally closed scale, while intensifiers such as *very*, which indicate divergence from the standard of comparison, are drawn to relative GAs such as *big*, which map objects onto an open scale and have a contextually based standard of comparison.

Second, the results of our word-learning experiments are consistent with the view that 30-month-olds are 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 relative GA and another best described by a maximal GA and heard the objects labeled by a novel adjective. This adjective was modified by either no adverb, or one of four different adverbs varied by their semantics and lexical frequency. At test, when the object properties were teased apart and children were asked about the interpretation they assigned to the adjective, they were more likely to turn their attention *toward* the relative GA object if the adjective had been modified by *very*, but *away* from this object and toward the maximal GA object if the adjective had been modified by *completely*. Children in the no adverb and two control adverb conditions (*extremely* and *penticly*) remained at chance. 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 children who are reported to not yet be producing these adverbs seem to have capitalized on their distribution in the language environment to acquire something about their meaning, and moreover to use this information as a cue to the meaning of their adjectival arguments.

We might take our conclusion one step further and suggest that adverbs may be doing for adjectives what 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 (cf. Fisher, Gleitman, & Gleitman, 1991). It has been argued that language learners may 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) in order to reach conclusions about that verb's meaning. In parallel fashion, learners may 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 order to reach conclusions about adjectival meaning. The process of acquisition may be a reciprocal word learning process, whereby children use what they have learned about adverbs to infer something about the meaning of the adjectives they modify, and at the same time use what they have learned about adjectives to infer something about the selectional restrictions of the adverbs that modify them.

Learners must have some prior knowledge (be it language-specific knowledge or knowledge about more general cognitive representations) 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 bring some expectations about the range of possible adjectival subclasses to bear in using the distributional features of the exposure language as a cue to adjective meaning. While it is not clear from our results whether this expectation is innate or has simply developed by the time children begin tracking the distributional information that enables them to succeed at our task, what is clear from our results is that by 30 months of age children know that adjectives can vary in their scalar structure and they expect this variability to leave a distributional signature in the exposure language. In English, one place where this distributional signature can apparently be found is in adverbial modification. Moreover, it appears that children can utilize this distributional signature as a cue to the meaning of a novel adjective.

ACKNOWLEDGMENTS

This work was made possible by funding from a Northwestern University Presidential Fellowship to the first author, a NSF grant BCS-0418309 and NIH grant DC006829 to the second author, and a NIH grant HD30410 for the Project on Child Development to Sandra Waxman. We are grateful to the parents and children who came to our laboratory to participate and to Rebecca Baier and Christy Call for their technical assistance and help in coordinating participants. This work benefited from many insightful discussions with Sandra Waxman, comments from members of the Northwestern Linguistics-Psychology Acquisition Lab (Joshua Viau, Ann Bunger, and Elisa Sneed German, in particular), and feedback from the audiences at the 2006 Midwest Semantics

Workshop, a 2006 CNL lunch a the University of Maryland, the 2007 Boston University Conference on Language Development, the 2008 Annual Meeting of the Linguistic Society of America, and the 2008 Language Acquisition Workshop at Princeton University. Finally, we thank two anonymous reviewers and Letitia Naigles, whose comments greatly improved the quality of our analysis and presentation.

REFERENCES

Alishahi, A., & Stevenson, S. (2005a). The acquisition and use of argument structure constructions: A Bayesian model. In *Proceedings of the second workshop on psycho-computational models of human language acquisition* (pp. 82–90), Ann Arbor, Michigan.

Alishahi, A., & Stevenson, S. (2005b). A probabilistic model of early argument structure acquisition. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 97–102), Stressa, Italy.

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

Bartsch, R., & Vennemann, T. (1972a). *Semantic structures*. Frankfurt, Germany: Athenäum.

Bartsch, R., & Vennemann, T. (1972b). The grammar of relative adjectives and comparison. *Linguistische Berichte*, *20*, 19–32.

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

Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer* (Version 4.5.14) [Computer program]. Retrieved from hhtp://www.praat.org/.

Bogusławski, A. (1975). Measures are measures: In defense of the diversity of comparatives and positives. *Linguistische Berichte*, *36*, 1–9.

Booth, A. E., & Waxman, S. R. (2003). Mapping words to the world in infancy: Infants' expectations for count nouns and adjectives. *Journal of Cognition and Development*, *4*, 357–381.

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

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/

Brown, R. (1957). Linguistic determinism and parts of speech. *Journal of Abnormal and Social Psychology*, *55*, 1–5.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Bunger, A. (2006). *How we learn to talk about events: Linguistic and conceptual constraints on verb learning*. Doctoral dissertation, Northwestern University.

Clark, R. L. (1970). Concerning the logic of predicate modifiers. *Noûs*, *4*, 311–335.

Cruse, D. A. (1980). Antonyms and gradable complementaries. In D. Kastovsky (Ed.), *Perspektiven der lexikalischen semantik: Beitraäge zum wuppertaler semantikkolloquium vom* 2–3 (pp. 14–25). 1977. Bonn: Bouvier, Verlag.

Ebeling, K. S., & Gelman, S. A. (1988). Coordination of size standards by young children. *Child Development*, *59*, 888–896.

Ebeling, K. S., & Gelman, S. A. (1994). Children's use of context in interpreting big and little. *Child Development*, *65*, 1178–1192.

Fara, D. G. (2000). Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics*, *28*, 45–81. (Originally published under the name "Delia Graff")

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *58*, 228–231.

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, *72*, 1003–1015.

Fisher, C. (2002). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, *5*, 56–65.

Fisher, C., Gleitman, L. R., & Gleitman, H. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, *23*, 331–392.

Gelman, S. A., & Ebeling, K. S. (1989). Children's use of nonegocentric standards in judgments of functional size. *Child Development*, *60*, 920–932.

Gelman, S. A., & Markman, E. M. (1985). Implicit contrast in adjectives vs. nouns: Implications for word-learning in preschoolers. *Journal of Child Language*, *6*, 125–143.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules. *Psychological Science*, *17*, 684–691.

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

Goldberg, A. E., Casenhiser, D., & Sethuraman, N. (2005). The role of prediction in construction-learning. *Journal of Child Language*, *32*, 407–426.

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.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*, 310–344.

Hall, D. G. (1991). Acquiring proper names for familiar and unfamiliar animate objects: Two-year-olds' word learning biases. *Child Development*, *62*, 1142–1154.

Hall, D. G. (1994). Semantic constraints on word learning: proper names and adjectives. *Child Development*, *65*, 1299–1317.

Hall, D. G., & Graham, S. A. (1999). Lexical form class information guides word-to-object mapping in preschoolers. *Child Development*, *70*, 78–91.

Hall, D. G., & Moore, C. E. (1997). Red bluebirds and black greenflies: Preschoolers' understanding of the semantics of adjectives and count nouns. *Journal of Experimental Child Psychology*, *67*, 236–267.

Hall, D. G., Waxman, S. R., & Hurwitz, W. (1993). How 2- and 4-year-old children interpret adjectives and count nouns. *Child Development*, *64*, 1651–1664.

Hollich, G., Rocroi, C., Hirsh-Pasek, K., & Golinkoff, R. (1999). *Testing language 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.

Kamp, H. (1975). Two theories about adjectives. In E. Keenan (Ed.), *Formal semantics of natural languages* (pp. 123–155). Cambridge, MA: Cambridge University Press.

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, *57*, 129–191.

Katz, N., Baker, E., & Macnamara, J. (1974). What's in a name? A study of how children learn common and proper names. *Child Development*, *45*, 469–473.

Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Outstanding dissertations in linguistics. New York, NY: Garland Press. (1997). Doctoral dissertation, University of California, Santa Cruz.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*, 1–45.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable adjectives. *Language*, *81*, 345–381.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, *4*, 1–45.

Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorization. *Journal of Experimental Child Psychology*, *98*, 69–93.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates. Available from http://childes.psy.cmu.edu/

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken-word recognition. *Cognition*, *25*, 71–102.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576–585.

Matsuo, A., Kita, S., & Naigles, L. (2008). *Japanese two-year-olds use morphosyntax to learn verb meanings*. Poster presented at the XVIth Biennial International Conference on Infant Studies (ICIS), March 27–29, Vancouver, Canada.

Montague, R. (1974). English as a formal language. In R. H. Thomason (Ed.), *Formal philosophy: Selected papers of Richard Montague* (pp. 188–221). New Haven, CT: Yale University Press.

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

Naigles, L., Fowler, A., & Helm, A. (1992). Developmental changes in the construction of verb meanings. *Cognitive Development*, *7*, 403–427.
Naigles, L., & Hoff-Ginsberg, E. (1995). Input to verb learning: Evidence for the plausibility of syntactic bootstrapping. *Developmental Psychology*, *31*, 827–837.
Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*, 5–15.
Paradis, C. (1997). *Degree modifiers of adjectives in spoken British English*. Lund, Sweden: Lund University Press.
Parsons, T. (1972). Some problems concerning the logic of grammatical modifiers. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 127–141). Dordrecht, The Netherlands: Reidel.
Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics*, *12*, 259–288.
Rusiecki, J. (1985). *Adjectives and comparison in English*. *Longman Linguistics Library* (Vol. 31). London: Longman Group Limited.
Siegel, M. E. A. (1979). Measure adjectives in Montague grammar. In S. Davis & M. Mithun (Eds.), *Linguistics, philosophy, and Montague grammar* (pp. 223–262). Conference on Montague grammar, philosophy, and linguistics, State University of New York at Albany. Austin, TX: University of Texas Press.
Smith, L. B., Cooney, N., & McCord, C. (1986). What is *high*? The development of reference points for *high* and *low*. *Child Development*, *57*, 583–602.
Soja, N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's inductions of word meanings: Object terms and substance terms. *Cognition*, *38*, 179–211.
Spelke, E. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, *15*, 626–636.
Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*, 103–114.
Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*, 73–108.
Syrett, K. (2007). *Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.
Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics*, *27*(1), 1–35.
Taylor, M., & Gelman, S. A. (1988). Adjectives and nouns: Children's strategies for learning new words. *Child Development*, *59*, 411–419.
Thorpe, K., & Fernald, A. (2006). Knowing what a novel word is not: Two-year-olds "listen through" ambiguous adjectives in fluent speech. *Cognition*, *100*, 389–433.
Unger, P. K. (1975). *Ignorance: A case for skepticism*. Oxford, UK: Clarendon Press.
von Stechow, A. (1984). Comparing semantic theories of comparison. *Journal of Semantics*, *3*, 1–77.
Wagner, L. (2006). Aspectual bootstrapping in language acquisition: Telicity and transitivity. *Language Learning and Development*, *2*, 51–76.
Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*, 217–242.
Waxman, S. R., & Markow, D. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*, 257–302.
Waxman, S. R., Lidz, J., Braun, I., & Lavin, T. (2009). Twenty-four-month-old infants' interpretation of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*, 67–95.
Woodward, A. (2000). Constraining the problem space in word learning. In *Becoming a word learner: A debate on lexical acquisition* (pp. 81–114). In M. Marschark (Ed.), *Counterpoints: Cognition, memory, and language* series. Oxford, UK: Oxford University Press.