NORTHWESTERN UNIVERSITY

Learning about the Structure of Scales:

Adverbial Modification and the Acquisition of the Semantics of Gradable Adjectives

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Kristen L. Syrett

EVANSTON, ILLINOIS

December 2007

ABSTRACT

Learning about the Structure of Scales:

Adverbial Modification and the Acquisition of the Semantics of Gradable Adjectives

Kristen L. Syrett

This work investigates children's early semantic representations of gradable adjectives (GAs) and proposes that infants perform a probabilistic analysis of the input to learn about abstract differences within this category. I first demonstrate that children as young as age three distinguish between relative (e.g., *big*, *long*), maximum standard absolute (e.g., *full*, *straight*), and minimum standard absolute (e.g., *spotted*, *bumpy*) GAs in the way that the standard of comparison is set and how it interacts with the discourse context. I then ask if adverbs enable infants to learn these differences. In a corpus analysis, I demonstrate that statistically significant patterns of adverbial modification are available to the language learner: restricted adverbs (e.g., *completely*) are more likely than non-restricted adverbs (e.g., *very*) to select for maximal GAs with bounded scales. Non-maximal GAs, which are more likely to be modified by adverbs in general, are more likely to be modified by a narrower range, predominantly composed of intensifiers (e.g., *very*). I then ask if language learners recruit this information when learning new adjectives. In a word learning task employing the preferential looking paradigm, I demonstrate that 30-month-olds use adverbial modifiers they are not necessarily producing to assign an interpretation to novel adjectives. Adjectives modified by *completely* are assigned an interpretation corresponding to an absolute property, while adjectives modified by *very* correspond to a relative property. Infants presented with an adjective modified by no adverb, a

novel adverb (*penticly*), or a low-frequency intensifier (*extremely*) pattern at chance. I argue that a form-meaning correspondence similar to the one discussed in verb learning is active in adjective learning. Infants are guided by their conceptual representations when attending to the distributional patterns of adverb-adjective bigrams in the exposure language, and expect that these surface-level cues will partition the class of GAs according to differences in scalar structure. I finish by connecting the structure of scales and paths in linguistic representations, and suggest that given that infants attend to the structure of paths when parsing events, it is optimistic to expect that they might also attend to the structure of scales when attending to properties.

Acknowledgements

Above all, I am deeply grateful to Stan, who has been endlessly supportive and thoughtful all along. More than I can express, I am also thankful to Anna Sophia, for bringing great joy into my life as I witness all of her newest discoveries and developments every day, and, of course, to Maggie. For years of encouragement and love, I thank my Grandmother and, most especially, my Mom, who should have been here to share this with me.

I was fortunate to have chosen Northwestern University for graduate school. In many ways, it was the perfect fit for me. I have had a fabulous experience with my committee members, especially my two advisors Jeff Lidz and Chris Kennedy, and Sandy Waxman. Their insights, encouragement, sense of humor, and rigorous intellectual standards have been welcome over the years, and have influenced me tremendously. Even with the physical distance between us at the end, Jeff and Chris bent over backwards to make sure everything went as smoothly as possible and were extremely generous with their time and feedback. Stefan Kaufmann also provided very helpful feedback, from which this document benefited a great deal.

A very special thanks goes to the students who formed the core of our language acquisition group, who have provided years of friendship, good discussion, delicious food, amusement at conferences, and generous help with experiments – in particular, Ann Bunger, Erin Leddon, Elisa Sneed German, and most especially Josh Viau. Other students working on acquisition during this time – Flo Anggoro, Robert Daland, Jessica Peterson Hicks, Heather Norbury, Tom Piccin, and Liza Ware – offered a much welcomed fresh perspective to discussions. I am also thankful to Catherine Anderson for being such a good officemate and to the rest of my cohort, Yongeun Lee, James German, and Lewis Gebhardt, for enlightening conversations and camaraderie in the early years of graduate school.

I have also been lucky to be in a department where there has been a great deal of creativity and collaboration among students and faculty, and eagerness to communicate across disciplines. I am thankful to both Gregory Ward and Bob Gundlach for being such great advocates for both the Department and the students in their role as department chair. I have benefited from participating in Linguistics-Psychology Acquisition lab meetings, Syntax-Semantics Lab meetings, the Chicago Syn-Sem Circle, journal communes, and reading groups (in all of their various forms). I am deeply grateful that I was welcomed into the homes of joyful and loving families in the department, from whom I learned an immeasurable amount about parenting and balancing priorities in life. Thanks especially to Chris, Hillary, Julian, and Sterling; Jeff, Tonia, and Sammy; Ann, Glen, Louisa, and Evan; and Laura, Mike, Nolan, and Alana.

A Northwestern University Presidential Fellowship made all the difference in my graduate school experience. It provided me with two years of funding for my dissertation work and supplemental travel funding, but also allowed me to discover what life is like in a wide range of other disciplines beyond my own, learn about administrative decisions that affect both students and faculty, and come to appreciate the common issues that arise in the pursuit of answering questions through research. I am especially thankful for conversations with Dean Andrew Wachtel, Simon Greenwold, Rhiannon Stephens, Ben Ponder, Alan Brothers, Chris Campbell, Carmen Niekrasz, Rana Ozbal, and John Pham. Mary Pat Doyle in The Graduate School has also been wonderfully kind and supportive.

Beginning with a Bloch fellowship, I became involved in the Linguistic Society of America. I cannot say enough how beneficial this fellowship was, how invaluable my interactions with senior linguists in the field of Linguistics have been, how much I have learned about the discipline through my involvement, and how worthwhile the Institutes I attended were. I am especially indebted to Maggie Reynolds, for being a tireless model of poise and professionalism, and to Gregory Ward, for his unflagging energy, vision, and refreshing wit.

Invaluable assistance running subjects, coding data, and preparing experimental stimuli was provided by Rebecca Baier, Stefanie Brody, Evan Bradley, Irena Braun, Christy Call, Jessica Mershon, Shyaam Ramkumar, Carol Sweeney, Adriana Weisleder, Melissa Baese, Jessica Bobula, and Anne Gooch. There would, of course, be no experimental results had there not been children who were willing to share their knowledge with me, parents who were willing to bring their children into the lab, or trusting parents and staff at area preschools, in particular Warren W. Cherry Preschool, Chiaravalle Montessori, Northbrook Community Nursery School, and the School for Little Children.

Other thanks go to my family and friends, for their support and advice. Thanks to Shirley and Tim for their limitless kindness; to Tierra, who I can always count on for uplifting words and entertaining anecdotes; to Dani for her spirit; to Joslyn for the unexpected; to Jamesie, Tricia, Jess, and Sean for many memories and taking such good care of Stan in Ann Arbor; to the Waltons for being such a warm extended family; and to Melissa, Noam, Sonja, and Rémi for semi-weekly diversions. Countless hours spent in front of the computer were made much less onerous by good background music, supplied by internet radio. Thanks also go to the coffeeshops of Andersonville, where way too many evenings were spent. A semester spent as a visiting scholar at the UMass Department of Linguistics was also extremely worthwhile. Along the way, I have had some amazing and inspiring yoga teachers, who have helped me to appreciate the interplay between balance, strength, and flexibility in many aspects of life. Namaste.

This work was presented in various forms at a number of different venues. I am grateful to audiences at the 2004 Generative Approaches to Language Acquisition of North America Conference, the LSA Annual meetings in 2005 and 2007, the 2006 CUNY Human Sentence Processing Conference, the 2006 Regional Meeting of the Chicago Linguistic Society, the 2006 Midwest Workshop on Semantics, and the 2007 Biennial Meeting of the Society for Research on Child Development. Travel to these conferences was funded in part by various grants from the Northwestern University Graduate School. Generous funding was provided by an NSF grant BCS-0418309 and an NIH grant DC006829 to Jeffrey Lidz, an NSF grant BCS-0094263 to Christopher Kennedy, and an NIH grant HD30410 to Sandra Waxman for the Project on Child Development.

Table of Contents

Table of Tables

Table of Figures

Table of Appendices

Chapter 1: Introduction

This dissertation is concerned with children's acquisition of the scalar structure of Gradable Adjectives (GAs), adjectives such as *big* and *full*, which are felicitous in comparative constructions, such as *X is bigger/fuller than Y*.[1,2] What allows these adjectives to appear in such constructions, and what makes adjectives such as *dead* seem odd in them, is that GAs and their corresponding properties in the world admit of degrees. What is especially interesting about this topic from the semantic perspective is that there are differences within this class of adjectives with respect to aspects of their meaning in the positive (non-comparative) form. For example, to say that something is *big* requires, among other things, knowing what comparison class it is a member of, in a way that saying that something is *full* does not. These differences among GAs can be captured by appealing to the structure of the scales occupied by the degrees – whether

---

[1] A note on terminology: the adjectives with which I am concerned have gone by many names in the literature: (*affirmative) non-predicative* (Kamp, 1975; Parsons, 1972); *attributive* (Ludlow, 1989; Platts, 1979; Wheeler, 1979); *comparative* (Cresswell, 1976); *degree* (Bolinger, 1972; Klein, 1980, 1982); *gradable* (Bierwisch, 1989; Kennedy, 1999, 2001, 2007; Klein, 1991; Rusiecki, 1985); and Seuren, 1978); *linear* (Klein, 1980); *measure* (Givón, 1970; McConnell-Ginet, 1973; Siegel, 1979); *measuring* (Vendler, 1968); and *relative* adjectives (Bartsch & Vennemann, 1972a, b; McConnell-Ginet, 1973). Within these classifications, there may be slight differences among the adjectives that are included, based on the specific issue at hand, how color terms are treated, and so forth. *Dimensional* adjectives are a subset of this group.

[2] Throughout this dissertation, my focus is on gradable adjectives in English. Where appropriate, I include examples from other languages in order to highlight similarities or differences in meaning or surface environments licensing their appearance. See Bhat (1994) and Dixon (1977) for a discussion of the adjective category cross-linguistically, and Kim (2002), Olawsky (2004), Wetzer (1992), and Wojdak (2001) for selected discussions of the status of adjectives in specific languages other than English.

they are open or closed (i.e., whether or not they have endpoints that serve as the standard of comparison). What is especially interesting from the language acquisition perspective is that in order for children to know the meaning of any given GA, they must assign it the right semantic representation, which requires knowing what kind of scalar structure it should be assigned. While there has been growing interest in the structure of scales – both in how they are encoded in the representation of adjectives and how they are informative about the measurement of events – and while children's acquisition of adjectives has been a focus in the word learning literature, little or no work has brought these two lines of research together and asked how they can inform each other. The aim of this dissertation is to do just that.

My interest in this topic is in determining what children know about the range of GA meanings at an early age, and identifying the process by which they came to appreciate these semantic distinctions. Specifically, I am interested in investigating whether young children understand that there are differences in the way the standard of comparison can be set for different kinds of GAs (and therefore that there are differences in the structure of scales), and what role adverbs play as a surface-level cue that allows children to assign the correct representation to a new GA. The mapping between form and meaning has had a key place in discussions of how the child can take advantage of the syntactic partitioning of lexical items into semantic classes when learning about the range of verb interpretations and the count/mass status of nouns (i.e., how sentence structure reflects transitivity, causativity, and other aspects of a verb's meaning, or how the presence and type of determiner reflects the way a nominal is measured). This work extends the discussion to the domain of adjectives, giving us a broader understanding of how surface structure informs abstract representations in word learning. At the same time, it provides empirical support for proposals in the theoretical literature concerning scalar distinctions among GAs and the way these distinctions are highlighted by adverbial modification.

Now before going further, I would like to clarify that role I am giving adverbs is not that children are actually using them to *acquire* scalar structure. Rather, they are taking advantage of cues in the input – such as adverbs – to recruit pre-existing conceptual representations in the word learning process in order to assign the right representations to new adjectives they encounter. This dissertation is not about how children *acquire* scalar structure, but rather how they learn to *classify* GAs with the correct scalar structure.

In Chapter 2, I describe how previously-proposed adjectival distinctions from the experimental literature can be reconciled with the gradable/non-gradable distinction, and clarify the focus of the current work with respect to previous investigations. I review the course of children's acquisition of adjective meaning, and then narrow the scope to two aspects of children's competency with respect to GA meaning: their ability to take contextual factors into account when making judgments of size, and their ability to allow the standard of comparison to shift from context to context. Having demonstrated that children age two to five are adept in both areas when it comes to adjectives of physical dimension, which are *relative* GAs, I ask whether this behavior is a more general reflection of how they treat the entire set of GAs (in which case, they exhibit non-adult-like behavior) or whether they recognize these features of meaning as specific to relative, but not *absolute*, GAs.

In Chapter 3, I highlight the importance of the degree argument of GAs, distinguishing them from non-gradable adjectives and allowing them to appear in a range of surface environments. I then discuss characteristics of the positive (non-comparative) form of GAs such as *tall* related to this aspect of their meaning. Assuming that degrees occupy scales, I discuss some fundamental aspects of scales, and then discuss at length how differences in scalar structure (open/closed status of scales) can capture differences in entailment patterns, felicity in comparatives, and – perhaps most importantly – the standard of comparison. Three distinct GA interpretations corresponding to these differences in scalar structure emerge. *Relative* GAs (e.g., *big*/*small*, *long*/*short*) map objects onto open-ended scales and have contextually-determined standards. *Absolute* GAs map onto scales closed at one or both ends. *Minimum standard absolute* GAs (e.g., *wet*, *dirty*, *spotted*) signal the (minimal) presence of a relevant property (i.e., that it is manifested at least minimally). *Maximum standard absolute* GAs (e.g., *dry*, *clean*, *full*) signal the (maximal) absence of a property. These differences motivate the experiments presented in the following chapter, which offer an answer to the question asked at the end of Chapter 2.

In Chapter 4, I present the findings from series of experiments designed to investigate whether children age three to five partition the class of GAs into distinct subclasses based on how the standard of comparison is set. In a pre-experiment Scalar Judgment task, in which participants are asked for judgments about properties of objects along a seven-point scale, differences among the three subclasses of GAs emerge with respect to how the standard of comparison is set. Consistent with their dependence on context, for relative GAs, the standard is is oriented somewhere towards an average or midpoint of the series, while for absolute GAs the standard is oriented towards the maximal or minimal value. Next, in a series of experiments employing the Presupposition Assessment Task, the differences between relative and absolute GAs are highlighted by tracking whether or not the standard shifts when the context does. Analysis of children's reaction times sheds light on an anomalous pattern of responses with the *full* items, providing evidence for a sophisticated level of pragmatic reasoning on the part of the child participants, and a distinction between the vagueness associated with the meaning of relative GAs and the imprecision licensed by language usage related to the maximal standard.

In Chapter 5, I investigate how the input can provide surface-level cues to within-GA distinctions based on scalar structure. I first review how both resultatives in English and adverbial modification are potentially informative in this respect, but then focus on the latter, given the cross-linguistic presence of adverbial modification. Looking first to a set of transcripts from the CHILDES database to get a picture of the presence of adverbial modifiers in child-directed speech, I then turn to the British National Corpus as a model of the broader exposure language, capturing both child-directed and ambient speech. I target two sets of adverbs – those that select for a closed scale (e.g., *completely*) and those that do not (e.g., *very*) –and classify the adjectives appearing with them as *maximal* or otherwise. Likewise, I also target two sets of GAs – maximum standard absolute GAs and relative GAs – and classify the adverbs appearing with them as *restricted* or *non-restricted*. The combined results of a series of statistical analyses demonstrate that adjectives classified as *maximal* are more likely to be modified by restricted adverbs than are non-maximal adjectives, and also appear with a wide range of adverbs. By contrast, although relative GAs appear to be much more frequent in the input than the maximum standard absolute GAs and are more likely overall to be modified by adverbs, they are modified by a narrower range of adverbs, which provide the learner with evidence that they are gradable. These findings therefore demonstrate that the input is rich with information about scalar structure and can be informative about underlying semantic representations *if* the infant is attending to patterns of adverbial modification.

In Chapter 6, I investigate whether infants at 30 months of age *are* actually attending to adverbial modification when learning new adjectives. In a set of experiments using the preferential looking paradigm, infants demonstrate in their behavioral responses that they recruit information provided by adverbial modifiers when learning adjectives. They are more likely to assign a novel adjective a relative interpretation when it is modified by *very* and an absolute interpretation when it is modified by *completely*. Importantly, they display no default interpretation in a no-adverb condition, and are not drawn to either interpretation when a novel adjective is modified by a novel adverb or when it is modified by a low-frequency non-restricted adverb. This pattern of responses provides support for the claim that infants are aware of the differences in distribution and selectional restrictions of adverbs in the input and that these lexical items may play an important role in the partitioning of GA subclasses based on their semantic representations.

Finally, in Chapter 7, I consider my findings in a broader context, asking how the process I have proposed for learning about the semantic representation of gradable adjectives relates to other processes in language learning. I begin by making a connection to the verb learning literature, and draw an analogy between the informativity of sentence structure to verb meaning and the informativity of adverbial modification to adjective meaning. In each instance, a probabilistic analysis of patterns in the input (in this case, the range and type of adjectives appearing with a class of adverbs, and the range and type of adverbs appearing with any given adjective) allows infants to deduce something about the meaning of the lexical item in question. I outline specifically what the learning process looks like both when infants know and when they do not know the meaning of an adverb, and how what they *do* know can enable them to posit semantic representations for both adverbs and adjectives in the input. Thus, what may at first seem circular is actually a very tight interplay between two kinds of lexical items in the word learning process. Finally, having proposed that children rely upon pre-existing conceptual representations to assign the correct scalar structure, I make a connection between the representations of scales in properties and paths in events. I finish by reviewing evidence that infants attend to the structure of paths at an early stage, and suggest that this evidence invites us to be optimistic that infants are also attending to the structure of scales.

Chapter 2: Previous Research on Gradable Adjectives in Language Development

"I woke in bits, like all children, piecemeal over the years.
I discovered myself and the world, and forgot them, and discovered them again.
I woke at intervals until…the intervals of waking tipped the scales,
and I was more often awake than not."

Annie Dillard
An American Childhood

2.1. The Acquisition of Adjectives

If you pick up an article on word learning, chances are that it will begins with a discussion of how nouns are disproportionately represented in early productive vocabularies, leaving other grammatical categories in the dust (cf. Dromi, 1987; Gentner, 1978, 1982; Gentner & Boroditsky, 2001; Goldin-Meadow, Seligman, & Gelman, 1976; Woodward & Markman, 1998). There are a number of possible accounts for the prominence of nouns (some mutually exclusive, some compatible), and these accounts focus on factors such as change in the information that can be recruited in word learning (Gillette, Gleitman, Gleitman, & Lederer, 1999; Piccin & Waxman, in press; Snedeker, Brent, & Gleitman, 2001), challenges presented by form-meaning mapping (Gleitman *et al*., 2005), conceptual change (Huttenlocher, Smiley, & Charney, 1983; Smiley & Huttenlocher, 1995), a whole object bias (Markman & Wachtel, 1988), an object-kind bias (Snedeker, Brent, & Gleitman, 2001), a shape bias (Samuelson & Smith, 1999), accessibility of concrete objects (as opposed to actions or events) in real-world experience (Gillette, Gleitman, Gleitman, & Lederer, 1999; Piccin & Waxman, in press), and so on.[3] Accounts explaining why nouns overshadow adjectives in particular in early lexicons typically

[3] See Choi and Gopnik (1995), Gentner (1982), and Tardif, Shatz, and Naigles (1997) for discussions of how the degree of the noun advantage can vary depending on aspects of the language being acquired, such as relative frequency of nouns and verbs, word order, morphological transparency, and so on.

mention one or more of the following: infants have an initial expectation that words will pick out individuals (whole objects) (cf. Markman & Wachtel, 1988); infants need to know something about how words label objects (or object kinds) before they can master how words label the properties of objects in the world (cf. Gentner, 1978; Maratsos, 1988; Macnamara, 1982); and infants expect words to highlight commonalities among objects that lead to categorization, and specifically expect words cued as count nouns on the surface to link to object categories (cf. Waxman, 1999; Waxman & Booth, 2001; Waxman & Markow, 1995, 1998).

Given that one or more of these possibilities accounts for nouns preponderance in infants' early productive vocabularies, it's not surprising that it's not until nearly 2 years of age that infants begin consistently mapping adjectives to dimensions such as COLOR or TEXTURE, instead of object kinds such as DOG, ANIMAL, or FRUIT (Booth & Waxman, 2003; Waxman & Booth, 2001; Waxman & Markow, 1995, 1998). A number of experiments have already been run and others are being planned to address (1) how children at this age differentiate adjectives (a) from nouns, which appear in different syntactic environments (e.g., *a zav-ish one* vs. *a zav*) or (b) from words that appear in apparently similar structural positions, such as proper names (e.g., *This is brown/James Brown*.); (2) how the extension of adjective meaning is influenced (a) by the role of object familiarity, (b) by the form and content of the modified noun, or (c) by object kind taxonomic level; (3) how comparison of object properties can help in word learning; (4) what the influence of the syntactic position of the adjective is, whether in prenominal or predicative position; and (5) how children process noun phrases with adjectives in real-time and incorporate information from the referential context; and many other topics.[4] The collected

[4] (1a) Booth & Waxman, 2003; Gelman & Markman, 1985; Hall & Moore, 1997; Hall, Waxman, & Hurwitz, 1993; Taylor & Gelman, 1988; (1b) Hall, 1991, 1994; Hall & Graham, 1999; (2a) Gelman & Taylor, 1984; Hall, 1991,

findings from this wide range of studies have enriched our understanding of what children know about this grammatical category – specifically, how and when the adjectival category is treated as distinct from or related to the nominal category.

In addition to distinguishing adjectives from other parts of speech and learning how their meaning relates to these other categories, children also need to know that not all adjectives are alike in their meaning and that these differences in meaning require adjectives to be assigned to different subclasses. Here, I will focus on one class of adjectives in particular, *gradable adjectives*. In the next chapter, I will present a much more in-depth semantic analysis of these adjectives, and describe how the broad class of gradable adjectives can be broken up into distinct subclasses based on their semantic representations. Here, I'll talk about how to identify these adjectives and then discuss what we have learned from previous research about young children's comprehension of these lexical items.

2.2. Terminology

2.2.1. An Introduction to Gradable Adjectives

Gradable adjectives (GAs) are adjectives that can appear felicitously in comparative constructions and can appear with adverbial degree modifiers. Consider the examples in (1). The examples in (a) involve a comparative construction: two (kinds of) entities are compared

---

1994; Hall, Waxman, & Hurwitz, 1993; Taylor & Gelman, 1988; (2b) Mintz, 2005; Mintz & Gleitman, 2002; (2c) Klibanoff, 2000; Klibanoff & Waxman, 1998, 2000; Sharpe, Fonte, & Christe, 1998; Waxman & Markow, 1995, 1998; Graham, Cameron, & Welder, 2005; Graham, Welder, & McCrimmon, 2003; (3) Klibanoff, 2000; Namy & Gentner, 2002; Waxman & Klibanoff, 2000; Sandhofer, 2002; (4) Diesendruck, Hall, & Graham, 2006; Nadig, Sedivy, Joshi, & Bortfeld, 2003; Prasada & Cummins, 2001; (5) Nadig, Sedivy, Joshi, & Bortfeld, 2003; Thorpe, Baumgartner, & Fernald, 2006; Thorpe & Fernald, 2006.

using an adjective with a degree morpheme (here, *-er*). In the (b) examples, the adjective is modified by *very*, a degree adverbial. The examples in (1a-b) are felicitous, while those in (2a-b) are not. This is because *big* is a gradable adjective, but *dead* is not.

1. a. Elephants are bigger than penguins.

   b. That elephant is very big.

2. a. #Attila the Hun is deader than Marie Antoinette.

   b. #The king is very dead.

The fact that GAs can appear in comparative constructions is a consequence of their semantic representations (which will be spelled out in detail in the following chapter). At a very basic level, what allows for gradable adjective to appear in such constructions – and what makes the appearance of non-gradable adjectives in such constructions anomalous – is a difference in meaning. In the above examples, you can compare degrees of *bigness* (or size), but not degrees of *deadness*. Degree constructions involve a comparison between degrees, amounts, or extents of a relevant property, and while the domains of GAs can be ordered with respect to the relevant property, the domains of non-gradable adjectives cannot (cf. Heim, 2000; Kamp, 1975; Kennedy, 1999; Klein, 1980; McConnell-Ginet, 1973; *inter alia*). The class of GAs is extremely large, represented (at least) by adjectives from all the categories of Dixon (1977)'s typology: dimension, physical property, color, human propensity, age, value, speed, and position. Other non-gradable adjectives similar to *dead* include those such as *American*, *pregnant*, *carnivorous*, and *wooden*.

2.2.2. Disclaimers

Before proceeding, I wish to make three disclaimers. First, while GAs have the distributional characteristic that they can appear in comparative constructions, I am not going to review here all that has been said about children's knowledge of comparative constructions (cf. Donaldson & Wales, 1970; Feider, 1973; Finch-Williams, 1981; Gathercole, 1979; Gitterman & Johnston, 1983; Goede, 1989; Graziano-King, 1999; Graziano-King & Cairns, 2005; Layton & Stick, 1978; Moore, 1999; Ryalls, 2000; Ryalls, Winslow, & Smith, 1998; Smith, Rattermann, & Sera, 1988; *inter alia*). In this dissertation, I will only be focusing on the positive (i.e., non-comparative) form of GAs (e.g., *big*, not *bigger*), because I am interested in children's acquisition and knowledge of these lexical items rather than their knowledge of the syntactic constructions in which these items appear or what consequences such knowledge of these items has for explicit comparisons. As I review the relevant literature, the reader will see that there is also plenty to be said on this topic.

Second, I do not say anything about possible asymmetries in the acquisition of the members of antonymous pairs of GAs (e.g., *big*/*small*) or about antonyms in general. Throughout this dissertation, when I refer to GAs, I am referring only to the positive (non-comparative) form of GAs taken from the positive pole (i.e., *big*, not *bigger*, and *big*, not *small*). Readers interested in children's acquisition of antonyms and polar asymmetry should consult the following articles: Barner & Snedeker, 2007; Brewer & Stone, 1975; Clark, 1972; Coots, 1976; Eilers, Kimbrough Oller, & Ellington, 1974; Huttenlocher & Higgins, 1971; Klatzy, Clark, & Macken, 1973; Markowitz, 1975; Tanz, 1977; Townsend, 1976; and Williams, 1977; *inter alia*.

Finally, readers familiar with previously-investigated divisions among adjectives in the experimental literature want to reconcile these divisions with the gradable/non-gradable distinction. To clarify the adjectival classification I am interested in investigating in child language, I will address three other types of adjectival classifications.

2.2.2.1. Dimensional Adjectives

First, as I mentioned above, GAs are – at a very basic level – adjectives that can appear in comparatives. Canonical examples include so-called *dimensional* adjectives or *polar* terms such as *big*/*small* or *tall*/*short*. However, these adjectives that map to a physical dimension are only a proper subset of the larger set of GAs, which span well beyond physical measurement. That is, all dimensional adjectives are GAs, but not all GAs are dimensional adjectives, since many (e.g., *happy*, *full*, *wet*) do not map onto physical dimensions.

2.2.2.2. Relative and Absolute Adjectives

Second, a 'relative/absolute' distinction among adjectives has been widely discussed over the years and has guided a number of experimental designs. Relative (or *scalar*) adjectives are said to relate to a standard or norm, and have been said to contrast with so-called absolute (or *predicative* or *intersective*) adjectives, which are said to denote properties where the adjective-noun phrase picks out an entity that has the property described by the adjective and belongs to the category named by the noun – that is, it represents an intersection of the two properties (e.g., a *wooden table* is both *wooden* and a *table*). At first blush this seems like the gradable/non-gradable distinction; however, once we examine the line that has been drawn for relative and absolute adjectives, we will see that we need to re-draw it slightly for gradable and non-gradable adjectives.

Two cases in particular illustrate this point. Rips and Turnbull (1980) examined adults' error rates and reaction times for evaluating the truth value of sentences with the classes of adjectives in (3).[5]

---

[5] Rips and Turnbull (1980) note that their dichotomy ignores details that would be captured in a semantic analysis and suggest that these adjectives may instead lie on a continuum. While such a possibility exists for the processing

3. Adjective classes identified by Rips and Turnbull (1980) (including their descriptions and examples)

   a. Relative

      Adjectives based on an underlying ratio scale of physical measurement (e.g., *large*, *small*, *narrow*, *wide*) and those that depend on ordinal judgments (e.g., *safe*, *dangerous*, *strong*, *weak*, *happy*, *sad*)[6]

   b. Absolute

      Adjectives that denote qualitative properties such as color (typical) (e.g., *pink*), shape (e.g., *square*), physical composition (e.g., *wooden*), nationality (e.g., *Chinese*), and others (e.g., *six-legged*)

Likewise, Nelson and Benedict (1974) examined children's error rates and reaction times for judgments of stimuli described by positive and comparative forms of adjectives in the three classes of adjectives as presented in (4).

4. Adjective classes identified by Nelson and Benedict (1974) (including their descriptions and examples)

---

of these adjectives, which calls upon the individual to incorporate information from the referential context in order to make a judgment about whether a property holds of an entity, the distinction we are interested in pursuing here is a grammatical one encoded in the representations of these adjectives.

[6] This description of scales of ratio and ordinal judgments comes from Huttenlocher and Higgins (1971), who also distinguish between adjectives that "fall into single continuous dimensions allowing for the ordering of various items within them" – whether the dimension is continuous or discrete – and dichotomous classes (e.g., *male-female*, *dead-alive*) "where the adjectives cannot fit on a single dimension" (e.g., shape, color) (pp. 487-488).

a. Relative

Adjectives that involve comparison between two antonymous poles on a scale that is relative to a perceptual array or stored class standard (e.g., *big*, *little*)

b. Absolute

'Categorical' adjectives in which the entity either does or does not have the property, allowing for the possibility of a continuum with a threshold above which the property may hold; not antonymous (e.g., *striped*, *furry*, *round*, *red*)

c. Contrastive

Adjectives that are antonymous, but dichotomous, since the two poles are mutually exclusive, and there is a neutral middle ground between the two poles (e.g., *happy*, *sad*, *clean*, *dirty*)

Looking at these two examples (as representative of a larger body of research), it is clear that the authors were interested in classifications that (broadly speaking) distinguished between words that rely on scalar judgments and those that were more categorical; however, the felicity of an adjective's appearance in a comparative construction was never used as a diagnostic for classification. Thus, the 'gradable/non-gradable' distinction does not align neatly with the 'relative/absolute' distinction.

For example, color terms such as *red*, which are often labeled as absolute adjectives, are clearly gradable, as highlighted in (5a-b).[7]

---

[7] See discussion in Bolinger (1967b), McConnell-Ginet (1973), Rusiecki (1985), Sapir (1944), and especially Wheeler (1972) on this point.

5. a. The *redder* the light, the longer the wavelength, and the longer the wavelength, the more deeply it can penetrate body tissues, Dr. Whalen says.

(“Healing Light,” article on LED therapy on www.nasaexplores.com, April 19, 2001)

b. Aunt Marge had already had quite a lot of wine. Her huge face was *very red*.

(Rowling, J. K. (1999). *Harry Potter and the Prisoner of Azkaban*. New York, NY: Scholastic Press, pg. 27)

Similarly, adjectives such as *striped* and *happy* can also appear in comparatives, as in (6).

6. a. He is especially remarkable in being *more profusely striped than* his sire (the zebra Matopo)…In being profusely striped, Romulus differs greatly from the quagga hybrid bred by Lord Morton, in which the stripes were fewer in number than in many dun-coloured horses. (II. Experiments with West Highland Ponies)

(Ewarts, J. C. (1899). Experimental contributions to the theory of heredity. A. Telegony. *Proceedings of the Royal Society of London*, *65*, pg. 249)

b. The newspapers are responsible for our giving Hon. Jesse Felix West’s middle name as Jesse *L*. in our February number. We are glad to make the correction and hope the Judge may always be *as happy as* his middle name indicates.

(*The Virginia Law Register* 8 (1), Errata: Editorials, pg. 59, 1922)

Some shape terms (e.g., *round*) are also gradable. Thus, in this dissertation, when I refer to *gradable adjectives*, I mean adjectives that can appear in comparatives.

2.2.2.3. Properties and Kinds

Finally, the ‘gradable/non-gradable’ distinction cuts across the ‘property/kind’ distinction discussed by Prasada (1992). Prasada was interested in the difference between adjectives such as *big*, *red*, and *clean*, which he said name properties of things, and adjectives such as *former*, *alleged*, and *corporate*, which he said name kinds of things (e.g., senator, crook, or lawyer). One diagnostic he used for classifying an adjective was whether the adjective could appear in both prenominal and predicative position. The former can appear in both positions, but the latter are restricted to prenominal position. Thus, we observe the difference between *big* and *former* in examples (7) and (8). (See also discussion in Bolinger (1967).)

7. a. That is a *big* elephant.

   b. That elephant is *big*.

8. a. He is a *former* president.

   b. *That president is *former*.

Note, however, that *dead*, a non-gradable adjective, can also appear in both positions, as seen in (9).

9. a. That is a *dead* plant.

   b. That plant is *dead*.

Thus, both gradable and non-gradable adjectives can appear in both structural positions, and would therefore both be considered 'property' adjectives in Prasada (1992)'s classification. The 'property/kind' classification is also rather confusing, since 'kind' adjectives are also called 'property-modifying', or 'non-intersective', adjectives elsewhere, and are said to denote functions from properties to properties.

2.3. Previous Work on the Acquisition of Gradable Adjectives

Having clarified what constitutes a (positive form, positive-pole) gradable adjective, we can begin to review what has been said about children's semantic knowledge of these lexical items. I will begin broadly and gradually narrow my focus in order to ask questions that get at the heart of the semantic representations of GAs.

Gradable adjectives – more specifically, terms such as *big* and *little* – are among the first adjectives to be produced by children (cf. Blackwell, 1998; Nelson, 1976).[8] Blackwell (1998) conducted an extensive analysis of the adjectives appearing in the three transcripts represented in the Brown corpora from the CHILDES database (Brown, 1973; MacWhinney, 2000). She found 27 adjectives appearing in the speech of Adam, Eve, and Sarah in an MLU (mean length of utterance) range of 1.5 to 2.5 and an age range of 18 to 35 months. The majority of these adjectives are gradable (only the last few in the list are non-gradable): *big*, *little*, *cold*, *hot*, *warm*, *dirty*, *pretty*, *busy*, *happy*, *sleepy*, *poor*, *nice*, *good*, *bad*, *blue*, *orange*, *purple*, *red*, *brown*, *old*, *new*, *round*, *stuck*, *broken*, *ready*, *another*, *other*. Examples of some of these utterances are included in (10).

10. a. *Big* tower (Eve, 19 months, file #4)

 b. Eve tapioca *hot*. (Eve, 20 months, file #20)

 c. Zip it *open*. (Adam, 30 months, file #9)

 d. Hey you get all *dirty*. (Sarah, 42 months, file #67)

As language learners continue to build up their productive repertoire, they exhibit a number of ostensibly non-adult-like behaviors. First, children have been known to overgeneralize, often substituting the positive polar term for the negative polar term (e.g., *big* for *little*, *wide* for *narrow*) and more general terms for more specific ones (e.g., *big* for *long*) (cf. Clark, 1972, 1973; Sera & Smith, 1987). Second, when shown a series of objects, two- and three-year-olds seem to prefer to label only the extremes of the series (cf. Clark, 1970; Ehri, 1976; Smith, Cooney, & McCord, 1986).

[8] This early production is modeled in part on the input to children, but not entirely. Conceptual accessibility and generality of the semantic features of adjectives have also been identified as contributing factors.

Third, a much-discussed difference between size and color terms concerns the fact that although children can perceive differences in color, can correctly map color terms to their real-world correlates early on, and in infant language studies demonstrate an ability to use color as a commonality leading to object categorization, preschoolers' production of color terms lags behind many size terms and they do not always respond correctly when asked to identify a color, often providing an incorrect color term. (see Backsheider & Shatz, 1993; Bornstein, 1985; Clark, 2006; Kowalski & Zimiles, 2006; O'Hanlon & Roberson, 2006; Sandhofer, 2002; Sandhofer & Smith, 1999, 2001; Sandhofer & Thom, 2006; Schwartz, 1977; Soja, 1994). (Because of the unique path of acquisition exhibited by color terms with respect to other adjectives, I will not say anything more about them in the rest of this chapter.)

Fourth, around four years of age, children's knowledge of size terms in particular takes an interesting turn. Although at three years, children can correctly pick out a big object from a perceptual array, one year later, they appear to rely too heavily on the vertical dimension (i.e., height) in their judgments of the overall size of objects. This pattern, which has been termed a 'decrement' in their understanding of these terms, does not persist for too long, and by five or six years of age, children are able to take into account the correct dimensions when judging the size of objects using GAs (cf. Coley & Gelman, 1989; Gathercole, 1982; Harris & Folch, 1095; Harris, Morris, & Meerum Terwogt, 1986; Lumsden & Poteat, 1968; Maratsos, 1973, 1974; Ravn & Gelman, 1984; Sena & Smith, 1990).

2.3.1. Standards of Comparison

We have seen that the hallmark of GAs is that they can appear in comparative constructions, because GAs involve reference to a standard of comparison. Let's now take another look at example (1a), repeated here as (11).

11. Elephants are bigger than penguins.

In this sentence, the size of penguins serves as the *standard value* to which the size of elephants (the *reference value*) is compared. The sentence does not entail that both animals are big, or even that the animal exceeding the standard is big (although in this case, it is).[9] Such a comparison does not even require that either of the entities comes from the same category, as illustrated in (12). The degree morpheme *-er* simply requires that the entities stand in the "greater-than" relation in the order indicated with respect to the meaning of the gradable adjective that heads the degree construction (cf. Kennedy, 1999).[10]

12. Penguins are bigger than tomatoes.

Note, however, that even when stripped of their comparative clothing, bare positive GAs still involve a comparison to a standard, although in this case the standard is implicit (cf. Bartsch & Vennemann, 1972; Bierwisch, 1967; Bresnan, 1973; Cresswell, 1976; Kamp, 1975; Katz, 1967; Klein, 1980; Ludlow, 1989; Montague, 1974; Parsons, 1972; Sapir, 1944; Seuren, 1978; Siegel, 1979; von Stechow, 1984; Wallace, 1972; *inter alia*). For example, (13) means that elephants are big relative to some relevant standard of size (e.g., other animals). This implicit standard has consequences for determining the truth value of the sentence in which the gradable adjective appears (i.e., the sentence can be true or false depending on the size standard to which we are comparing elephants).

13. Elephants are big.

---

[9] The picture is slightly different when a negative gradable adjective (e.g., *short*) is the main predicate, but I will not be discussing such cases here.

[10] A further stipulation determining the felicity of comparisons is that the entities must be commensurable – that is, they must exemplify a property that can be measured along the same parameter or dimension.

An implicit standard (or *comparison class*) can be set by the semantic content provided by the noun phrase (e.g., elephants, animals) or by the context or shared knowledge between the discourse participants (e.g., (13) may be offered in a conversation about the sizes of various things in the world) (cf. Kamp, 1975; Platts, 1979; Siegel, 1979; Vendler, 1968; Wheeler, 1972). This comparison class serves to sharpen the meaning of such predicates: absent a context for interpreting (13), we are unable to assign a truth value to the sentence. For example, compared to other animals, elephants are big, so the sentence is true, but compared to the Sears Tower, Dhaulagiri Mountain, or the planet Mercury, elephants are not big.[11] In the following chapter I draw on a long history of research in semantics and philosophy to analyze this phenomenon and discuss another factor contributing to the vagueness of these predicates (e.g., borderline cases). For now, working with a very basic description of sentences such as (13) allows me to pose two fundamental questions related to children's knowledge of the semantics of GAs.

1) Can children use contextual information to set the standard of comparison?

2) Do children allow the standard of comparison to shift?

In the paragraphs that follow, I will present experimental evidence from a range of studies that have addressed both of these questions for children as young as three years of age, and in some cases as young as two.

---

[11] This view of the role of the comparison class is in contrast to a view which sees the comparison class as a function that takes the context as its argument and partitions the domain of discourse in order to assign a truth value to the sentence. See Kennedy (1999) for further articulation of differences between a scalar analysis (the one adopted here) and a vague predicate analysis.

2.3.2. Can children use the context to set the standard of comparison?

The combined results of a number of experiments demonstrate that young children are able to appeal to a variety of aspects of the context when setting the standard of comparison. To begin, recall that some researchers have claimed that around age four, children go through a period in which they rely too heavily on the vertical dimension in judgments of object size (e.g., when asked to locate *the big one*, they may select the tallest one). Whatever the reason for this behavior – e.g., spreading of lexical features from newly acquired dimensional terms such as *tall* (cf. Clark, 1972, 1973), perceptual salience of the vertical feature of the objects (cf. Maratsos, 1973, 1974), etc. – the fact remains that children are basing their judgments on how an object relates to others in the context given a salient dimension. That is precisely what we expect for them to do if they understand that the positive form of a gradable adjective calls upon them to make an implicit comparison.[12] However, more direct evidence of children's ability to use the context in assessing object properties corresponding to GAs comes from experiments in which children are asked to make judgments about members of a set of objects that can be ordered linearly, or an object that is assigned positions in a linear stepwise progression.

---

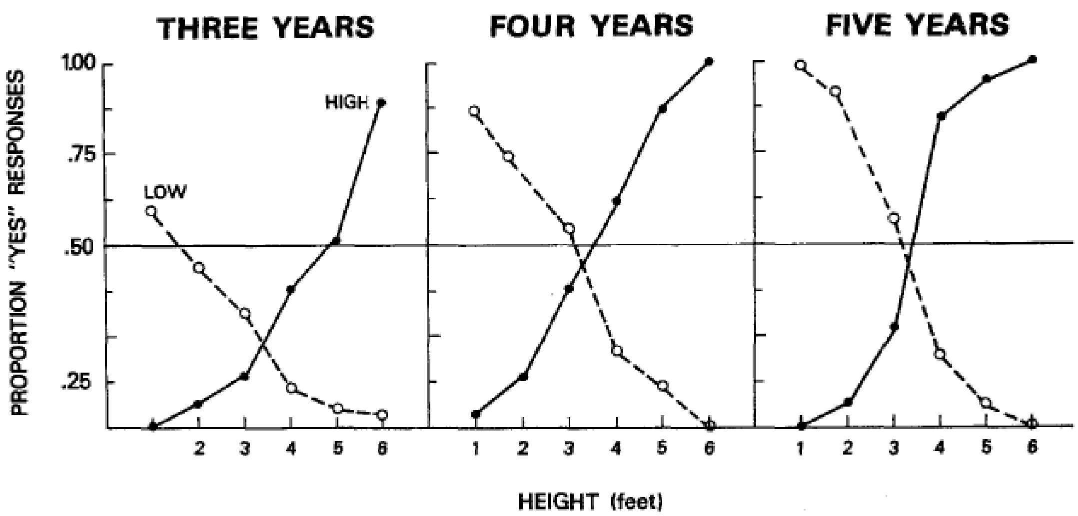
[12] Further evidence of children taking contextual information into account comes from experiments in Dutch by Krämer (2005), who found that children age four to eight were overwhelmingly biased to take into account the size of the container rather than the size of the set of objects when judging the truth value of sentences with *veel* ('many'), regardless of its syntactic position in the sentence (i.e., subject/topic vs. postverbal subject in an existential *there* construction). For example, in contexts in which five (of five, or of seven) objects were placed in a rather large container (e.g., eggs in a basket), child participants did not think that 'a lot' were in there, referencing the sizable empty space in the container, rather than how many objects were left over.

It is well known that, when given a series of objects, children as young as age three tend to label only the extremes (i.e., the objects at the positive and negative poles) and resist labeling the objects in the middle of the series. This behavior was initially interpreted as their *nominal* treatment of GAs (cf. Clark, 1970) – basically, that children saw these terms as picking out more stable, contextually invariant categories. However, as Sera and Smith (1987) have pointed out, such behavior actually shows that children have a relative understanding of these terms. In order to make a determination about what counts as *big* or *little* and label the poles categorically across experiments with sets composed of various numbers of objects, children must be able to take the entire perceptual array (i.e., the context) into account and set a standard based on this information.

Another way of thinking of younger children's refusal to label objects in the middle of the series is to say that children at this age have difficulty assigning truth values in the space between the clear-cut cases (Klein, 1980; cf. also Fine, 1975), or that this is simply a result of GAs qua vague predicates giving rise to borderline cases that are difficult to classify. Under Klein (1980)'s account, a set of objects able to be described by an adjective such as *big* or *tall* will be treated as the relevant context. This context is the comparison class and is partitioned into three disjoint subsets – definitive cases of *big*, *not big*, and a middle group about which a decision is pending. Klein terms this grey area the *extension gap* and proposes resolving it by successive application of the predicate, in this case *big*, until it is fully resolved. In this way, every extension gap remaining from one application will be the context fed into the function the next time around, and so on. If this account is right, then children may have difficulty successively reapplying the predicate to partition the entire context into two groups without an intermediary extension gap. Even if Klein's treatment of such adjectives is not correct – in fact, in the next chapter, I adopt a semantics of GAs which is distinctly different from his approach – it may be that younger children are at least 'more comfortable' than older children and adults leaving a (sometimes considerably large) range of objects in the middle of the series without definite labels.

Smith, Cooney, and McCord (1986) sought to determine whether children partitioned a range categorically for bipolar *high*/*low* judgments. They positioned a red disk, 11" in diameter in six 1' intervals against a 6'-high backdrop, and asked three-, four-, and five-year-olds for judgments about its height (*high*/*low*). The authors used two forms of measurement to assess child responses: the crossover point (at which judgments exceeded 50% *yes*) and the slope of the curves at the crossover. Having observed three-year-olds have a higher crossover point for *high* and a lower crossover point for *low*, and that slopes are sharper for the four-and five-year-olds (cf. Figure 1, their Figure 1 and Table 1), they concluded that three-year-olds are more narrow and extreme in their judgments, while the older children have broader, more 'categorical' categories.

Figure 1: Results from Experiment 1 of Smith, Cooney, & McCord (1986) (figure and table reproduced from source)



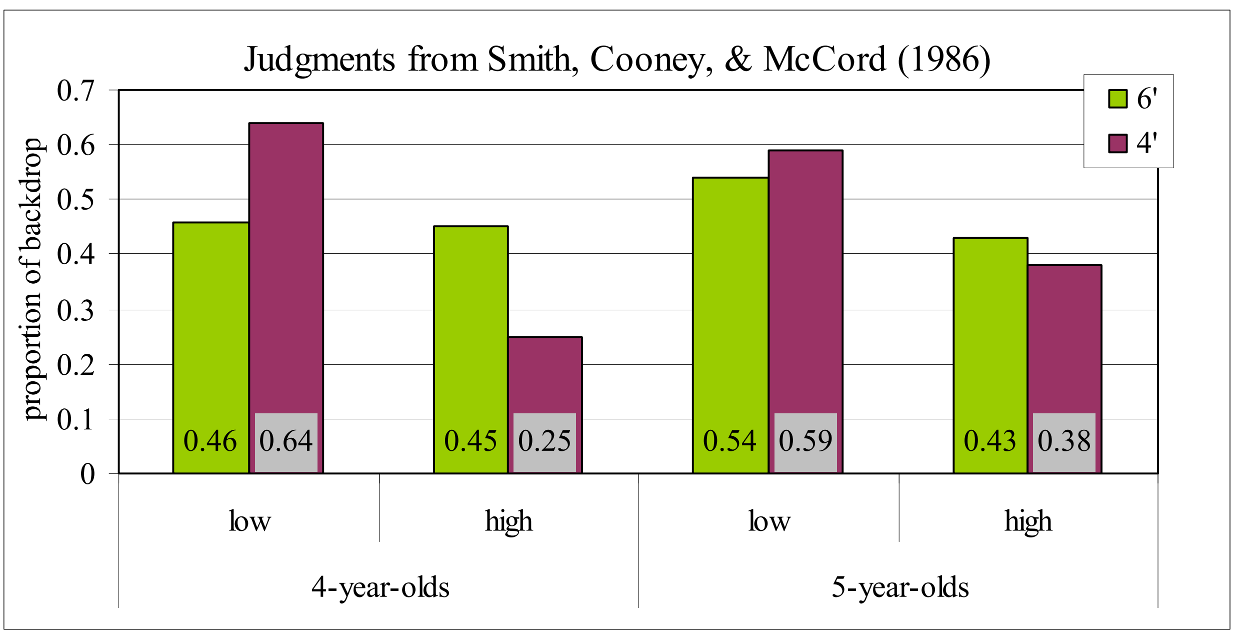
MEAN CROSSOVER POINTS AND MEAN SLOPES AT CROSSOVER FOR "HIGH" AND "LOW" AT THE THREE AGE LEVELS

	M CROSSOVER POINT		*M* SLOPE	
AGE (Years)	High	Low	High	Low
3 (*N* = 7)	5.07 (.92)	2.18 (.99)	.341 (.121)	.270 (.148)
4 (*N* = 9)	3.20 (.92)	2.91 (.89)	.403 (.096)	.420 (.137)
5 (*N* = 10)	3.31 (.47)	3.32 (.64)	.507 (.117)	.482 (.131)

NOTE.—Standard deviations are in parentheses.

In Experiment 4 of their study, when a 4' backdrop was used in place of the 6' backdrop, the authors found differences among all three age groups. For five-year-olds, the crossovers for both terms were higher when judging the height of the disk against a 6' than a 4' backdrop, reflecting the fact that they made use of the entire range in making judgments. However, the crossovers for three-year-olds were higher in the 6' than in the 4' condition for *high* but did not differ for *low*, and the crossovers for the four-year-olds did not vary with height of the backdrop (crossovers of 3.28 ' *high* and 2.79' *low* for the 6' backdrop, and 3.01' *high* and 2.28' *low* for the 4' backdrop). Another way of describing this difference, which makes the difference between the four- and five-year-olds more striking, is by looking at the proportion of the backdrop judged *high* or *low* (derived from the crossover points). For five-year-olds, the proportions judged *high* or *low* did not change between the 6' and 4' backdrop, while for four-year olds, they did. Four-year-olds appeared to apply absolute judgments, instead of recalibrating as the five-year-olds did, so that as the crossover points remained constant, but the backdrop decreased in size, the proportion of the backdrop judged *high* decreased from the 6' to the 4' condition at the same time as the proportion of *low* judgments was higher in the 4' condition. This difference in proportions is illustrated in Figure 2.

Figure 2: Judgments for *high* and *low* in Experiment 4 of Smith, Cooney, & McCord (1986)



The authors interpreted these results as providing evidence that, "For 5-year-olds, the physical height at which objects are *high* and/or *low* varies with changes in extent of variation but the *proportion* of extent of variation at which objects are *high* and/or *low* is constant. In contrast, for 4-year-olds it is the *physical* height at which an object is high that is invariant across the two contexts, not the proportion of variation" (pg. 597). In fact, the results of their Experiment 5 demonstrate that the five-year-olds perform similar to adults in this task. The ability to impose context-dependent reference points and recalibrate the standard when the range of comparison changes in size thus appears to develop with age.

Barner and Snedeker (2007) have, however, demonstrated that four-year-olds can take the entire range of objects into account when delivering judgments of *tallness*. In their Experiment 1, children were presented with 9 novel objects (labeled 'pimwits') ranging in one-inch intervals from 1" to 9" in height in one of two pseudorandomized orders. Children were asked to place all of the tall pimwits into a plastic circle.[13] In this baseline condition, the average minimum selection for *tall* judgments from 16 children was 7.19". In Experiment 2, the authors included four additional pimwits (0.5", 1", 1.5", and 2"), reducing the mean height of the objects from 5" to 3.85". The authors predicted that if children perform a statistical analysis over the entire array of objects, the average minimum height for *tallness* would decrease accordingly. This is just what happened: the average minimum height for judgments for *tall* pimwits shifted downward to 5.44".[14]

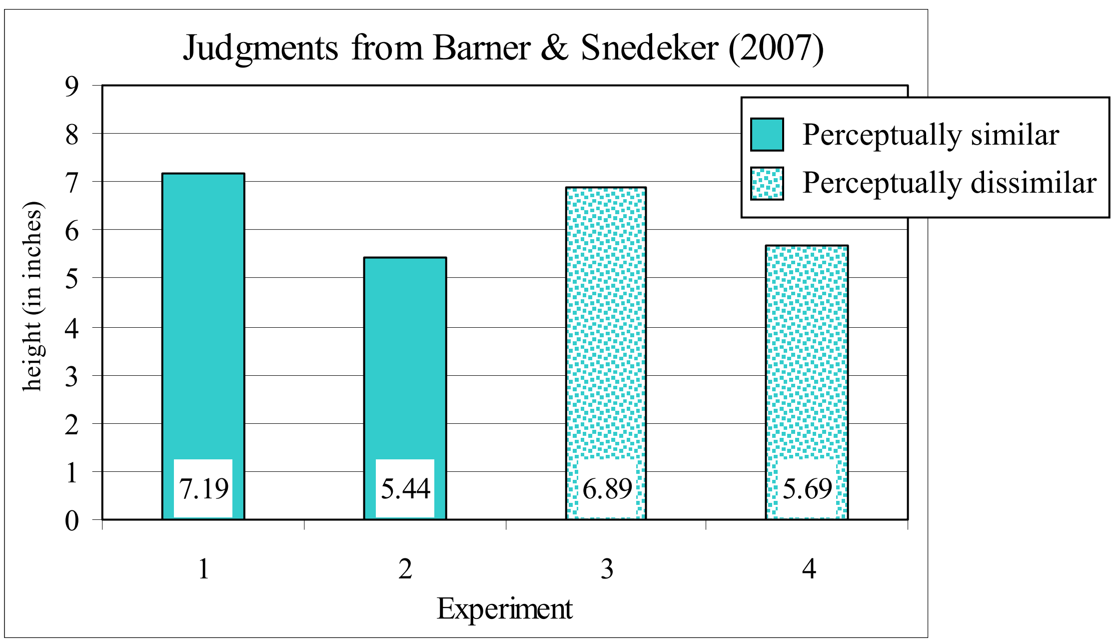
---

[13] Children were also asked to give judgments about *short*. However, since Barner and Snedeker observed an asymmetry between judgments for *tall* and *short*, and since the judgments for *tall* alone capture children's abilities to set and shift the standard of comparison appropriately given the composition of the array of objects, *short* judgments are not reported here.

[14] In this paper, the authors did not report on a follow-up experiment where a similar set of four objects *increased* the mean height (e.g., additional pimwits 9.5", 10", 10.5", and 11" in height), leaving it an open question whether this behavior can be generalized to both polar directions. However, in a subsequent version of the paper, they include such a condition, and find that the standard shifts upward, as predicted. In addition, they stated the findings in terms of children's statistical computation of the standard based on the mean height of the object. But since they

In their Experiment 3, when four new same-type objects were given a new label ('tulvers') and mixed in with the original nine objects, the same shift was not observed. The average minimum height for *tallness* was not significantly different from Experiment 1, 6.89". However, in Experiment 4, when these four objects were also labeled 'pimwits' and the experimenters emphasized that despite perceptual differences, all objects were of the same kind, the average minimum height shifted to 5.69", and was significantly different from Experiment 1 (9 objects) but not Experiment 2 (13 perceptually similar items). The judgments from these four experiments are captured in Figure 2.

Figure 3: Average minimum height of *tall* judgments from Barner & Snedeker (2007)



---

did not control for this feature with respect to other possible statistical factors (e.g., mode), it also remains an open question what statistical information is fed into the calculation of the standard of comparison for children at this age and for adults.

Together, these experiments demonstrate that in setting the standard of comparison for *tall*, four-year-olds do take into account information from the entire set of same-kind objects in an experimental setting.[15] The difference in findings for four-year-olds' performance between Smith *et al*. (1986) and Barner and Snedeker (2007) appears to lie in the nature of the task and/or the way in which the responses are measured.[16] Four-year-olds are thus not as 'absolute' as they initially appeared to be, but still may not be as agile at shifting as the five-year-olds.

In addition to taking into account the number of items in an the array, children are also able to incorporate other information about the objects in the array when delivering judgments. Evidence of this sort was noted above in the context of the discussion of Experiments 3 and 4 in Barner and Snedeker (2007). Smith, Cooney, and McCord (1986) also included an experiment in which they assessed whether children could use object class/kind information when establishing reference points. In Experiment 2, the 6' backdrop was painted with a scene of a tree and some grass, and the large red disk was replaced by a bird and a bunny, each of the same size as the disk. The authors hypothesized that what is high for a bird (an animal that is typically found in high places) and a bunny (an animal that is typically found on the ground) should differ. The task proceeded similarly to the experiments described above. Although three-year-olds were

---

[15] Since Barner and Snedeker (2007) only tested one age group, it is not possible to conduct a direct comparison between their study and Smith, Cooney, and McCord (1986) with respect to developmental differences in the ability to recalibrate when the size of the range is altered.

[16] In Smith, Cooney, and McCord (1986), one object was placed in various heights along one backdrop, whereas in Barner and Snedeker (2007) judgments ranged over a set of multiple objects. In addition, the objects in Barner and Snedeker (2007) were in a pseudorandomized order, whereas the disk/bunny/bird in Smith, Cooney, and McCord (1986) was either raised or lowered in ascending or descending order.

somewhat confused by this task, overall they appeared to render judgments of *high* and *low* independent of object class/kind information. The two older age groups differed from each other dramatically.

Four-year-olds judged the bird to be *high* more often than the bunny and judged the bunny to be *low* more often than the bird. In contrast, five-year-olds judged the bunny to be *high* more often than the bird, and the bird *low* more often than the bunny. The authors interpreted these results as indicating that for four-year-olds, a characteristically high object is more likely to be judged *high* than a characteristically low object, while for five-year-olds, the reference point for characteristically high objects is higher than the reference point for characteristically low objects. The difference between the age groups, then, appears to lie in how they incorporated object kind information into the equation. Four-year-olds seemed to think of judgments of *high* as *high/low because it is an X*, while five-year-olds were more adult-like and thought of judgments of *high* as *high/low for an X*. Again, we see a difference between four- and five-year-olds that reflects an increasing ability to make use of the context when making judgments about gradable properties/adjectives.

Object class/kind is also important for judgments of the size of individual items. Ebeling and Gelman (1988) and Gelman and Ebeling (1989) have shown that children as young as two years of age are better at judging the size of an object when it is familiar to them than when it is novel. For example, when presented with a mitten, young children can rather easily label it *big* or *little*. However, when given a novel shape for which they have no 'normative' standard, but which is matched in size with the mitten, they are less successful.

Children of this same age are also able to perform judgments of 'functional size' – that is, determine the size of an object in relation to its intended use (e.g., a hat for a head, a dress for a doll). Shown dolls of different sizes and an object (either clothing or a tool) held up to the doll, and asked if the object is *big* or *little*, children as young as age three are able to judge the size of the object relative to the doll's size (Ebeling & Gelman, 1994, Experiment 1; Gelman & Ebeling, 1989; cf. also discussion in Carey, 1978). Their performance improves significantly when their attention is explicitly drawn toward the mismatch in size (e.g., by the experimenter attempting to try the clothing on the doll or asking explicitly, "Is this hat too big or too little for the doll to wear?") (Gelman & Ebeling, 1989, Experiments 2, 3).

The evidence gathered across these experiments indicates children's developing ability from age two to five to incorporate a variety of information into making judgments about the relative property of an object. They are able to evaluate an object's size (e.g., label it *big*/*little*, *tall*/*short*) in relation to other perceptually salient objects, other same-kind objects from a series that can be ordered linearly, its object class/kind, or another object highlighting its intended use. They are also able to evaluate an object's height (e.g., label it *high*/*low*) relative to a given range of heights. The true test of children's competence with respect to the GAs targeted in these experiments comes in the form of experiments that ask children to shift the standard of comparison. The set of experiments by Smith, Cooney, and McCord (1986) described above touch upon children's ability to do just this. What happens, though, when an individual object, whose size has been previously determined relative to one standard is placed in another context that prompts a reevaluation? In the next section, I review a number of experiments which demonstrate that children – in some cases, children as young as age two – are able to abandon previous judgments and shift to a new standard of comparison.

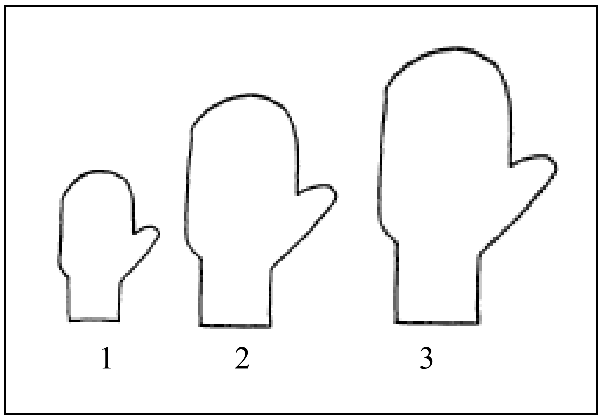
2.3.3. Do children allow the standard of comparison to shift?

Two quotations encapsulate the issue addressed in this section. First, a language learner "does not understand *tall* if he can only say of objects that he has previously encountered whether or not they are tall, if he only knows the extension of *tall* as it intersects with objects in the immediate world of his experience. To understand *tall* he must have some notion of how to decide whether the predicate applies to the newly encountered object" (McConnell-Ginet, 1973, pg. 124). Second, "The speaker must not only know the different uses of a word…and know the contexts that elicit each usage…, but must also be able to 'switch gears' – to reframe the context as necessary. Contexts are continually shifting, so the speaker needs to be flexible enough to respond to these shifts" (Ebeling & Gelman, 1994, pg. 1179). Thus, if a child really understands that the meaning of an adjective such as *tall* relies upon a set of salient discourse entities, she will allow this domain (i.e., the context) to be variable and will therefore also allow the standard for *tall* to vary accordingly.

Recall that for the positive form of a gradable adjective such as *big* or *high*, the standard of comparison is implicit and depends on contextually salient information. The standard for such adjectives is therefore predicted to shift from context to context. Do preschool-age children demonstrate a facility in shifting the standard of comparison in this way? Two lines of research demonstrate reliably that they do. The first line focuses on cases in which they are asked to relabel an object – sometimes immediately – when it is compared to a new same-kind object, the second on cases in which standards of a different nature conflict with each other.

Ebeling and Gelman (1988) asked children two and a half to four and a half years of age to make 'perceptually' based judgments about a series of pairs of objects (pictures of mittens or similar novel shapes). Each set of objects was made up of nine items divided into three size groups – small, medium, and large. Within each set of three objects, the middle item was targeted for comparison; when presented with a smaller size object from the triad, it was *big*, but when presented with the larger size object, it was *small*. See, for example, mitten #2 in Figure 4.

Figure 4: Example of mitten stimuli used in Experiment 1 of Ebeling & Gelman (1988) (figures adapted from source)



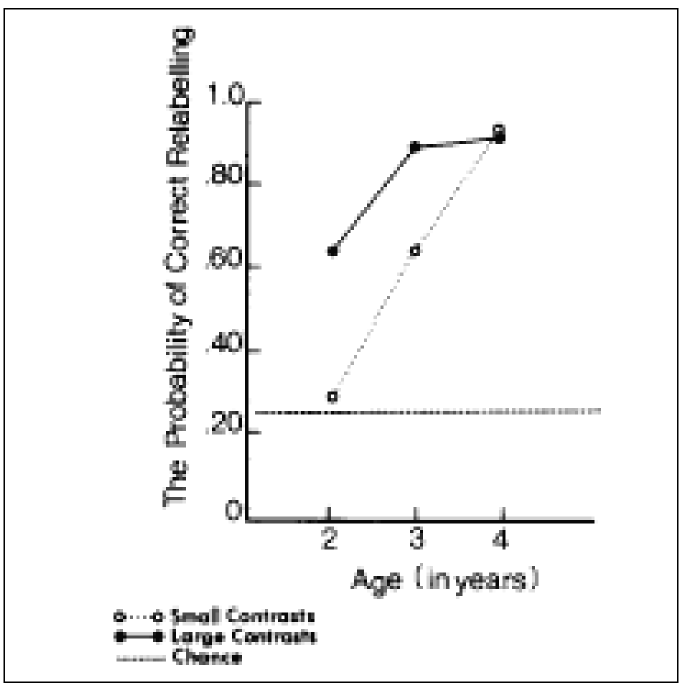
Children performed above chance in all but one condition: two-year-olds exhibited some reluctance to call an otherwise big mitten *little* when compared with an even bigger mitten, but were not reluctant when judgments concerned big novel shapes. Performance for all children improved when, in a variation of this task, children were not asked for *big* or *little* judgments, but were asked to give the experimenter *the big one* or *the little one*.[17]

In the tasks described above, the pairs were presented in a pseudorandomized order, with *big* and *little* questions also pseudorandomized. Sera and Smith (1987) placed children two to four years of age in a somewhat more challenging situation, asking them to immediately relabel the object. In their Experiments 1 and 2, children were shown a series of pairs of circles ranging in size from 1" to 12" in diameter. For each pair in 'Phase 1', one object was singled out and

[17] Ebeling and Gelman (1988) referred to the first task as a production task and the second as a comprehension task.

children were prompted with, “This one is _____.”[18] Immediately afterwards in ‘Phase 2’, one of the objects was removed and another put in its place. Children were then asked again for judgments about the original object. By analyzing the probability of a correct response on the judgment in Phase 2 given a correct response on the initial judgment in Phase 1, the authors showed that even the youngest children were able to perform significantly above chance on this task. Performance did improve with age and, for the two- and three-year-olds, performance also improved when the magnitude of difference between the two objects was larger (1" versus 4"). (See Figure 5.)

Figure 5: Probability of correctly relabeling a circle from Sera & Smith (1987) (figure reproduced from source)



[18] A practice session with other materials familiarized children with the goals of the game and made sure that size was the salient feature to be labeled.

In a series of three papers, Karen Ebeling and Susan Gelman offered a detailed picture of children's ability to render judgments about stimuli when standards conflict. The three types of standards they investigated were those mentioned above: normative (i.e., relative to object class/kind), perceptual (i.e., relative to another object), and functional (i.e., relative to intended use with another object). Their combined results reveal the following. First, when children as young as two are asked to switch from a normative to a perceptual standard, the shift proceeds smoothly (Ebeling & Gelman, 1988, Experiment 2; Ebeling & Gelman, 1994, Experiment 1). For example, children can accurately judge the size of individual articles of clothing as *big* or *little*, and then later evaluate their size with respect to another article of clothing that may or may not result in relabeling the original item. Second, children as young as age three shift from a normative to a functional standard relatively smoothly (cf. Gelman Ebeling, 1989, Experiment 1), although less smoothly than when switching to a perceptual standard (Ebeling & Gelman, 1994, Experiment 1). For example, children can judge an article of clothing with respect to other clothing and then with respect to a doll for which it was intended.

Finally, the direction of the shift (i.e., from one type of standard of comparison to another) can make a significant difference (Ebeling & Gelman, 1994, Experiments 2, 3). While children shift somewhat easily *from* a normative standard *to* either a perceptual or functional standard, when asked to shift *from* a perceptual or functional standard *to* a normative standard, they experience some difficulty. Specifically, they appear to ignore the new normative context and perseverate with their earlier perceptual or functional judgment. Thus, an object that should be judged *big* individually (and ordinarily is judged as such in normative tasks) is labeled *little* if, in an earlier comparison to a bigger object or to a bigger doll, the child found it too small to perform its function. This pattern holds for two- and three-year-olds for normative-perceptual/perceptual-normative shifting, and for four-year-olds for normative-functional/perceptual-functional shifting (the youngest age groups tested for each). Interestingly, Ebeling and Gelman (1994) found that adults often appeared to ignore the perceptual context and judge objects solely based on a normative standard.

In sum, it appears that preschoolers demonstrate a facility to shift a standard of comparison, both in terms of recalibrating the standard with respect to a different cutoff point and in terms of taking into account a different type of standard. However, both the contextual information incorporated into judgments and the ease with which this shift occurs increase developmentally. How much of these developmental differences should be attributed to general cognitive or processing expertise and how much is specific to the semantic representation of these terms remain open questions. For example, when similar judgments are called for in the absence of adjectives labeling these properties, do similar results surface? Moreover, we would want to know how much of the difference among age groups is due to development of cognitive and/or semantic expertise as opposed to task effects, and whether an apparent willingness to take contextual variables into account is actually an inability to refrain from relating terms such as these to the discourse context when it is not appropriate to do so. I will focus my attention on this last point.

2.4. Gradable Adjectives and Standards of Comparison: It's not all relative.

It may appear that we have a complete picture of what children know about GAs, but while the studies I have reviewed give us a good idea of what young children know about setting and shifting contextually-based standards of comparison, these experiments have only focused on a *subset* of a much larger set of GAs, because they focused on terms pertaining to physical dimensions. By restricting their focus to size terms, previous researchers have neglected a second set of GAs. This point is significant, because for this second set of adjectives, the standard of comparison is set much differently.

Adjectives such as *big*, *little*, *high*, and *low* have been treated as canonical GAs, but they actually belong to one category of GAs, one in which the positive form relies upon the context to set the standard of comparison. To illustrate this aspect of their meaning, I have pointed out that a sentence such as (13) (*Elephants are big*.) can be understood as entailing an implicit comparison to some contextually relevant comparison class. But to see that this is not the case for all GAs, let's consider two other cases.

First, look at the example in (14). In contrast to (13), this example does not mean that the lion is awake compared to other lions or other animals. Instead, it means that lion is not asleep – that it exhibits some degree of being awake (which would matter a great deal if you were a zookeeper or a gazelle). Likewise, (15) does not mean that the reservoir is full *for a reservoir* or *container*; it means that the reservoir is not empty.

14. The lion is awake.

15. The reservoir is full.

We know that these adjectives are gradable, since they can appear in comparative constructions (cf. (16)-(17)).

16. a. The lion is more awake than it was 5 minutes ago.

    b. The lion is wide awake.

17. With all of the rain lately, the reservoir is fuller than usual.

These examples require us to broaden the notion of how the standard of comparison can be assigned for GAs.

What these examples compel us to recognize is that all GAs make reference to a standard of comparison, but not all GAs depend on the context to set the value of the standard. GAs such as *big*, *long*, *high*, and *low* do, but GAs such as *awake* and *full* do not. In the chapter that follows, I will be much more explicit about these differences and what the implications are for the child acquiring language. For now, it is enough to know that the experiments that have investigated children's understanding of GAs up to this point have only provided insight into one (albeit rather large) subset of GAs.

To appreciate children's knowledge of the full range of GAs, we must move beyond physical dimensions and investigate adjectives whose standard is not dependent on a contextually-relevant comparison class. If children have full semantic competence with respect to GAs, they should display the behaviors described in the experiments above with adjectives such as *big*, *long*, *high*, and *low* – that is, they should take into account object kind information, the series of objects, and other features of the discourse context – but they should *resist* doing so for GAs (like *awake* and *closed*) whose standard does not rely on the context. For these adjectives, the standard for the positive form should be set independently of the context (e.g., *awake* means having some minimal degree of awakeness, *closed* will mean having no aperture), while explicit comparisons between two objects with respect to a mutual property that admits of degrees should proceed in the usual manner. If, however, children do not have the correct semantic representations for the entire class of GAs, and treat all GAs as contextually variable in the way that physical dimension terms such as *big* are – or even if they have the correct semantic representations, but have a more general bias to take context into account whenever possible – then children should exhibit behavior seen with context-dependent GAs with terms like *awake* and *closed*.

Anticipating the results from Chapter 4, I present evidence from a series of experiments that demonstrate that children as young as age three do indeed have adult semantic competence for the wider range of GAs. They shift the standard for GAs such as *big* and *long*, but do not do so for adjectives such as *spotted*, *bumpy*, *full*, and *straight*. Based upon these results, we have reason to believe that in the representations of the former type of adjectives, there is room for the context to determine the standard, whereas in the representations of the latter type, the standard is oriented towards some other value. Given this picture of linguistic comprehension at the preschool age, we can then step back, consider the acquisition process, and ask what kind of information would enable children to learn about semantic differences that exist within this class of adjectives. That question is addressed in Chapters 5 and 6.

Chapter 3: The Semantics of Gradable Adjectives and Mapping to Scalar Structure

3.1. Introduction

In the previous chapter I reviewed what we know about children's acquisition of adjectives. Starting with a general discussion of what infants know about the kind of properties adjectives denote, I then described in greater detail what preschoolers know about gradable adjectives (GAs). Presenting evidence concerning young children's knowledge of GAs whose meaning is related to physical dimensions, such as height. Although we have a solid knowledge base about children's ability to take contextual factors into account and assign a contextually-based standard of comparison, we lack complementary evidence showing that they behave differently with other kinds of GAs, those whose standard of comparison is not context-dependent. To understand more clearly what is at stake, in this chapter, I review previous semantic theoretic analyses of GAs.

In § 3.2, I present the traditional diagnostic procedures used to identify GAs, and I discuss what about these syntactic environments licenses the appearance of GAs and what about the meaning of GAs allows them to appear in such environments. In § 3.3, I focus on the positive (non-comparative) version of GAs with respect to their dependence on a contextually-relevant comparison class in order to distinguish between these and GAs that are not context-dependent. I follow other researchers in assuming that a scalar analysis (one in which GAs establish relations between entities and degrees on a scale) best accounts for the relevant facts. In § 3.4, I present certain assumptions about the structure of scales that are crucial for the language acquisition questions at hand. Finally, in § 3.5, I discuss differences in scalar structure that give rise to different classes of gradable adjectives based on how the standard of comparison is computed. This discussion then sets the stage for the set of experiments presented in Chapter 4.

3.2. Identifying Gradable Adjectives

3.2.1. Appearance in Comparative Constructions

In the previous chapter I said that GAs are those adjectives which can appear felicitously in comparative constructions, and that this is a consequence of their meaning. Here, I will begin to elaborate on this claim by looking at a range of typical comparative constructions.[19] In each of the sentences in (18), a GA appears in a comparative. These constructions differ with respect to a number of features: the GA may appear in predicative (18a-e) or attributive (prenominal) position (18f); the GA may be the positive or negative polar member of an antonym pair ((18a) vs. (18b)); the comparative may be a superlative (18b) or an equative (18d); a measure phrase may be present (18a, c); and the comparative form itself may differ (cf. (18a) vs. (18c), etc.).

18. a. Rose is (two inches) <u>taller</u> than Lizzie.
    b. Valerie is the <u>shortest</u> of all of them.
    c. Julia is (much) more <u>intelligent</u> than Doris.
    d. Martha is as <u>enthusiastic</u> as Joan.

---

[19] I will not be concerned with comparison involving nominals (e.g., *X is a better student than Y*, *X has less leaves than Y*, *X is more of a vegetable than a fruit*) since I am concerned with the adjectives appearing in comparatives, not the comparative constructions themselves. Sapir (1944)'s claim is that every 'quantifiable' (e.g., object, event, property) can be compared with respect to 'mores' and 'lesses' of different features and so may appear in comparatives (e.g., *more*/*much of*/*half of the house*). See also Morzycki (2006) for an extensive discussion of cases such as *big idiot*/#*small idiot*, which leads him to propose a nominal degree head similar to the one proposed for adjectives.

e. Prudence was so tired that she fell asleep on the bus.

f. Lucy drove a longer distance than Loretta.

Non-gradable adjectives are infelicitous in such constructions (19).

19. a. #Eleanor is deader than Magil.

b. #Vera's clock is more biological than Michelle's.

c. #Pam is more pregnant than Nancy.

The comparative environment has been used as a diagnostic for gradability for as long as GAs have been studied (cf. Clark, 1970; Sapir, 1944; Seuren, 1978; Wheeler, 1972; *inter alia*), in part because it makes such a clear distinction between these two classes of adjectives.

More specifically, constructions such as those in (18) involve a comparison among at least two entities with respect to the extent or amount of a property (cf. Heim, 1985). The comparative (and more specifically, the comparative morpheme, e.g., *-er*) sets up an ordering relation between two degrees related to this property – the reference value (in the sentences above, the property predicated of the subject) and the standard value (cf. Bartsch & Vennemann, 1972a, b; Kennedy, 1999, 2001; Kennedy & McNally, 2005).[20] What allows GAs to appear in such constructions, and what bars non-gradable adjectives from appearing in them is that the former admit of degrees, or allow for a partial ordering of the domain of the adjectives, while the latter do not (cf. Bierwisch, 1989; Cresswell, 1976; Hellan, 1981; Kamp, 1975; Kennedy, 1999;

[20] See Bartsch and Vennemann (1972a), Kennedy (1999), and von Stechow (1984) for a combined introduction to the syntax and semantics of comparative constructions. Though I will not say anything about the syntactic configuration of adjectival phrases or the semantic compositionality of adjectives with measure phrases, there is a rich literature devoted to this topic (cf. Abney, 1987; Bierwisch, 1989; Cinque, 1999; Corver, 1990; Grimshaw, 1991; Kennedy, 1999, 2007; Kennedy & McNally, 2005; Klein, 1980; Schwarzschild, 2005; *inter alia*).

Klein, 1980; McConnell-Ginet, 1973; Seuren, 1978; Wallace, 1972; Xiang, 2006; *inter alia*).[21] That is, entities may differ in their degree of height or intelligence, but not their degree of deadness.[22] Any reading of the comparatives in (19) which seem to be acceptable is coerced.[23] For example, to the extent that (19c) sounds good, it is because something else is being compared (e.g., the baby's gestational age, or the size of the women's stomachs) and not the actually degree of pregnancy.[24]

3.2.2. Appearance in Other Environments

Apart from the traditional comparative constructions, GAs can also be identified by their appearance in a number of other environments. They participate in *too*/*enough* constructions (20) (Hacquard 2006; Heim, 2000; Meier, 2003); they can be modified by intensifiers, also known as 'adverbs of degree' (21) (Bolinger, 1967b, 1972; Klein, 1980; Wheeler, 1972); and

---

[21] The ordering is partial, and not total, because elements of the set may be assigned the same value.

[22] I follow Cresswell (1976), Heim (1985), Hellan, (1981), Seuren (1978), and more recently Kennedy (1999, 2001, 2007), Kennedy and McNally (2005), and Schwarzschild and Wilkinson (2002) in assuming the ontological status of degrees. See Klein (1980) for a critique of analyses based on degrees, specifically Cresswell (1976), which he identifies as a prototypical case in point. For an explicit comparison of 'vague predicate' analyses (cf. Klein 1980; McConnell-Ginet, 1973) with those based on degrees on scales, see Kennedy (1999).

[23] See Bierwisch (1989) for discussion of gradable vs. non-gradable predicates.

[24] Take, as another example, the following sentence from an online NY Times article by Edward Rothstein published May 24, 2007, "this pastoral scene is a glimpse of the world just after the expulsion from the Garden of Eden, in which dinosaurs are still apparently as herbivorous as humans, and all are enjoying a little calm in the days after the fall." *Herbivorous* should be a canonical non-gradable adjective, so what is it doing in this equative construction? In this instance its meaning has been coerced into a gradable one in order to highlight the author's skepticism about a scene one can witness at a newly-constructed Creation Museum in Kentucky, where dinosaurs and humans are displayed as co-existing peacefully on the earth.

they routinely allow pitch accenting for intensification (22) (Bolinger, 1967b).[25] While it may be that some non-gradable adjectives share these features, they do not pattern this way as a class, as gradable adjectives do.

20. a. Sadie is old enough to drive.

    b. Lil was clever enough to solve the riddle.

21. a. *very big*, *overly cautious*, *less interesting*, *particularly impressive*, *incredibly good*

    b. *#very biological*, *#overly wooden*, *#less Russian*, *#particularly plastic*, *#incredibly dead*

22. a. It's (so) DEEP / LONG!

    b. *It's BIOLOGICAL! (Bolinger, 1967b)

What these environments have in common with comparatives is that they also relate the reference value (e.g., Sadie's age in (20a)) to a standard value (e.g., the age at which one is able to drive). GAs are therefore easily identified by their appearance in a number of surface environments, and where they appear they are able to because they satisfy the requirements of such environments by allowing differences in degree of the relevant property.

In a model-theoretic analysis based on degrees, this difference in meaning between GAs and non-gradable adjectives is captured in their semantic representations. Following Bartsch and

---

[25] Geach (1956) and Bolinger (1967b, 1972) draw a distinction between adjectives which appear freely in predicate position and those which do not. Compare, for example, (i) and (ii). However, this is not a hard and fast test, as Bolinger (1967b) notes, since there are numerous non-gradable adjectives which are licensed in predicative position (cf. *All of the houseplants were dead*, *The furniture was wooden*, etc.).

(i) a. The waitress is nice.

    b. That is a nice waitress.

(ii) a. *The debate was parliamentary.

    b. That was a parliamentary debate. (Bolinger, 1972)

Vennemann (1972a, b), Cresswell (1976), Kennedy (1999, 2007), and von Stechow (1984) I will assume that GAs are measure functions of type ⟨*e, d*⟩, mapping individuals onto degrees on a scale.[26, 27] By contrast, non-gradable adjectives are of type ⟨*e, t*⟩: the truth value of a sentence in which they appear is simply based upon whether or not the property holds of the individual or not.[28] This analysis of GAs has an important implication: even in the positive, non-comparative form, GAs have a comparative meaning.

3.3. The Positive Form of GAs

Abstracting away from semantic types and frameworks, the observation that GAs in the positive form are implicitly comparative is a long-standing one. In his analysis of the positive form of GAs, von Stechow (1984) cites the following examples adapted from Aristotle's

---

[26] Bartsch and Vennemann (1972b) actually describe their measure function (f^M) as a two-place function, with individuals as the first argument and *dimensions* as the second argument.

[27] Different semantic types have been proposed for GAs. In this dissertation I assume that they are of type ⟨*e, d*⟩, since all that matters here is that they incorporate a degree variable in their denotation in contrast to non-gradable adjectives, which do not. Alternatively, they could be of type ⟨d, ⟨e, t⟩⟩, in which case their domain is not individuals but degrees, and their range is not degrees but properties of individuals. See Kennedy (1999), Kennedy & McNally (2005), and references cited in footnote no. 2 of Kennedy (2007) for brief discussions.

[28] As a consequence, although both meanings take an individual of type e as their first, argument, the composition in the two instances results in a very different meaning. The expression resulting from composition with a GA yields a degree variable, which must then compose with another function that takes a degree as the first argument. Composition with a non-gradable adjective results in a proposition-level truth value.

Categories (23) demonstrating this fact.[29] In (23a, b) the sizes of the entities in subject position are compared to members of the same category (i.e., other corn or mountains), while in (23c), the number of people in Athens and the village is compared to the number at other times of year.

23. a. This corn is big.

    b. This mountain is small.

    c. In the summer there are only [a] few people in Athens, but there are a lot of people in the village.

This type of implicit comparison associated with the positive form has also played a central role in a number of analyses over the years (cf. Bierwisch, 1967, 1989; Cresswell, 1976; Kamp, 1975; Kennedy, 1999, 2007; Klein, 1980, 1982; Ludlow, 1989; McConnell-Ginet, 1973; Montague, 1974; Sapir, 1944; Seuren, 1978; Vendler, 1968; Wallace, 1972; Wheeler, 1972). The reason for

---

[29] Actually, what Aristotle writes is much more informative with respect to the role of the comparison class in assigning a truth value to a sentence in which a GA appears, and to the relative/absolute terminology I will discuss later in the chapter. In his discussion of contraries, he says that

> A man might, indeed, argue that *much* was the contrary of *little*, and *great* of *small*. But these are not quantitative, but relative; things are not great or small absolutely, they are so called rather as the result of an act of comparison. For instance, a mountain is called small, a grain large, in virtue of the fact that the latter is greater than others of its kind, the former less. Thus there is a reference here to an external standard, for if the terms *great* and *small* were used absolutely, a mountain would never be called small or a grain large. Again, we say that there are many people in a village, and few in Athens, although those in the city are many times as numerous as those in the village: or we say that a house has many in it, and a theatre few, though those in the theatre far outnumber those in the house. The terms *two cubits long*, *three cubits long*, and so on indicate quantity, the terms *great* and *small* indicate relation, for they have reference to an external standard. It is, therefore, plain that these are to be classed as relative. [online resource]

this interest derives from two main aspects of this comparison related to their status as vague predicates,. To fully appreciate these aspects, which I will maintain as distinct from each other, let's look at a sentence containing a GA in a positive form, as in (24).

24. Rita is tall.

We can first observe that even if we know Rita's height, it is not possible to assign a truth value to the proposition expressed in this sentence without also knowing the standard (or norm) to which we are comparing her height. If Rita is 6' tall, she is tall for a human female, but not for an apatosaurus. This aspect has to do with the fact that GAs are interpreted with respect to a contextually-relevant *comparison class* (cf. Cresswell, 1976; Kennedy, 1999; Kennedy & McNally, 1999, 2005; Klein, 1980, 1991; Siegel, 1979; Wheeler, 1972).[30] Second, even when a comparison class has been identified, it is not necessarily clear where the cutoff should be drawn between those things that are tall and those that are not. There are no 'sharp boundaries' (cf. Wright, 1975). The choice of an exact number to serve as the boundary between the two categories 'tall' and 'not tall' is rather arbitrary, and the boundary can shift depending on the comparison class. (See Kennedy (2007) and Rusiecki (1985) and references cited therein for a review of related work.)

Before discussing each of these aspects in turn, it is necessary to make one clarification. Both of these aspects of vagueness should be distinguished from the problem of *indeterminacy*, which has to do with an inability to settle upon the interpretation of an adjective without further information being supplied by the modified noun or the discourse context. For example, in (25),

[30] Klein (1980) attributes the earliest published reference to a 'comparison class' to Hare (1952).

what it means to be *good* or *talented* depends on the noun (cf. Kamp, 1975; Katz, 1964; McConnell-Ginet, 1979; Ross, 1930; Vendler, 1968; *inter alia*).

25. a. That is a good knife/dessert/guess/magazine.

    b. He is a talented chef/game theorist/yogi/violinist.

To explain these examples, we can say that some GAs – such as *good* and *talented* – may be associated with more than one ordering on a domain, so in these cases, not only can the boundary shift, but so can the relevant dimensions (Kennedy, 1999; McConnell-Ginet, 1979). This issue is therefore a problem of ambiguity in meaning rather than uncertainty about the domain of reference or the delineation of categories. Thus, when comparing, e.g., how *long* something is, it is necessary first to determine whether one is interested in making a comparison with respect to temporal duration, physical length, or wagers.[31]

3.3.1. The Role of the Comparison Class

I said above that in order to assign a truth value to (24), it is necessary to identify the domain. Determining whether or not Rita is tall depends on her membership in a relevant

---

[31] The issue of settling the dimension of comparison feeds into the observation that comparison cannot take place across two different and incompatible dimensions. This problem of *incommensurability* can be observed in the following example, which Kennedy (1999) notes is conceptually accessible and should have a meaning such as the one in (iiib), but is barred linguistically.

(iii) a. #My copy of *The Brothers Karamazov* is heavier than my copy of *The Idiot* is old. (104)

b. My copy of *The Brothers Karamazov* is higher on a scale of heaviness than my copy of *The Idiot* is on a scale of age. (107)

Likewise, the truth value of the following sentence depends on whether the calculation is based on population or land area. See also Kennedy (2007).

(iv) Chicago is larger than Dallas.

comparison class. This type of sentence differs from those in (26), where we simply have to look to the individual in the real world to determine whether the property predicated of Rita holds in order to assign a truth value to the proposition.

26. a. Rita is four-legged.

    b. Rita is a brown animal.

This difference between predicates such as *tall* and *four-legged* has been cast as a fundamental difference in meaning: while the former make reference to a comparison class, the latter do not. Put another way, in order to decide whether Rita is tall, it is necessary to know *what counts as tall* in the context at hand. This is not the case for deciding whether she is four-legged.[32] The comparison class may be supplied by a noun in the sentence, as in (27a,b), extrasententially, or from the discourse context or common ground (cf. Barker, 2002; Bartsch & Vennemann, 1972b; Bierwisch, 1989; Higginbotham, 1985; Ludlow, 1989; Platts, 1979; von Stechow, 1984; Warren, 1988).

27. a. That is a big flea. (Montague, 1974)

    b. Jumbo is a large elephant. (Vendler, 1968; Wallace, 1972)

Adjectives such as those in (26) are therefore termed *intersective*.[33] For example, to account for (26b) we can say that Rita is a member of the set that stands at the intersection of two sets –

[32] Parsons (1972) discusses the ambiguity associated with a sentence such as (v), which can mean that *X is small for a four-footed animal*, or that *X is a small for an animal, and is also four-footed*.

(v) X is a small four-footed animal.

[33] But note that not all non-gradable adjectives are intersective (cf. *former*, *alleged*).

animals and brown things.[34] By contrast, adjectives such as *big* and *large* are *subsective*: (27a) does not mean that the animal in question is both big and a flea; rather it is big *for a flea*. The precise contribution and syntactic status of this qualificational *for*-PP (and others like it, such as *as an elephant* or *as elephants go*, cf. Vendler, 1968) has played an important role in the discussion of such adjectives (cf. Kennedy, 2007; Rusiecki, 1985).[35] At the most basic level, it provides the comparison class relevant for the interpretation of the GA.

As noted by Kennedy (1999), both the 'vague predicate' analysis and the 'scalar' analysis of GAs rely on the existence of a contextually-salient comparison class, and for both analyses, a change in the comparison class may affect the truth value of a sentence. However, the role of the comparison class differs in the two analyses.[36] In the 'vague predicate' analysis, the comparison class is a function that picks out a subset of the discourse context, which the adjective then partitions into three disjoint subsets: one for which the predicate holds, one for which it does not, and an extension gap, for which it is indeterminate.[37] Klein (1980) proposes that successive reapplication of the predicate function on the extension gap can narrow the set down to two

---

[34] Simply for expository purposes, I treat the color term *brown* as an intersective adjective here, but this can be misleading. (See Bolinger, 1967b; Edgington, 2001; McConnell-Ginet, 1973; Raffman, 1994; Rusiecki, 1985; Sapir, 1944; Wheeler, 1972; and Wright, 1975 for (more philosophical) discussions of vagueness among color terms.) A host of empirical investigations looking at the perception and categorization of color and numerous analyses of the various linguistic, cultural, and cognitive perspectives on color have been carried out over the years, the list of which is too long to include in a footnote.

[35] There are ambiguous cases in which a GA may have both types of interpretations – e.g., *happy agent* (Bolinger, 1967a), *beautiful dancer* (Larson, 1988; Siegel, 1976; Vendler 1968) – which I do not discuss here.

[36] See Staab and Hahn (1998) for a model of how comparison classes are constructed 'on the fly'.

[37] Cruse (1986) calls this the 'pivotal region'.

subsets, thereby rendering the adjective more precise. In the 'scalar' analysis, the comparison class is used to fix the value of the standard of comparison, which is expressed as a degree. The predicate denotes a relation between the entity in question and this standard degree. A sentence such as (24) therefore means something like *The degree to which Rita is tall is at least as high as the standard degree*.

3.3.2. Sliding Scales and Vagueness

Now, even when a comparison class has been established, the meaning of a GA in the positive form is still vague. To understand this, let's look again at (27b). Recall that this sentence means something like *Jumbo is large for an elephant*. Knowing this does not help to sharpen the boundary for what counts as *large* for elephants. As Platts (1979) remarked, offering an interpretation of (27b) such as (28a) (cf. Langford, 1968, after G. E. Moore) or (28b) (cf. Fillmore, 1971, pp. 536-537; Katz, 1972) will not do, because such comparatives render (27b) more precise than it actually is, giving us a clearer idea of where Jumbo stands with respect to the size of other elephants. (See also Bogusławski (1975).) In fact, appearance in any comparative construction will have this effect, given the relation it expresses between a reference degree and a standard degree.

28. a. Jumbo is larger than most elephants.

    b. Jumbo is larger than the average elephant.

An analysis that relates Jumbo to a degree, which is then compared to the standard degree offered by the comparison class, does not run up against this problem, and neither does an analysis that leaves an extension gap between elephants that are not large and those that are. This issue is basically this: where does one draw the line to separate cases where the property holds from those where it does not?

As a further illustration, consider the following sentence, a version of which is discussed in Bogusławski (1975) and Kennedy (2007).

29. Lizzie's height is greater than the average height of a female gymnast, but she is still not tall for a female gymnast.

If *tall* simply meant 'greater than the average', this sentence would be a contradiction. However, it is possible that Lizzie's height only just exceeds the average height of a gymnast, while there are other gymnasts who are several inches above the average height, for whom it is clearly true that they are tall gymnasts. (Dominique Dawes did not stand out for being mere millimeters taller than the other female gymnasts on her team.) Adding a measure phrase is another way to precisify the meaning of a GA (cf. Kennedy, 2007; Klein, 1980). In (30) *tall* is no longer vague.

30. Rita is 5'6" tall.

The meaning of GAs such as *tall* is vague with respect to the demarcation of the cutoff point. Wallace (1972) describes this situation by saying that there is 'slack' in the application of gradable predicates.[38] More typically, these predicates are described as giving rise to 'borderline cases', cases where it seems reasonable to describe something both as, e.g., *tall* or *not tall* (or where classification in one or other category is not clear-cut). That is, there are no sharp boundaries. The philosophical literature is replete with vigorous and often contentious discussions of vagueness and the paradoxes arising from such cases[39], and I do not intend to

---

[38] Kamp (1975) notes that vagueness is also possible with categories other than adjectives, and offers the example of *cat*. See Engel (1989) and Sapir (1944) for discussion of how categories other than adjectives admit of degrees.

[39] Two paradoxes, in particular, are usually discussed. The first, and the most popular, is the Sorites paradox, or the paradox of the heap: 1 grain of sand is not a heap, and adding one more grain does not make it a heap, but at what point after adding grains of sand do you have a heap? Another version of this paradox is the 'falakros' paradox, or

become involved in those debates. What is important here is that GAs in the positive form such as *tall* give rise to borderline cases precisely because they make reference to a standard of comparison that is based on a contextually relevant comparison class. (For further discussions of the vagueness of GAs in the semantics literature, see Barker, 2001; Fara, 2000; Kamp, 1975; Kennedy, 1999, 2007; Klein, 1980, 1982; and McConnell-Ginet. 1973.)

3.3.3. Degrees in the Positive Form

Having explored the inherently comparative and vague nature of GAs in the positive form, we are now in a position to examine how the semantic representation of GAs can account for these cases. As I stated above, I will adopt a scalar analysis of GAs. Following a number of previous analyses, I will make three main assumptions about the lexical representations of GAs, and consequently about sentences such as (24).

First, even in the bare, positive form, GAs encode an implicit degree variable (Cresswell, 1976; Kennedy, 1999, 2007; Kennedy & McNally, 2005; Seuren, 1978). In (24), we are therefore saying something like, *Rita is* d *degrees tall* or d-*much tall*, *which is greater than the standard degree* d. Seuren (1978) argues that GAs can never occur without some sort of 'degree phrase'. If there is not an overt one (e.g., *much*, *3 inches*, *as ever*, *-er*, etc.), then there must be a covert one. (See Cresswell (1976) for a similar argument.) Evidence for this claim

---

the paradox of the bald man: a man with one hair on his head is bald, a man with 2 hairs on his head is bald, but a man with 50,000 hairs on his head is not bald, so where is the line drawn? The second paradox is known as Wang's paradox (Dummett, 1975; Wright, 1975): 1 is a small number; if *n* is a small number, then *n*+1 is a small number, and therefore all numbers ($\geq n$) are small. These examinations are also especially interesting because of what they bring to bear on the Law of the Excluded Middle (P or ¬P, with no third option). See Rolf (1984) for a rather clear discussion of these issues.

comes from Chinese. Xiang (2006) shows that dimensional adjectives such as *tall* are barred in their bare form and require some sort of degree modifier: (0a) is only permissible if a reference and standard of comparison have already been made salient in the discourse.

31. a.??Ta gao (Xiang, 2006, (49))

he tall

'He is tall.'

b. Ta hen / you dian er / feichang / bu zenme / tai gao

he very / a little / very / not how / extremely tall

'He is very / a little / not very much / extremely tall.'

Second, I will assume that GAs in the bare positive form are actually not *bare*: they contain a null positive degree morpheme (POS). The function of this morpheme is similar to the comparative morpheme: it encodes a relation between the degree argument in the representation of the GA and the degree of the standard of comparison. Following Kennedy (2007), I will adopt the representation in (32) for this POS morpheme (although see Bartsch & Vennemann, 1972b; Bogusławski, 1975; Fara, 2000; Kennedy & Levin, to appear; Kennedy & McNally, 2005; Kennedy, 2007; and von Stechow, 1984 for further discussion of and variations on the representation of the positive morpheme).[40] In (32) x is variable ranging over individuals; g represents a GA measure function, taking objects and mapping them onto degrees; and s is a context-sensitive function from measure functions (GAs) to degrees, which returns a standard of comparison compatible with both the context and the properties of the GA (e.g., domain, dimension). This morpheme ensures that the degree of the individual stands in a 'greater-than-

[40] See Rett (2007) for discussion of the choice between a POS or an EVAL morpheme, based in large part on work by Bierwisch (1989).

or-equal-to' relation to the standard of comparison, giving us the right kind of meaning for a sentence like (24): *tall*(Rita) $\succeq$ s(*tall*) indicates that the degree to which Rita is tall meets or exceeds the standard for what counts as *tall* in the context.

32. $[\![[_{\text{Deg}}\ \mathbf{pos}]]\!] = \lambda g \lambda x. g(x) \succeq s(g)$ (Kennedy, 2007, (27))

Third, I will assume that degrees are elements of scales (cf. Bierwisch, 1989; Kennedy, 1999, 2001; Schwarzschild & Wilkinson, 2002; Seuren, 1978; von Stechow, 1984). This may seem rather obvious, but this much must be clear before we go on to discuss additional aspects of scalar structure. The need for a scale for comparisons and measurement arises from the intuition that when we perform mental comparisons, we are establishing an ordering relation between multiple entities with respect to 'more' or 'less' along a particular parameter (cf. Cresswell, 1976; Bierwisch, 1989; Huttenlocher & Higgins, 1971; Sapir, 1944). As observed by Rusiecki (1985) among others, the concept of a scale is necessary for comparisons in domains other than adjectives ((33)-(35)).

33. She is more a wife than a mother. (Rusiecki, 1985)

34. Carol earns more money than Anna (does). (Rusiecki, 1985)

35. Molly arrived (shortly) before dinner began/just as we sat down to eat.

Once we express the properties which are predicated of these entities as degrees, we can think of these degrees as occupying positions on a scale, and we can therefore speak of them (or the values they represent) in terms of 'greater-than', 'lesser-than', and 'equal-to' relations, as well as talk about the properties of these relations (e.g., transitivity, (a)symmetry, etc.). I will also assume that degrees (or entities represented by degrees) are instantiated as extents or intervals on a scale rather than points (cf. Bierwisch, 1989; Bolinger, 1967b; Kennedy, 1999, 2001; Schwarzschild & Wilkinson, 2002; Seuren, 1978; von Stechow, 1984), although that distinction will not substantially affect my analyses, since my focus is on the structure of the scales, not the comparisons that take place between objects associated with the scale.

3.4. Scales

I began this chapter by presenting a number of diagnostics for identifying GAs, the appearance of GAs in comparative constructions among them. I then explained the features of these environments that license GAs and what it is about the lexical-semantic representations of GAs that makes it possible for them to appear in these environments. Comparatives involve an ordering relation among degrees, and only those elements that permit variation in degrees can appear in them. GAs encode a degree argument in their representation, while non-GAs do not. I showed that even in the positive (non-comparative) form, GAs are implicitly comparative – they make reference to a contextually-relevant comparison class. A POS morpheme was responsible for the implicit comparison observed in the non-comparative form by ensuring that the degree is related to the standard. The need to identify the comparison class and the fuzzy boundaries for categorizing items in this comparison class with respect to the gradable property make these lexical items inherently vague. Finally, I noted that the degrees that play a central role in these comparisons map onto a scale. In this section, I lay out the fundamental characteristics of scales, which will lead us into a discussion of differences in scalar structure.

Scales are straight-line (one-dimensional) functions constructed out of a set of degrees ordered in relation to each other according to some dimensional 'parameter' or 'field', which determines how the objects in the adjective's domain are ordered (cf. Bierwisch, 1989; Cresswell, 1976; Givón, 1970; Kennedy, 1999; Kennedy & McNally, 2005; Osgood, Suci, Tannenbaum, 1957; Rusiecki, 1985). For example, we can construct a scale for height, temperature, intensity, and so forth, which allows us to set up a partial order of individuals with respect to these values. We can think of positive values extending in one direction, and negative values extending in the other, as with real numbers (e.g., in the *big*/*small* antonym pair, *big* extends in a positive direction, *small* in a negative direction). Scales, then, are directional. The points on the scale may be assigned names for the units (e.g., inch, degree Fahrenheit, decibel, Hertz, etc.), and differentials measure the distance between these points (e.g., *3 inches*, *9 degrees*, *5 hours*, *a lot*, etc.). In addition, the dimension, by indicating the kind of measurement, has implications for the kind of measure phrases that are licensed (e.g., *90-degree water*/**90 degrees of water*, *2-foot cable*/*2 feet of cable*) (cf. Schwarzschild, 2002). Finally, a scale has parts (subsets, or intervals), which stand in a 'part-of' relation to larger intervals on the scale that contains them. In this respect a scale is a mass-like object (cf. Schwarzschild & Wilkinson, 2002). Adjectives can therefore differ with respect to any of these parameters – dimension, order/direction, units, and the differentials and measure phrases that are licensed.

Aside from the *composition* of the scale, we can also talk about the *structure* of the scale. These two aspects are closely connected, but two points in particular are relevant when we talk about the structure of a scale, that are relevant. First, we can talk about the presence or absence of endpoints (i.e., a *zero value* and/or a *maximal* value[41, 42]) – that is, whether the scale is open or closed. Given the options for a unary dimension presented to us by geometry, we have three

---

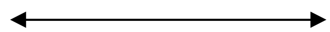
[41] The term 'maximal' used should not be confused with Bierwisch (1989)'s MAX, which is an element in the lexical entry of a GA specifying the maximal dimension for that adjective.

[42] Note that with scales composed of real numbers there is no maximal element, since numbers range to infinity. In such cases, maximality is imposed based on the set in question, not the full possible range.

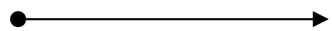
options: a line, which extends into space on both ends; a ray, which is closed at only one end; and a segment, which is closed on both ends.[43] This set of possibilities is illustrated in Figure 6.

Figure 6: Three possible scalar structures

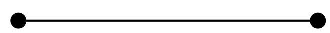
a. completely open-ended



b. closed at one end



c. closed at both ends



Second, the presence or absence of endpoints has implications for the standard of comparison. When one or more endpoints are present, the endpoint(s) will serve as the standard. When there are no endpoints (i.e., when the scale is symmetrically open-ended) the standard will be located somewhere along the scale and will require contextual information in order for the value to be fixed.

3.5. Differences in Scalar Structure

In this section, I will first demonstrate that entailment patterns of sentences involving GAs can capture differences among adjectives, and more specifically among GAs. The relevant sentences here include those with: a GA as a prenominal adjectival modifier; a GA in a

---

[43] The directionality of scales allows us a fourth option, if we assume that a scale can be closed at either the minimal or maximal end. Kennedy & McNally (2005) pursue this possibility (pp. 354-355). I will leave the issue of directionality aside, since we are able to make strong generalizations and predictions without entertaining this possibility. In fact, their examples in (26)-(27) could be collapsed into one type of scale, as in (c) above, by referencing the presence or absence of a particular property, a point I discuss below.

comparative construction; a negated GA; and a GA modified by adverbials such as *almost* and *slightly*. After observing how these classes of GAs emerge through entailment patterns, I will review how previous analyses have appealed to differences in scalar structure to capture their representational differences. The main focus here will be on the presence or absence of scalar endpoints. Finally, I will review how the presence or absence of endpoints directly relates to differences in how the standard of comparison is computed.

Because my aim is to lay the theoretical foundation for the experimental that follow, which will allow us to pose more sophisticated, theoretically-motivated questions about the language acquisition process, I will have to ignore some of the more fine-grained observations made by previous researchers. For example, there may be further classifications within a particular grouping, exceptional adjectives that challenge or support certain classifications, and contextual variables that exhibit interesting effects. For reasons of clarity and space, I will simply encourage interested readers to go directly to the sources I cite for further discussion.

3.5.1. Entailment Patterns

Hamann (1991), following Bartsch & Vennemann (1972a), distinguishes between three classes of adjectives through entailment patterns generated by a sentence such as the one in (36), where *A* is an adjective and *N* is a noun.[44]

36. X is an A N

37. X is a <u>carnivorous</u> / <u>four-legged</u> / <u>Greek</u> / <u>dead</u> N

[44] I state these patterns in terms of entailments, but one could just as easily state them in terms of the intersection of denotations and part-of relations, as in Kamp & Partee (1995) (i.e., *X is an A N* ⇒ *X is an N, X is A* would be recast as ⟦carnivorous N⟧ = ⟦carnivorous⟧ ∩ ⟦N⟧).

a. ⇒ X is an N

b. ⇒ X is A

38. X is a <u>tall</u> / <u>large</u> / <u>small</u> / <u>heavy</u> / <u>good</u> N

a. ⇒ X is an N

b. ⇏ X is A

39. X is a <u>fake</u> / <u>former</u> / <u>reported</u> / <u>counterfeit</u> N

a. ⇏ X is an N

b. ⇏ X is A

Privative adjectives such as those in (39) fall outside the scope of the current discussion, and I do not discuss them here. For my purpose, the relevant distinction here is the one between the adjectives in (37) and (38). The intersective adjectives in (37) have been called *absolute* adjectives and the subsective adjectives in (38) *relative* adjectives (Bartsch & Vennemann, 1972a; Rusiecki, 1985; Siegel, 1979; Unger, 1975). (See discussion of (26) and (27) above.) Siegel (1979) describes the difference by saying that absolute adjectives are extensional and modify individuals, while relative adjectives are intensional and rely upon additional information related to the noun they modify.[45] Thus, according to her, the ambiguity in the phrase *beautiful dancer* arises because *beautiful* has both a relative and an absolute reading. (Note, then, that despite the labels assigned to *adjectives* here, what is actually being classified is the possible *readings* of these adjectives.)

---

[45] Siegel (1979) used this terminology in her discussion of *measure* adjectives. In her work she drew a distinction between measurement and comparison, which she saw as a much more widespread phenomenon.

Notice, though, that the set of examples in (37) includes non-gradable adjectives.[46] However, some non-gradable adjectives also license the same entailments (40).

40. X is a clean / straight / wet / spotted N

a. ⇒ X is an N

b. ⇒ X is A

Because we are interested in teasing apart differences in scalar structure, and non-gradable adjectives do not map onto scales, from here on out, I will only be exploring differences between *gradable* adjectives that license entailments such as those in (37)/(40) and (38).

Other tests of entailment allow us to probe the difference between absolute and relative GAs further. For example, given that both types of adjectives are gradable, they are able to

[46] It is hard to say whether Bartsch and Vennemann (1972a) or Siegel (1979) intended for the absolute/relative distinction to map onto a non-gradable/gradable one, or if they meant this distinction to signal a difference based on the role of context-dependence. Their sets of examples leave room for ambiguity but always include clearly non-gradable adjectives such as those listed above. For example, Bartsch and Vennemann note that absolute adjectives can have a secondary relative interpretation, which becomes accessible when the positive form is modified by an intensifier such as *very*, and they also talk about relative adjectives relying on the context for interpretation. Siegel's Table 11 (pg. 227), where she lists adjectives that have absolute or relative interpretations or both, is a mixed bag. Her list of absolute adjectives includes *speckled*, *nude*, *naughty*, and *drunk*, and she allows adjectives such as *beautiful*, *clever*, *skillful*, *crazy*, *true*, and *old* to have both absolute and relative interpretations. Furthermore, in the list of relative adjectives, she includes *former*, *ostensible*, and *alleged*. I choose to follow Bartsch and Vennemann in viewing the context as the distinguishing factor when focusing solely on differences within GAs. This choice seems to make sense, given the way the terminology is used later by Rusiecki (1985), Kennedy and McNally (2005), and Kennedy (2007).

appear in comparative constructions. However, they license different entailments in a typical comparative such as the one in (41).

41. NP_1 is Aer than NP_2

Comparatives with absolute GAs such as those in (42) entail that the first, but not necessarily the second entity has the property in question. By contrast, relative GAs such as those in (43) have no such entailments.

42. Absolute GAs

a. My towel is wetter then yours. ⇒ My towel is wet. (Rusiecki, 1985)

b. The gold is more impure than the platinum. ⇒ The gold is impure.

(cf. Kennedy & McNally, 2005, (35a))

43. Relative GAs

a. John is older than Peter. ⇏ John is old. (Rusiecki, 1985)

b. John is taller than Peter. ⇏ John is tall.

This difference is also highlighted when the first proposition is explicitly contradicted, as in ((44)-(45)).

44. Absolute GAs

a. #My towel is wetter then yours, but it is still not wet.

b. #The gold is more impure than the platinum, but the gold is still pure.

45. Relative GAs

a. John is older than Peter, but neither one of them is old.

b. John is taller than Peter, but both of them are pretty short.

Sentences in which one member of the pair is negated show a similar difference in entailment patterns between the two types of GAs, as seen by comparing the examples with absolute GAs in (46)-(47) with the examples with relative GAs in (48) (cf. Cruse, 1986; Rotstein & Winter, 2004; Yoon, 1996).

46. Absolute GAs (*clean*/*dirty*)

a. It's clean. ⇒ It's not dirty.

b. It's not clean. ⇔ It's at least slightly dirty.

c. It's dirty ⇒ It's not clean.

d. It's not dirty. ?⇒ It's clean.

47. Absolute GAs (*open*/*closed*)

a. It's open. ⇔ It's not closed

b. It's not open. ⇔ It's closed

48. Relative GAs (*long*/*short*)

a. It's long ⇒ It's not short.

b. It's not long. ⇏ It's short.

c. It's short ⇒ It's not long.

d. It's not short. ⇏ It's long.

Absolute and relative GAs also differ in their ability to be modified by adverbials such as *almost* as in (49)-(50). Here, in fact, an interesting difference emerges: modifiers such as *almost* are generally not felicitous with relative GAs (49), but are felicitous with only some absolute GAs (compare (50a-b)). (See also (19)-(20) in Kennedy & McNally (2005).)

49. Relative GAs

a. #It's almost long. (Cruse, 1986; Rotstein & Winter, 2004)

b. #It's almost short. (Cruse, 1986; Rotstein & Winter, 2004)

50. Absolute GAs

a. It's almost clean. (Cruse, 1986; Rotstein & Winter, 2004)

b. #It's almost dirty. (Cruse, 1986; Rotstein & Winter, 2004)

Following the characterization of the members of these absolute pairs by Yoon, 1996; Kamp & Roßdeutscher, 1994; and Rotstein & Winter, 2004, and borrowing terminology from Kennedy & McNally (2005) and Kennedy (2007), I will call absolute GAs such as *dirty* or *wet minimum* standard absolute GAs, because a sentence such as (46c) is true if there is a minimal degree of dirt, and absolute GAs such as *clean* or *dry maximum* standard absolute GA, because a sentence such as (46a) is true if the object is maximally free of dirt. This distinction, then, boils down to the simple question of the existence or absence of the property in question. This fact can be illustrated with the denials in (51).

51. a. #My hands are not wet, but there is some water on them. (*minimum* standard)

(Kennedy & McNally, 2005, (36a))

b. #The children's toys have dirt on them, but they are clean. (*maximum* standard)

(cf. Yoon, 1996)

We also find a difference between maximum and minimum standard GAs in comparative constructions. The examples of absolute GAs in (44) and (51a) above involve minimum standard GAs, and the sentences are infelicitous (e.g., *wet*, *impure*). In contrast to the above examples, maximum standard GAs in the same environments are permissible (52), because the entities being compared can differ with respect to how close they are to the maximal standard without having reached it.

52. Maximum standard absolute GAs

   a. My towel is drier than yours, but it is still not (100%) dry.

   b. The gold is purer than the platinum, but the gold is still not (100%) pure.

   c. The children's room is cleaner than it was 10 minutes ago, but it is still not (totally) clean.

In sentences such as (50a), we also see informative entailment patterns among maximum and minimum standard absolute GAs when the maximum standard member of the pair is modified by *almost* (53). Approaching but not yet reaching the maximal standard, which signals the absence of the relevant property, means that the minimal standard, which signals the presence of the property, still obtains.

53. a. It's almost clean. ⇒ It's (still) dirty.

   b. It's almost dry. ⇒ It's (still) wet.

Modification of absolute GAs by *slightly* creates a pattern opposite that observed in (50a-b) above. While *almost* felicitously modifies *maximum* standard absolute GAs (54), *slightly* felicitously modifies *minimum* standard absolute GAs (55) (exx. from Rotstein & Winter (2004)).

54. a. The work is *almost* complete/#incomplete. (9a)

   b. The explanation is *almost* clear/#unclear. (9c)

   c. The patient is *almost* healthy/#sick.

55. a. The work is *slightly* #complete/incomplete. (37a)

b. The jar is *slightly* #whole/cracked. (37c)

c. The line is *slightly* #straight/curved. (37d)

In support of this difference, Frazier, Clifton, and Stolterfoht (2007) have demonstrated that the processing of such sentences is impaired when the adverbial modifier clashes with the default interpretation of the adjective (whether it is a maximum or minimum standard absolute GA). Such a difference follows from the characterization I offered above.

Based on these patterns, we arrive at the classes of GAs in (56).

56. a. Relative GAs (e.g., *long*/*short*, *old*/*young*)

b. Maximum standard absoluteGAs (e.g., *clean*, *dry*, *healthy*, *full*)

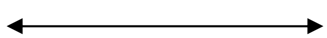
c. Minimum standard absolute GAs (e.g., *dirty*, *wet*, *sick*, *empty*)

3.5.2. Open and Closed Scales

Appealing to scalar structure enables us to account for these differences in entailment patterns among GAs. Recall the scalar structures presented in Figure 6. We can now map these structures to actual adjectival classifications (cf. Figure 7).

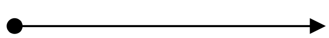
Figure 7: The three scalar structures of GAs

a. Relative GAs (open-ended scale)



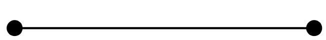
(e.g., *big*/*small*, *old*/*young*)

b. Absolute GAs (scale closed on one end)



(e.g., *clean*/*dirty*, *dry*/*wet*, *straight*/*bent*)

c. Absolute GAs (scale closed on both ends)



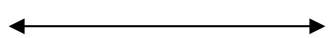
(e.g., *full*/*empty*, *open*/*closed*)

Taking each of these structures in turn, I will review the motivation behind the structure and explain how it accounts for the entailment patterns observed above.

3.5.2.1. Relative GAs and Open Scales

I began our discussion of GAs with relative GAs, adjectives such as *tall*/*short*, *big*/*small*, and *old*/*young* – vague predicates that depend on a comparison class for interpretation and have borderline cases or extension gaps. Characterizations by Cruse (1986); Kennedy (2007); Kennedy & McNally (2005); Rusiecki (1985); and Seuren (1978) isolate the same aspects of relative GAs and their corresponding scalar structure.[47] The dimension of a relative GA must have a value greater than zero. It is not possible to have 0 height, 0 size, or 0 age (so while zero is a lower bound, it cannot be the minimal element). As a result, the scale approaches the zero value asymptotically, but never reaches it. Likewise, there is no upper limit; there is no value that is greater than all others in the set. There are therefore no words indicating the top and bottom of the scale. A comparison class can delimit a range along the scale, but these values ‘move along’ the scale depending on the context. The scalar structure for relative GAs is therefore the one captured in Figure 8.

Figure 8: Relative GAs (open-ended scale)



Given this scalar structure, the reasons underlying the entailment patterns observed above should be clear. Comparisons in the form of *NP_1 is Aer than NP_2* do not entail that *NP_1 is A* because of the role of the comparison class: in comparing the size of two entities relative to each

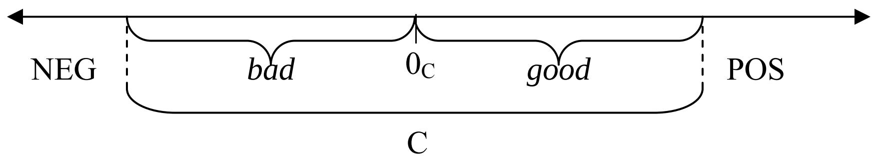
[47] A note on terminology: *relative GAs* also correspond to Seuren (1978)’s *neutralizers* (as opposed to *pointers*), Rusiecki (1985)’s *relative GAs* with an *open symmetric scale* (as opposed to a *closed* scale, or an *asymmetric* scale), and Cruse (1986)’s *antonyms* (as opposed to *complementaries*).

other, it is not necessary that either one be, e.g., big or tall. Likewise, to say that one entity is tall may entail that it is not short, but saying that it is not tall does not entail that it is short, since there are no sharp boundaries for the categorization of entities. Finally, the lack of scalar endpoints and a context-sensitive comparison class has implications for which adverbial modifiers are licensed. Modifiers such as *almost* are not licensed, because there is no maximal endpoint; modified such as *slightly* are not licensed because there is no zero value.

I will forgo any in-depth discussion of differences between dimensional and evaluative adjectives (DAs vs. EAs) (cf. Bierwisch, 1989; Rett, 2007; Xiang, 2006).[48] Bierwisch (1989) argues that polar DAs – GAs that permit measure phrases (e.g., *tall*, *long*, *short*, *old*, *young*, *new*) – occupy different portions of the same scale, while EAs (e.g., *good*, *bad*, *industrious*, *lazy*, *pretty*, *ugly*) map onto different scales. Because Bierwisch allows EA scales to be joined at their zero point, extending in opposite directions, and allows this zero point to be contextually determined by a contextually relevant class (C) (cf. Figure 9), , they may behave like the rest of the relative GAs with respect to the entailment patterns, but may also have a minimum standard absolute interpretations, given that the use of a term such as *lazy*, *pretty*, or *industrious* presupposes divergence from the zero value and therefore the existence of that property (and not its antonym).

[48] Evaluative adjectives also seem to correspond to Rusiecki (1989)'s *unary*-scale adjectives.

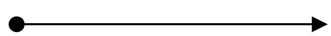
Figure 9: Evaluative adjective scales (Bierwisch, 1989)



3.5.2.2. Absolute GAs and Closed Scales

Absolute GAs differ from relative GAs in scalar structure by the presence of one or more scalar endpoints, which signal the absence versus the presence of a particular property.[49] As discussed in the previous section, there are two varieties absolute GAs – maximum and minimum standard. In each of the following pairs, the maximal member signals the absence of the property denoted by the minimal member: *clean*/*dirty*, *dry*/*wet*, *healthy*/*sick*, *safe*/*dangerous* (i.e., *clean* signals the absence of dirt(iness), *dry* signals the absence of moisture, *healthy* the absence of disease, and *safe* the absence of danger). We can capture this pattern with the scale in Figure 10. This type of scalar structure is proposed by Kennedy & McNally (2005) and Kennedy (2007). It corresponds to Rusiecki (1985)'s asymmetric bounded scale and, roughly, to Rotstein & Winter (2004)'s scale for total and partial predicates.

Figure 10: Absolute GAs (scale closed on one end)



The endpoint marks a zero value, indicating the absence of the property in question, while the presence of the property occupies the rest of the scale in increasing degrees. The opposite end remains open for the same reasons the relative scale is open on this end: there is no maximal

[49] *Absolute GAs* correspond to Cruse (1986)'s *gradable complementaries*, to Yoon (1996)'s and Rotstein & Winter (2004)'s *total*/*partial* predicates, and Kamp & Roßdeutscher (1994)'s *existential*/*universal* characterization, which they used discussing the German verb *heilen*, 'to cure/recover'.

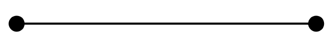
value, no degree that cannot be excelled (e.g., it is not possible to be maximally wet, sick, dangerous, dirty, etc.).

We can now see how the type of scale structure proposed in Figure 10 easily accounts for the entailment patterns discussed earlier. Beginning with comparatives such as *NP_1 is Aer than NP_2*, if A is *dirty*, then it should be clear that both entities should be dirty (or at a minimum that NP_1 is), since their degrees occupy the part of the scale that extends rightward into increasing degrees of dirtiness. If A is *clean*, then either NP_1 has no degree of dirtiness and is oriented at the endpoint, while NP_2 has some degree of dirtiness, or else NP_1 has some degree of dirtiness, but a lesser degree than NP_2 and is therefore oriented closer to the endpoint than NP_2. Likewise, to say that something is *clean* must entail that it is not *dirty*, because *clean* denotes the absence of dirtiness, and to say that something is *dirty* must entail that it is not *clean*, because it occupies a segment of the scale that signals the presence of dirtiness. Negating these sentences yields similar conclusions.

The fact that absolute GA scales are closed with an endpoint is particularly important, because this element that licenses modification by adverbs such as *slightly* and *almost*. Recall that *almost* can modify maximum standard absolute GAs (total predicates) but not minimum standard absolute GAs (partial predicates). This makes sense if *almost* indicates proximity to a maximal value or endpoint (cf. Amaral, 2006). By contrast, *slightly* can modify minimum standard absolute GAs (total predicates) but not maximum standard absolute GAs (partial predicates). Again, this makes sense if *slightly* indicates minimal degree of a property, and minimum standard absolute GAs occupy the portion of the scale signaling the same. Something that is *slightly A* will minimally extend away from the endpoint.

Note, however, that not all absolute GA pairs are like *clean*/*dirty*, *dry*/*wet*, *healthy*/*sick*, etc., with one member of the pair signaling complete absence of a property and the other presence to any degree, excluding a maximal degree. In absolute GA pairs such as *full*/*empty*, *open*/*closed* (at least on one interpretation), and *opaque*/*transparent*, the members of the pair indicate a range from complete presence to complete absence of a property. For example, a container that is *full* is completely filled, while one that is *empty* is not at all (or, stated another way, a container that is *full* has no degree of emptiness, while one that is *empty* is completely *empty*). Absolute GA pairs such as these are therefore assigned the scalar structure presented in Figure 11, where there is both a minimal value (a 0 point) and a maximal value. This type of scalar structure is proposed by Kennedy & McNally (2005) and Kennedy (2007), and corresponds to Rusiecki (1985)'s symmetric bounded scale.

Figure 11: Absolute GAs (scale closed on both ends)



We have seen that assigning different scalar structures to relative and absolute GAs accounts for the entailment patterns observed in § 3.5.1, and specifically, that the presence of one or more endpoints on these scales plays a crucial role. In fact, these differences in scalar structure also impact the standard of comparison for these GAs. This point is fleshed out in more detail in the following section.

3.5.3. Standard of Comparison

In § 3.3.1, I discussed the role of the comparison class in the positive form of GAs. In that section, I discussed the comparison class as if it were always contextually determined. In fact, having knowledge of different classes of GAs and differences among scalar structures in our arsenal, we now see that this characterization glossed over an important difference between two classes of adjectives – one in which the standard of comparison is contextually determined, and one in which it is determined independent of contextual factors by minimal or maximal values.

Kennedy & McNally (2005) explicitly linked scalar structure and the standard of comparison by assuming that GAs' lexical representations encode ensuring that their standards are fixed appropriately in the positive form. In the case of relative GAs, which map onto an open-ended scale, the context determines the standard of comparison in the manner discussed earlier. These predicates are therefore inherently vague. In the case of absolute GAs, the endpoint(s) of the scale plays a key role, providing the cutoff point that serves as the standard of comparison. Because assigning a truth value to a proposition expressed in a sentence is based on determining whether the property does or does not hold of an object – a question of the property's presence or absence – these predicates are not vague.

Following Kennedy (2007), I adopt the representations in (57) to distinguish between the three different kinds of GAs. Here, as in (32), g is the gradable adjective (a measure function), s is a context-sensitive measure function that returns a comparison class, min is the minimum scalar value, and max is the maximum scalar value.

57. a. relative GA $\lambda x.g(x) \succeq \mathbf{s}(g)$

b. minimum standard absolute GA $g_{min}(x) \succ \mathbf{min}(g_{min})$

c. maximum standard absolute GA $g_{max}(x) = \mathbf{max}(g_{max})$

(57a) indicates that for a relative GA such as *tall* to hold of an individual, the individual's degree must be greater than or equal to the contextually-relevant standard for *tall* (i.e., $tall(x) \succeq s(tall)$), while (57b) indicates that for a minimum standard absolute GA such as *spotted* to hold of an individual, the minimal degree (a zero value) must be exceeded (i.e., $spotted(x) \succ min(spotted)$), and (57c) indicates that for a maximum standard absolute GA such as *full* to hold of an individual, the maximal degree to be met (i.e., $full(x) = max(full)$). The difference among the three standards – relative, maximum standard, and minimum standard – is captured by basing the standard of comparison either on a contextually-relevant comparison class or the endpoint of a closed scale.[51]

3.6. Conclusion

I began this chapter by explaining that the degree argument encoded in the lexical representations of GAs allows them to appear in a variety of constructions which involve the comparison between a reference and a standard value. I next discussed important features of the positive (non-comparative) form of GAs, specifically their implicit comparative nature and

[51] There are cases in which imprecision in usage pulls the maximum standard off its perch, so to speak, and the GA takes on a relative reading (cf. Bolinger, 1967b; Kennedy & McNally, 2005; Sapir, 1944). (See also Bartsch and Vennemann (1972a) for related discussion.) This point becomes important in examining the experimental results reported in the following chapter.

vagueness. I presented a possible representation of a null morpheme, POS, which has been proposed for the positive form before discussing how degrees map onto scales, identifying basic scalar features, and presenting a series of entailment patterns, which led us to conclude that crucial differences underlie different classes of GAs. I then showed that differences in scalar structure – specifically, the presence or absence of one or more endpoints – can account for differences among relative GAs and maximum and minimum standard absolute GAs. I concluded this discussion by describing the important connection between the scalar structure of GAs and the standard of comparison.

We are now in a better position to understand the issue presented at the end of Chapter 2. Showing that young children can shift the standard of comparison and take contextual information into account when assigning an interpretation to a gradable adjective only demonstrates that they have one type of lexical representation – the representation for a relative GA, which maps onto an open scale and relies on the context to specify the standard of comparison. In order to demonstrate that children fully understand the all types of GAs, we must be able to also demonstrate that they make use of endpoint-oriented standards, and therefore distinguish between relative and absolute GAs. The experiments presented in Chapter 4 fill this gap.

The features of scales presented in this chapter also allow us to ask interesting questions about the process by which children acquire knowledge about scales: how do children learn about the directionality of scales, what allows them to deduce the relevant dimension, does their knowledge of comparisons within the classes of adjectives differ, what role do entailments play in their developing knowledge of GA meanings, and so on. In Chapters 5 and 6, I focus on one particular question that lies at the heart of GA representations: how do children learn about relative-absolute scalar differences? That is, what enables them to learn that some scales are closed, while others are not? In these chapters, I will propose that the same adverbs used to highlight differences among GAs in this chapter play an important role in this process.

Chapter 4: Children's Knowledge of the Semantics of Gradable Adjectives

4.1. Introduction

In Chapter 2, I began by introducing background research on children's initial acquisition of adjective meaning and then discussed their more advanced knowledge of the grammatical category of adjectives at preschool age, focusing specifically on gradable adjectives (GAs). GAs are a good case study for examining children's semantic knowledge (specifically with respect to the adjectival category) because some of the canonical GAs (e.g., *big*, *little*) are among the first adjectives produced by children (cf. Blackwell, 1998) and because with these lexical items, we see clear-cut differences among subclasses of GAs (e.g., entailment patterns, contextual variability) that arise from differences encoded in their semantic representations. In the previous chapter I outlined the principal semantic theoretic accounts of these distinctions: while relative GAs have a standard of comparison based on a contextually-determined comparison class, absolute GAs make reference to a standard of comparison based on the presence or absence of a relevant property. This difference in the standard of comparison encoded in the lexical representation is correlated with a difference in scalar structure. Relative GAs map individuals onto degrees on an open scale (i.e, with no endpoints), which allows the standard to shift when the context changes. Absolute GAs map them onto a closed scale. Maximum standard absolute GAs (also called total predicates) participate in comparisons in which the standard is oriented to an endpoint signaling the (maximal) absence of a property, while minimum standard absolute GAs (or partial predicates) diverge from this endpoint, indicating increasing degrees of the presence of a property.

Drawing out these differences among GAs allows us to build on the previous literature on children's understanding of adjectives presented in Chapter 2, and to ask whether children's semantic representations of GAs encode differences between standards of comparison and scalar structure. Two main questions highlight the crucial differences among GAs. First, do children distinguish GAs that depend on the context to set the standard of comparison (relative GAs) from those that do not (absolute GAs)? Second, within the class of absolute GAs, do they distinguish between those that have a minimum standard of comparison (e.g., *spotted*, *wet*) and those with a maximum standard of comparison (e.g., *full*, *dry*)?

Here, however, a third question arises, whose answer provides an additional way to distinguish the three GA subclasses and highlights an instance of the interaction between semantics and pragmatics in child language. Among the three subclasses of GAs are there differences in the role of context? That is, while relative GAs *rely* upon the context to identify the comparison class and set the standard of comparison, do children take aspects of the discourse context into account when making judgments about sentences in which absolute GAs appear?

In this chapter I present the results of two very different tasks that demonstrate that both adults and children as young as three years of age do indeed differentiate between three different kinds of GAs – relative GAs, maximum standard absolute GAs, and minimum standard absolute GAs.[52] To anticipate the results, we find that children establish the standard of comparison somewhere the middle of a series for relative GAs (e.g., *big*, *long*) and allow this standard to be

[52] This work is also presented with additional analysis of the results in Syrett, Kennedy, and Lidz (submitted). Various versions of this work were presented at GALANA 2004 (Syrett, Bradley, Lidz, & Kennedy, 2006), the 2005 LSA Annual Meeting (Syrett, Bradley, Lidz, & Kennedy, 2005), the 2006 CUNY Human Sentence Processing Conference (Syrett, Lidz, & Kennedy, 2006a), the 2006 Regional Meeting of the Chicago Linguistic Society (Syrett, Lidz, & Kennedy, 2006b), and the 2006 Chicago Workshop on Scalar Meaning (Kennedy, 2006).

shifted depending on the context. By contrast, the standard for minimum standard absolute GAs (e.g., *spotted*, *bumpy*) is invariantly oriented towards a minimal value or existence of a relevant property, while the standard for maximum standard absolute GAs (e.g., *full*, *straight*) is oriented towards a maximal value or absence of the relevant property. Furthermore, children allow a certain degree of flexibility regarding how precise this maximum standard must be, a pattern that helps to distinguish between the vagueness arising from the semantic representation of relative GAs and the pragmatic imprecision resulting from language usage in which the standard for maximum standard absolute GAs deviates minimally from the maximal value.

4.2. Pre-Experiment Scalar Judgment Task

4.2.1. Introduction

The purpose of the Scalar Judgment Task was to elicit judgments about properties of objects to determine whether children represent the three subclasses of GAs differently. This task also allowed us to establish children's judgments about the stimuli used in Experiments 1 and 2.

4.2.2. Method

4.2.2.1. Participants

Thirty-six children representing three age groups participated in this task: 12 three-year-olds (6 boys 6 girls, range: 3;3 to 3;11, M: 3;8); 12 four-year-olds (5 boys 7 girls, range: 4;1 to 4;11, M: 4;5); and 12 five-year-olds (5 boys 7 girls, range: 5;0 to 5;11, M: 5;5). In addition, 28 adult native speakers of English (Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course) served as controls.

4.2.2.2. Materials

The stimuli for this task consisted of four sets of seven items each, which are outlined in Table 1. Pictures of these stimuli are presented in Appendix A.

Table 1: Stimuli for Scalar Judgment Task

adjective	stimuli
relative	
big	7 wooden blocks painted blue, decreasing incrementally in size from 2^3″to $5/8^3$″
long	7 wooden rods painted green, 0.5″ in both width and height, decreasing incrementally in length from 8″ to 2″
absolute	
spotted	7 wooden disks painted red, all 3.75″ in diameter and 0.5″ thick, ranging from being covered with spots to having no spots
full	7 clear plastic containers with white lids, all 2″ in height and 1.5″ in diameter, and ranging from being full (of lentils) to empty

4.2.2.3. Procedure

Participants were presented with four sets of seven objects one after the other. For each item in each set, the experimenter asked the participant, "Is this A?" where A was a target adjective corresponding to the property exemplified by the objects in the set (*big*, *long*, *full*, *spotted*).[53] The experimenter always started at the positive, rather than the negative, pole (i.e., at the big end, not the small end). There were two conditions, based on the order of presentation of the sets. Half of the participants were randomly assigned to the 'Relative-Absolute' condition and saw the sets in the following order: *big*, *long*, *full*, *spotted* (i.e., relative GA, relative GA,

[53] We targeted the positive member of each antonymous pair in order to address the differences between the adjectives, and not get sidetracked by issues arising from asymmetry between the two poles, which has been the focus of a number of studies (cf. Bartlett, 1976; Brewer & Stone, 1975; Carey, 1978; Clark, 1972; Clark, 1973; Eilers, Kimbrough Oller, & Ellington, 1974; Keil & Carroll, 1980; Klatzky, Clark, Macken, 1973; Marschark, 1977; Townsend, 1976; *inter alia*).

absolute GA, absolute GA). The other half of the participants were randomly assigned to the 'Absolute-Relative' condition and saw the sets in the following order: *full*, *spotted*, *big*, *long*. The task took approximately 10 minutes.

4.2.3. Analysis

Results were analyzed in terms of the percentages of "yes" answers for each item in the series.

4.2.4. Results

The judgments from adults and children are presented in the following figures. The seven items are represented by numbers along the length of the x-axis (where the greatest degree = 1, and the least degree = 7) and the percentage of acceptance is on the y-axis.

Figure 12: Adults' scalar judgments

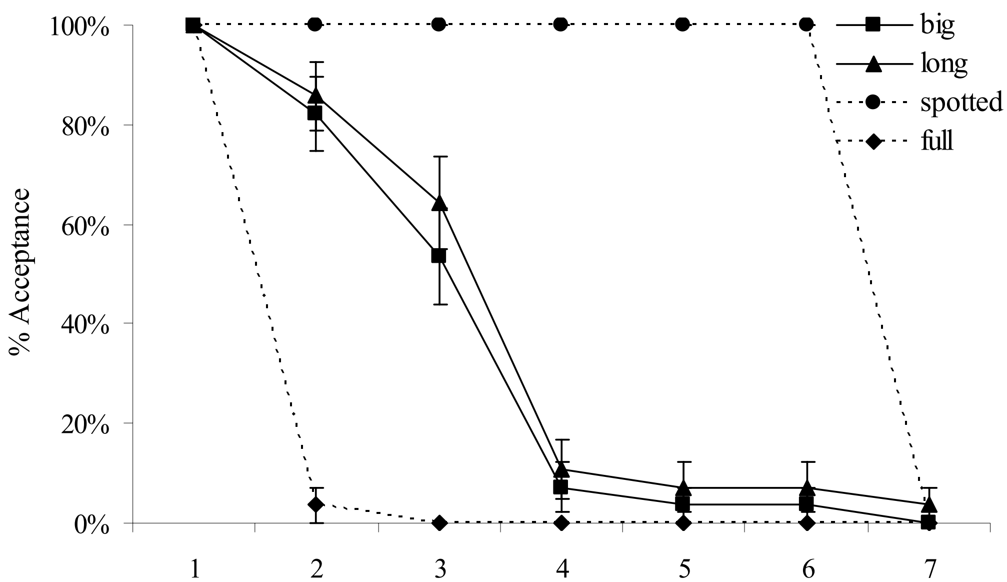
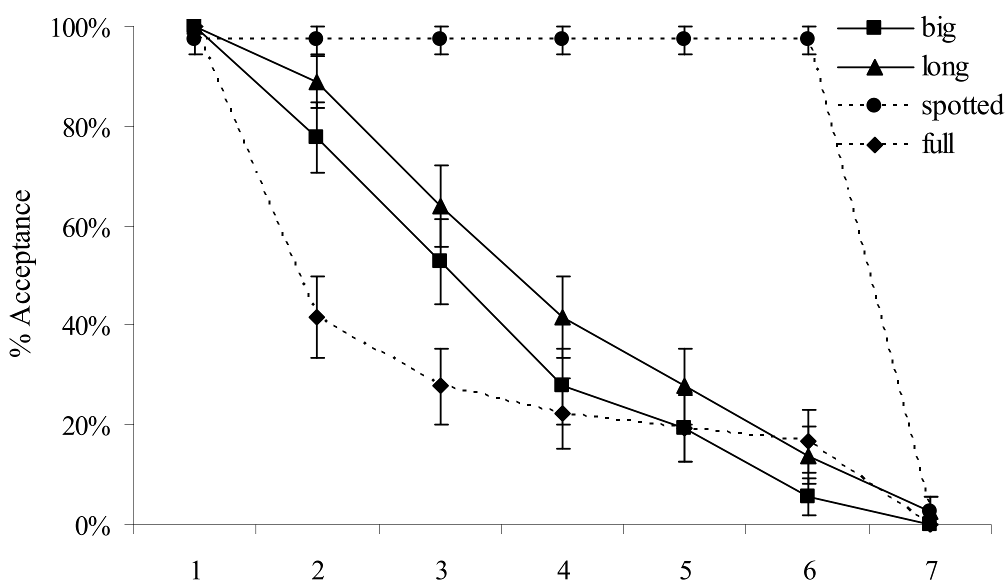


Figure 13: Children's scalar judgments



For both age groups the acceptance for both relative GAs decreased around the midpoint of the series.[54] Judgments for *spotted* were clear-cut for both children and adults. Both children and adults accepted as *spotted* any disk with any number of spots on it. Judgments for the maximum standard absolute GA *full* were noticeably different from the minimum standard absolute GA *spotted*. Adults only accepted item #1 as full. Children's judgments, however, were less sharp. In order to assess whether children were adult-like in their judgments of *full*, we conducted a comparison of *full* to the relative GAs *big* and *long*.

If children were adult-like in their judgments, they should have demonstrated a tendency to allow the second and third items to be *big* or *long* but not allow them to be *full*. We therefore targeted the second and third item in the series for statistical analysis. Even though the children were not at floor with *full*, as adults were, for both the second and third items, the difference between the *big* and *full* judgments and between the *long* and *full* judgments is significant in one-tailed t-tests (item #2, *big*/*full*: $t(35) = -3.654$, $p = 0.0004$; *long*/*full*: $t(35) = -5.596$, $p = 0.000001$; item #3, *big*/*full*: $t(35) = -2.485$, $p = 0.009$; *long*/*full*: $t(35) = -4.448$, $p = 0.00004$).[55]

There was also an effect of condition ('Relative-Absolute' v. 'Absolute-Relative'). Eleven of the fifteen children (73.3%) who judged a container other than the first one to be *full* were in the 'Relative-Absolute' condition, and saw the *big* and *long* sets before the *full* and *spotted* sets. These children were scattered across age groups (three 3-year-olds, four 4-year-olds, and four 5-year-olds). Put another way, eleven of the eighteen children (or 61.1%) in the

---

[54] Although judgments for the first and last items in the *big* and *long* sets were at 0% and 100%, this does not mean that their scales are closed at the endpoints. Rather, the percentage should be interpreted as the probability of being judged as *big* or *long* given a finite set of entities, which ranges from 0 to 1.0.

[55] We predicted that judgments for *full* and *big* would diverge, and therefore ran one-tailed t-tests.

‘Relative-Absolute’ condition judged the second container to be *full*, while only four of the eighteen children (or 22.2%) in the ‘Absolute-Relative’ condition did. This difference in condition for the *full* items that is captured by focusing on the second item is significant in a two-tailed t-test ($t(17) = 2.36$, $p = 0.03$). Six children across age groups seemed to interpret *full* as ‘filled to some degree’ and judged the first six items in the set to be *full*. Excluding these children from the condition analysis, we observe the same asymmetry: seven children in the ‘Relative-Absolute’ condition, and only two in the ‘Absolute-Relative’ condition judged the second item to be full. A two-factor repeated measure ANOVA (*big*/*full*, Relative-Absolute/Absolute-Relative) revealed a between-subjects effect ($F(1, 35) = 12.37$, $p = 0.001$), indicating a difference in responses for the two adjectives, a marginal within-subjects effect ($F(1, 36) = 3.24$, $p = 0.08$), and a marginal interaction ($F(1, 34) = 3.24$, $p = 0.08$), both driven by the differences in responses to *full*, since in both orders for *big*, 78% of the responses indicated the second object was *big*.

4.2.5. Discussion

There are three main results for the Scalar Judgment Task. First, we have evidence that for both children and adults there is a difference among the three subclasses of GAs, which is manifested in judgments along a scale. The fact that relative GAs such as *big* and *long* have a context-dependent standard (whereas absolute GAs do not) is reflected in judgments that divide the scale somewhere around the contextual midpoint for the former, but not for the latter.[56] Likewise, the fact that the standard for minimum standard absolute GAs such as *spotted* is

[56] A standard need not be assigned exactly at a midpoint of a range, but the mean or median seem to be reasonable targets for the standard. In our stimuli we predicted that participants would target the midpoint, since the relative GA items represented a one-step linear continuum.

oriented towards a minimal value, while the standard for maximum standard absolute GAs such as *full* is oriented towards a maximal value, is reflected in judgments that separate the minimal item from the rest of the series for *spotted* and the maximal item from the rest of the series for *full*. Second, judgments of relative size along a scale appear to become more categorical with development. This trend is evident in the comparison of the percentages of the third and fourth items of the relative GA sets for the two age groups. The judgments are therefore similar in form to those reported by Smith, Cooney, and McCord (1986), who found that adults had broader categories with steeper slopes than children did for objects described as *big* or *long*. Finally, most children, excluding a small subset who interpreted *full* as ‘filled to some degree’, share with adults the meaning of *full* as ‘maximally filled’. This finding was supported by a comparison of judgments for the *full* set with the two relative GA sets. However, this interpretation of *full* can be influenced by prior context, as seen in the ordering effect between conditions. This malleability of *full* judgments based on the context previews two results that emerge in the next set of experiments, one in which we see again that the order of presentation has an effect on *full* judgments, and another where we see that children allow some imprecision in the degree to which this maximal value must hold, evidenced in their reaction times.

4.3. Experiment 1a

4.3.1. Introduction

The Scalar Judgment Task provided evidence of differences among the three subclasses of GAs with respect to how the standard of comparison is set, indicating that for relative GAs, the standard is dependent on the context and will be oriented somewhere towards an average or midpoint, while for absolute GAs the standard will be oriented towards a maximal or minimal value. If this is where the difference between these adjectives lies, then we should see the standard of comparison continually shift from context to context for relative GAs, while it will always be endpoint-oriented for the absolute GAs. The goal of this experiment is to use a pragmatic task to highlight the differences among the three subclasses of GAs based on how the standard of comparison is computed. More specifically, we were interested in determining whether children would correctly shift the standard of comparison for relative GAs, but correctly resist doing so for absolute GAs.

4.3.2. Method

4.3.2.1. Participants

Thirty children representing three age groups participated in Experiment 1a: 10 three-year-olds (5 boys 5 girls, range: 3;5 to 3;11, M: 3;8); 10 four-year-olds (4 boys 6 girls, range: 4;1 to 4;11, M: 4;5); and 10 five-year-olds (3 boys 7 girls, range: 5;1 to 5;8, M: 5;5). In addition, 24 adult native speakers of English (Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course) served as controls.

4.3.2.2. Materials

The materials for Experiment 1a and 1b were the same, consisting of a series of pairs of objects, each sharing a salient dimension (e.g., color, shape, length, etc.). The experiment was divided into a training session and test session. Each session included both kinds of target stimuli, corresponding to the target relative and absolute GAs, and control stimuli. The complete set of materials is presented in the Appendix (B-D).

4.3.2.3. Procedure

Participants were invited to play a game. Children were introduced to a puppet (played by a second experimenter) and were told that that the purpose of the game was to help the puppet learn how to ask for things. They were then told that they would be shown two objects at a time, and that every time they saw two objects, the puppet would ask for something. Their job was to determine if they could give the puppet what he asked for based on his request, and if they couldn't, to tell him why not. Even the youngest participants followed these directions easily. Adult participants interacted with one adult experimenter instead of the puppet.

There was an important twist in this task that distinguished it from previous forced-choice studies in which a child was asked to act on a request with an adjective such as *big*.[57] In this task, the request was not always felicitous. This pragmatic manipulation was accomplished in the following way. Each request included a singular definite NP (e.g., *Please give me the A one*, where A was an adjective such as *red*). This type of description has received a great deal of attention in the semantics, pragmatics, and philosophy literature since at least the early part of the twentieth century (cf. Heim, 1990; Kadmon, 1990; Neale, 1990; Roberts, 2003; Russell, 1905; Strawson, 1950). It is generally accepted that a singular definite NP of this form presupposes both existence of such an object (e.g., there must *be* a red one) and uniqueness (e.g., there must be *only one* red one). Thus in order to use such a description felicitously, a speaker should be committed to these presuppositions. In this task, in determining whether or not they could give the puppet what he asked for, children were, in essence, assessing the context with respect to the presuppositions of the singular definite NP. For this reason, we refer to this task as the *Presupposition Assessment Task* (PAT).

Pairs of objects were designed so that they either satisfied or violated one or both of the presuppositions of the definite description. For some pairs, the request (e.g., *Give me the red*

---

[57] Cf. Bartlett, 1976; Ebeling & Gelman, 1988; Eilers, Kimbrough Oller, & Ellington, 1974; Gelman & Ebeling, 1989; Gelman & Markman, 1985; Harris, Morris, & Terwogt 1986; Ravn & Gelman, 1984; and Sena & Smith, 1990.

one) was felicitous, because exactly one object fit the description (e.g., there was a red object and a white object); for other pairs, the request was infelicitous, because both members of the pair fit the description (e.g., there were two red objects); and for still other pairs, the request was infelicitous, because neither member of the pair fit the description (e.g., there was a yellow object and a blue object). For our target pairs, whether or not the request was felicitous depended on how the adjective was interpreted. This aspect warrants further explanation.

Let's begin with an example using the positive form relative GA *big* and two pairs of blocks from the Scalar Judgment Task. The puppet's request for both pairs is *Please give me the big one*. Now, when both blocks come from the positive end of the series, one might expect there to be a violation of the presupposition of uniqueness, as with the two red objects in the above example. Likewise, when they both come from the negative pole, one might expect there to be a violation of the existence presupposition, as with the yellow and blue objects in the above example. However, recall that as a relative GA, *big* depends on the context to fix the standard of comparison: what is considered to be *big* depends on what the salient cutoff is. We can think of each of these pairs as the relevant context. Each time a new pair is presented, the child and puppet are presented with a new context, and as long as the members of the pair differ in size, there will always be a way of fixing the cutoff so that it lies between the two, thereby 'shifting' the standard of comparison each time, and accommodating the presuppositions of the request (cf. Lewis, 1968). Thus, regardless of whether a block was judged to be *big* or *not big* in the Scalar Judgment Task, it has the potential to be treated as 'the big one' in Presupposition Assessment Task, as long as it is the bigger member of the pair.

If children select the bigger of the two blocks (or the longer of two rods) each time a request is made, this would certainly appear to constitute evidence that they know that such GAs are sensitive to the context and that they are willing to accommodate the presuppositions of the singular definite NP – and therefore that they have the correct semantic representation for these GAs. However, there is an alternate explanation consistent with this pattern. It may be that children simply understand the request for 'the A one' as a request for 'the more A one'. On the surface, this appears to be the same sort of explanation offered above: given the meaning of relative GAs, when there are two different-size blocks, the bigger block can be labeled as 'the big one'. However, if children really do understand this phrase as an implicit comparison (for example, because they think the standard of comparison for all GAs is context dependent), they should also understand requests for 'the full one' and 'the spotted one' as implicit comparisons and should always select the object possessing the greater degree of the relevant property (i.e., 'the fuller one' and 'the more spotted one').

Including absolute GAs in the positive form therefore serves as a control for the relative GAs and helps to bolster our conclusion about children's adjectival representations. If, for children, *spotted* means 'has some degree of spots', they should reject a request for 'the spotted one' when presented two spotted objects – even if one is clearly more spotted than the other, because this request involves a violation of the presupposition of uniqueness. Likewise, if *full* means 'filled to the maximal degree' or 'has no degree of emptiness', then participants should reject requests for 'the full one' when presented two non-full containers, because the request violates the presupposition of existence. If instead we observe children consistently giving the puppet the object with the greater degree of the property, there is also a second explanation: children have the correct representations for relative and absolute GAs, but this is masked by a general willingness to accommodate, resulting in them overriding their own judgments and consistently finding a way to satisfy the request. Looking ahead to the results, we find that participants consistently accept requests involving relative GAs but reject requests involving absolute GAs with the pairs described above, therefore providing us with evidence for the relative/absolute distinction within GAs.

To ensure that participants understood the rules of the game and to help them feel comfortable rejecting the request, we had them participate in a brief training session before the test session began. This training session was composed of two felicitous and two infelicitous requests (cf. Appendix B). Once it was evident that participants felt comfortable with the task, we proceeded with the test session. If children were still hesitant to correct the puppet after the four training items, we introduced a fifth impromptu pair accompanied by an infelicitous request.[58] Items in the test session included target items whose salient property corresponded to one of the target GAs (cf. Appendix C) and control items (cf. Appendix D). (Pictures of stimulus

---

[58] One question that might arise at this point is whether children will be able to succeed at this task, given the understanding of the definiteness that it requires. It is true that children have been reported to overproduce the definite determiner (cf. Karmiloff-Smith, 1979; Maratsos, 1976) and have apparent difficulty processing definite NPs in real-time (cf. Trueswell, Sekerina, Hill, & Logrip, 1999). However, neither of these studies necessarily predicts that children will have difficulty succeeding in the Presupposition Assessment Task. First, non-adult-like production has been known to mask adult-like comprehension in general (see Naigles, 2002 for a review). Second, analysis of the results from the online studies has centered on children's difficulty incorporating information from the discourse context to restrict the reference of the NP (cf. Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Meroni, 2006), not their inability to understand the presuppositions of definite NPs. Finally, there is evidence that for children as young as three, singular definite NPs do pick out a relevant unique individual, and plural definite NPs the relevant maximal set, but that the difficulty for children may lie in determining how to restrict the domain of reference (Munn, Miller, & Schmitt, 2006). Thus, we have every reason to think that children may be successful in this task. And in fact, as we will see, they are.

examples from Experiments 1 and 2 are presented in Appendix E.) The presentation of the members of the pairs was counterbalanced so that the object fitting the description appeared on different sides of the pairs throughout the test session. The order of the pairs was also pseudorandomized with respect to three factors: the felicity of the request, the nature of the presupposition violation, and the adjective. Subjects were randomly assigned to one of two orders, a point that becomes important in the interpretation of the results. Specific details about the request, stimuli, and correct responses for the target GA items are outlined in Appendix F.

4.3.3. Analysis

For each pair, we predicted children would respond in one of three possible ways to the accompanying request: they would accept the request by giving one object, reject the request and give neither object, or reject the request and give both objects. Each of these three types of responses could potentially to be correct depending on the pair, and predictions about each pair were clear, given the above discussion. We therefore chose to capture the results in terms of percentage of correct responses. This method is in contrast to capturing the results in terms of the percentage of responses in which the participant gave only one object (since such a percentage would not indicate *which* object was given) or the percentage of responses in which the participant only gave the object with the greater degree of the property (since in this case and the former, the percentage would not indicate what happened when the participant did not give one object, but gave both or neither).

4.3.4. Results

The results are presented in Table 2.

Table 2: Percentage of correct responses in Experiment 1a

	control items			test items				
age	color	shape	mood	*big*	*long*	*spotted*	*full*	(nf/nf, f/nf)
3 yrs	93	95	93	95	90	90	70	(40, 100)
4 yrs	98	100	100	95	95	85	65	(30, 100)
5 yrs	100	100	100	100	100	95	65	(30, 100)
adults	100	100	100	98	98	98	92	(88, 96)

These results show that both adults and children accepted and rejected the requests accordingly: both age groups were at ceiling with the control items and also responded correctly with the test items. Recall that the members of the target pairs also appeared in the series seen in the SJT. In the SJT, children consistently judged items #1 and 3 in the relative sets (the bigger/longer pair in this task) as *big* or *long*, and rejected this label for items #5 and 7 in the same series (the smaller/shorter pair in this task). In spite of these judgments, however, in the PAT, children selected only item #1 as 'the big one' or 'the long one' in that pair, and also allowed item #5 to be 'the big one' or the 'long one' compared to item #7. It appears, then, that children were willing and able to shift the standard of comparison for the relative GAs. That our participants did not do so for the other items indicates that context-dependence is part of the lexical entry for these adjectives only and not the others, and that their success with the relative GA stimuli was not due to their treating the request as involving an implicit comparison.

Although children's responses generally aligned with adults', there was one exception – the items corresponding to the GA *full*. Recall that there was a pair in which one of the containers was full and the other filled only to some degree. (We will call this pair the 'full/non-full' pair, abbreviated in Table 2 as 'f/nf'). For another pair, there were two containers, neither of which was full, but one was fuller than the other. (We will call this pair the 'non-full/non-full' pair, abbreviated in Table 2 as 'nf/nf'). Children's correct responses to the 'non-full/non-full' pair are significantly lower than to the other target pairs (two-tailed t-tests t(29), 'non-full/non-full' v. *big* #1/3: t = -7.62, p < 0.00001; *big* #5/7: t = -5.29, p = 0.00001; *long* #1/3: t = -5.46, p < 0.00001; *long* #5/7: t = -7.62, p < 0.00001; *spotted* #1/4: t = -4.07, p = 0.0003; *spotted* #5/7: t = -7.62, p < 0.00001; 'full/non-full': t = -7.62, p < 0.00001). When presented with the 'full/non-full' pair, children consistently gave the puppet the full container; however, when presented with the 'non-full/non-full' pair, many children did not reject the request. Instead, they gave the puppet the fuller of the two containers.

This type of response would make sense if children were interpreting *the A one* as an implicit comparison (e.g., *the more A one*). We know from their responses to the spotted objects, though, that they were not uniformly interpreting the request this way. Given two spotted objects, where one is more spotted than the other (e.g., items #1 and 4), children consistently rejected the puppet's request for 'the spotted one', commenting that both objects were spotted, and did not give the puppet the more spotted member of the pair. The fact that children's responses for *full* and *spotted* were so different suggests that this pattern of results has to do with the way children are interpreting *full* rather than a general strategy for reanalyzing these unmarked adjectives as comparatives in contexts that would otherwise result in presupposition failure.

This response to the 'non-full/non-full' pair is also curious in light of what we learned from the SJT, which showed that children share with adults a meaning of *full* that has a maximally-oriented standard of comparison. Revisiting children's responses in the SJT makes this point even clearer. 18 children who participated in the Presupposition Assessment Task (PAT) also participated in the SJT within an interval of approximately three weeks. Their judgments of the *full* series in the SJT set are presented in Table 3. Note that only two children thought a container beyond the first two (i.e., the third container) was *full*.

Table 3: Judgments of *full* from children participating in the SJT and the PAT

response in SJT	children giving each response	ages 3 yrs	 4 yrs	 5 yrs
only #1 is *full*	12 (66.7%)	5	4	3
#1 and 2 are *full*	4 (22.2%)		2	2
#1, 2 and 3 are *full*	2 (11.1%)	1	1	

Recall that the 'non-full/non-full' pair was composed of items #4 and 6 from the set of seven items. *None* of these eighteen children judged item #4 to be full in the SJT; however, in the PAT, *eleven* of these eighteen children (61.1%) gave this item to the puppet to satisfy his request for 'the full one'. (See Table 4.)

Table 4: Children in the SJT and the PAT who gave the puppet item #4

response in SJT	children who gave item #4 as 'the full one'	ages 3 yrs	 4 yrs	 5 yrs
only #1 is *full*	8 (72.7%)	3	3	2
#1 and 2 are *full*	2 (18.2%)		1	1
#1, 2 and 3 are *full*	1 (9.1%)	1		

Given this pattern, we decided to probe children's responses to the 'non-full/non-full' pair further, asking if the condition to which participants were assigned had an effect. Indeed, upon further examination, we found that eight of the 11 children (72.7%) who gave the puppet the fuller container while having judged it to be *not full* in the SJT were in the condition in which this 'non-full/non-full' pair appeared early in the sequence. Moreover, every single one of the 15 children from Experiment 1a who were in this condition gave the puppet the fuller of the two non-full containers in response to his request for 'the full one' . Only five of the 15 children in the condition in which this pair appeared later in the sequence (after the 'full/non-full' pair) responded to the puppet's request in this way (a difference between 100% and 33.3% for the two conditions).

This difference between the two orders leaves open two possibilities. The first is that in setting the maximum standard, children benefit from exposure to the object instantiating the maximal value (e.g., seeing the full container). The second is that the presence of a relative pair earlier in the sequence simply influenced the children's responses, misleading them to treat *full* as relative rather than as absolute. In the order in which the 'non-full/non-full' pair appeared before the 'full/non-full' pair, the pair was also immediately preceded by a *long* pair. It is possible that in shifting the standard of comparison with the *long* pair, children were then influenced to respond in a similar way to the 'non-full/non-full' pair that followed. This possibility is explored in Experiment 1b.

4.3.5. Discussion

The results of Experiment 1a demonstrate that both adults and children distinguish between three subclasses of GAs. They shift the standard of comparison for relative GAs, but do not do so for absolute GAs. Relative GAs such as *big* and *long* have as part of their semantic representation a contextually-determined standard of comparison, and so allow for the standard to shift from context to context. Accordingly, when presented with a request for 'the big one' or 'the long one', participants willingly gave the bigger or longer object in the pair. In contrast, when presented with two spotted objects, participants did not give the more spotted object. This is to be expected, since *spotted* is an absolute GA with a standard that is oriented towards a minimal value and does not make reference to a cutoff point supplied by the context at hand. Adult participants similarly rejected requests for 'the full one' when given two non-full containers. However, some children gave the puppet the fuller container in this situation. This pattern was in contrast to our findings from the SJT, in which we found that children did not actually think that a nearly-full container was *full*. One possible explanation for this pattern of results is that a relative pair appearing earlier in the sequence influenced children's responses to the 'non-full/non-full' pair. This possibility is explored in Experiment 1b.

4.4. Experiment 1b

4.4.1. Introduction

The goal of Experiment 1b was to determine the source of the ordering effect observed in Experiment 1a. Specifically, we wanted to know if the prior presentation of a relative pair in the sequence of items influenced children to erroneously treat *full* as context-dependent, causing them to give the puppet the fuller of the two non-full containers in response to his request for 'the full one'.

4.4.2. Method

4.4.2.1. Participants

Seventeen children representing three age groups also participated in this task: 6 three-year-olds (3 boys 3 girls, range: 3;1 to 3;11, M: 3;5); 6 four-year-olds (2 boys 4 girls, range: 4;2 to 4;11, M: 4;6); and 5 five-year-olds (1 boy 4 girls, range: 5;2 to 5;10, M: 5;4). In addition, 10 adult native speakers of English (Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course) served as controls.

4.4.2.2. Materials

The same objects from Experiment 1a were used. The only difference was in the sequence of items. In Experiment 1a, the 'non-full/non-full' pair was almost immediately preceded by a *long* pair, with only one control pair intervening. To evaluate the influence of the relative GA pair, we made a minor change in the order of presentation, simply switching the order of the *long* and 'non-full/non-full' pair so that the latter was no longer preceded by the former. This change allowed us to determine whether participants would still give the fuller container when shown this pair, even when it was not preceded by a relative pair, thereby demonstrating that participants' willingness to treat a non-full container as 'the full one' was not due to them being cued to 'think relatively'.

4.4.2.3. Procedure

The procedure was the same as in Experiment 1a.

4.4.3. Results

Two analyses of the data were conducted. The first is an analysis of the percentage of correct responses, as in Experiment 1a. The second is an analysis of children's reaction times across conditions in Experiment 1 (the reasons for which are outlined in § 4.4.3.2.)

4.4.3.1. Percentages

The results for Experiment 1b are presented in Table 5. The pattern of responses is similar to the one observed in Experiment 1a[59]: children and adults responded correctly to the training, control, and test items, with the exception of the 'non-full/non-full' pair.

---

[59] The lower rate of correct responses to the *spotted* stimuli by the three-year-olds may be explained by the fact that in Experiment 1b, the experimenter encouraged to give the puppet verbal feedback, but, unlike in Experiment 1a, did not encourage the child to give the puppet both objects when both fit the description. This unintended change was due to a change in the experimenters administering the task.

Table 5: Percentage of correct responses in Experiment 1b

	control items			test items				
age	color	shape	mood	*big*	*long*	*spotted*	*full*	(nf/nf, f/nf)
3 yrs	83	100	83	100	83	67	50	(0, 100)
4 yrs	96	100	100	92	83	92	58	(17, 100)
5 yrs	95	100	100	100	90	90	70	(40, 100)
adults	100	100	100	100	100	95	85	(70, 100)

Recall that the purpose of conducting Experiment 1b was to identify the source of children's non-adult-like responses to the 'non-full/non-full' pair. The order of stimuli in Experiment 1a, where a *long* pair almost immediately preceded the 'non-full/non-full' pair left open the possibility that this relative pair influenced children to pattern as they did with the latter pair. Here, a direct comparison between participants' responses to this pair in both conditions of Experiment 1 is informative. As we observe in Table 6, both children and adults were at their best when this pair appeared later in the sequence.[60]

Table 6: Percentage of correct responses to 'non-full/non-full' pair

age	pair early, after relative pair (Experiment 1a)	pair early, before relative pair (Experiment 1b)	pair later, after f/nf pair (Experiments 1a, 1b)
children	0	18	67
adults	75	70	100

It appears that the prior presentation of a maximal standard (i.e., seeing the 'full/non-full pair' earlier in the sequence) allowed participants to better succeed in their judgments of the 'non-full/non-full' pair. Children were not misled by the prior presentation of a relative GA pair to treat the fuller container as 'the full one'. These conclusions are supported by the fact that,

[60] Adults who gave the fuller of the two non-full containers noted at the end of the experimental session without any prompting that they realized this mistake later in the experiment and wished to make clear to the experimenter that they knew what *full* means.

while the difference between either of the two ‘early’ orders and the ‘later’ order is significant, the difference between the two ‘early’ orders is not (Fisher’s Exact Tests, two-tailed probabilities: early 1a v. later: $p < 0.001$; early 1b v. later: $p = 0.01$; early 1a v. early 1b: $p = 0.24$). However, the fact that children benefited from exposure to the maximal standard does not fully explain the results from Experiment 1. Even when given this boost, children’s responses to the ‘non-full/non-full’ pair still deviated from their adult-like responses to the other pairs. Instead of rejecting the puppet’s request, as they did when there were two spotted objects and the puppet asked for ‘the spotted one’, they often accepted his request, giving him the fuller container. In order to explore this difference further we therefore turned to an additional analysis, in which we evaluated children’s reaction times during the experimental session.

4.4.3.2. Reaction Times

Our comparison of the two experimental orders in Experiment 1 shed light on children’s non-adult-like performance with the ‘non-full/non-full’ pair. To get a more fine-grained picture of this pattern, we turned to an analysis of their reaction times (RTs) during the experimental sessions. The purpose of the RT analysis was to determine if children who gave the puppet the fuller of the two non-full containers in response to his request for ‘the full one’ took longer to do so for this pair than for other pairs. We reasoned that children might have evaluated the puppet’s request given the experimental stimuli at hand and subsequently decided to allow the fuller container to be treated as ‘the full one’. Thus, although children might not have judged the fuller container to be *full* in the abstract, they might have taken some time to decide that it was close enough to the maximal standard to *count as* full given the discourse context, and subsequently decided to treat this container as ‘the full one’ to satisfy the speaker’s request.

Analyzing children's RTs therefore bears on the question of whether children actually do have the correct representations for the maximum standard absolute GA *full*, and in the experimental session, are influenced by the immediate context. This analysis also bears on the question of the role context plays for the different subclasses of GAs. By analyzing children's RT to key pairs in the experimental session, we can determine whether a longer RT is exhibited whenever children have to accommodate a request (e.g., when they are shown two little blocks and asked for 'the big one') or if it is a more specific behavioral response associated with deviations from a maximal standard. Thus, the RT analysis is a way of evaluating the difference between the vagueness associated with the semantic representations of relative GAs, which depend on the context for the standard, and the imprecision associated with the pragmatics of maximum standard absolute GAs, where language usage allows for minor deviations in the stringency of the standard.

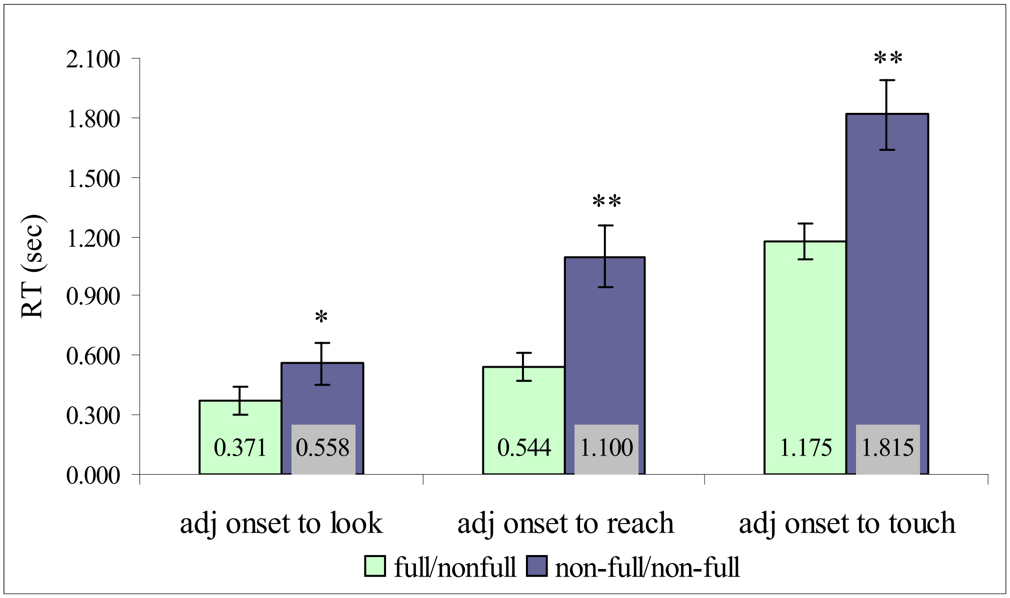
Experimental sessions with child participants were videotaped using a Sony Digital8 Handycam. Videotapes were imported from the camera into Apple Inc's iMovie program as .mov files. Videos were then coded offline by research assistants in our laboratory on a Macintosh computer using SuperCoder software (Hollich, 2003). For each item in which the child accepted the puppet's request (i.e., gave the puppet one of the two objects), the research assistants coded three measurements: the child's look to the object, the child's reach toward the object, and the child's touch of the object. These RT measurements were coded frame by frame, where one frame is equal to 1/30 of a second.

It is necessary to provide some additional details about the coding session. First, we excluded from analysis any items in which the children's eye movements could not be coded (e.g., if the eyes were occluded), the child was already looking at or touching the stimuli before the request was uttered, or any other experimental artifact prevented the coders from obtaining measurements (i.e., there was a distraction in the background). For this reason, the total number of children whose RTs were analyzed varies from analysis to analysis. This number is always provided in a footnote. Second, rather than coding the initial look to the object, since the child could have decided to inspect the second object before deciding to give the puppet the first one, we coded the look that immediately preceded the reach to the object. A reach was a movement that ultimately resulted in touching an object. We chose to target these measurements instead of the proportion of looking time, since we wished to measure latency of response across key items. Finally, at least two coders were assigned to each experimental session, with one coder arbitrarily chosen as the default. In case there was a disagreement of more than 5 frames for any of the three measurements, a third coder was brought in as a tiebreaker for that item. There was generally high agreement among coders.

We targeted two RT measurements for analysis. In the first RT measurement, we asked if children took longer to respond to the puppet's request for the 'non-full/non-full' pair than with the 'full/non-full' pair. These results are presented in Figure 14.[61]

[61] We analyzed the RTs for sixteen children who gave the puppet the fuller container for each pair.

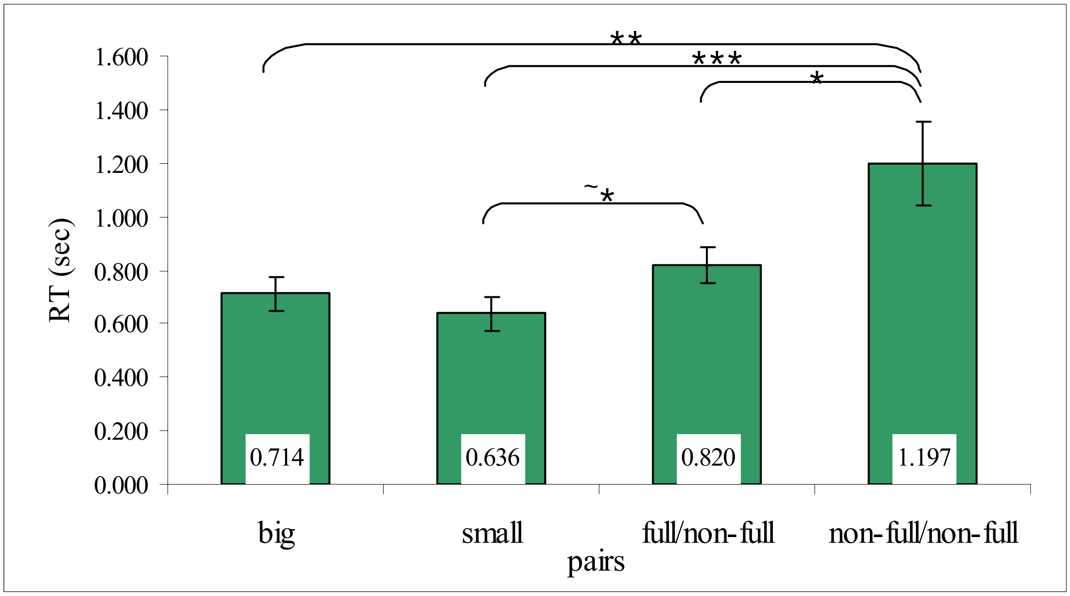
Figure 14: Reaction times for two *full* pairs in Experiment 1a



Indeed, differences between the look, reach, and touch are significant for these two pairs (one-tailed t-tests (look $t(15) = 1.71$, $p = 0.05$; reach $t(15) = 3.03$, $p = 0.004$; touch $t(15) = 3.47$, $p < 0.002$). This difference indicates that the computation associated with deciding to give the puppet the fuller container in response to his request for 'the full one' is different than the one associated with the decision to give him the container that is actually full and is indeed 'the full one'. That children take longer with this latter pair appears to indicate that something extra is involved in their decision. To determine the nature of this 'something extra', we performed a second analysis.

In the second RT measurement, we compared the RTs for the 'non-full/non-full' pair to those for other key pairs in order to determine if children generally took extra time to shift the standard of comparison, or if the extra time observed with the 'non-full/non-full' pair was unique to that pair. We targeted the difference between the look and the touch of the object for four key pairs, the two *big* pairs (two big blocks and two small blocks) and the two *full* pairs. These results are presented in Figure 15.[62]

Figure 15: Reaction times for four key pairs in Experiment 1a



The RTs for the 'non-full/non-full' pair clearly stand out from the three others. Indeed, while the RT for this pair is significantly longer than every other pair (post-hoc two-tailed t-tests: 'non-full/non-full' v. 'full/non-full': $t(48) = 2.42$, $p = 0.02$; 'non-full/non-full' v. big: $t(45) = 3.07$, $p = 0.004$; 'non-full/non-full' v. small: $t(51) = 3.79$, $p = 0.0004$), the other pairs do not differ significantly from each other (two-tailed t-tests: 'full/non-full' v. big: $t(53) = 1.11$, $p = 0.27$; big v. small: $t(56) = 0.87$, $p = 0.39$; 'full/non-full' v. small: $t(59) = 1.96$, $p = 0.05$, marginally significant). These results demonstrate that the shift in the standard of comparison for the *big* pairs – even for the pair in which both blocks were judged to be small in the series –

---

[62] We analyzed all codable responses for these items across all children. The number of children varied for each item: 26 for the big blocks, 32 for the small blocks, 29 for the 'full/non-full' pair, and 21 for the 'non-full/non-full' pair.

was automatic.[63] This is expected: given that a contextually-determined standard is part of the semantic representation for *big*, no additional time should be required to compute the standard of comparison. It is part of their meaning that they rely on the context, so accommodation of the request should not have presented any added difficulty. By contrast, GAs such as *full* do not rely on the context in this way, since their standard is maximally oriented. Thus accommodation of the request should not be automatic. That children took longer for the *full* than the *big* pairs is an indication that the role of the context differs between these two pairs, and consequently, that their semantic representations are different.

4.4.4. Discussion

The results of Experiment 1b elucidate those of Experiment 1a by demonstrating that the source of children's non-adult-like performance with the 'non-full/non-full' pair (i.e., giving the puppet the fuller container in response to his request for 'the full one') is not driven by the influence of a preceding relative GA pair. Children are significantly more likely to pattern like adults with the 'non-full/non-full' pair if it occurs later in the sequence, some time after they have seen the maximal standard exemplified, but even with this assistance, some children are still inclined to allow the fuller container to count as 'the full one'. The RT analysis demonstrated that children take significantly longer to give the puppet the fuller of the two non-full containers (and that their response time does not increase when giving the puppet the bigger

[63] One might have expected a semantic congruity effect with the small pair of blocks (i.e., longer RTs resulting from the fact that the adjective in the request was *big*, but the blocks were judged to be not big). The lack of a semantic congruity effect here may be due to the fact that the difference in magnitude between the two pairs was not extraordinarily large.

of two small blocks), perhaps indicating that they are able to suspend their own judgments of fullness and allow the fuller container to be treated as ‘the full one’ in that context.

What we seem to see here (children being more likely to give the puppet the non-full container when it appears earlier in the series, and when taking longer to treat a container that is nearly full ‘the full one’ than they do to identify a container that is actually full or for the bigger of two small blocks) is pragmatic reasoning that arises not from the lexical representation of these GAs, but from language usage, which permits some variation in how stringent the maximality requirement must apply. If children see the full container early in the series, they know that the puppet has also seen it, and therefore that he has seen what the maximal standard is. Given this order, they are less likely to relax their standard and allow the fuller non-full container to be treated as ‘the full one’. Their decision to give this container in response to the puppet’s request, whatever the order of stimuli, consistently takes additional time, indicating that this behavior does not arise from what is encoded in the semantic representation of the term but from how they see this term being used in the discourse context.

4.5. Experiment 2a

4.5.1. Introduction

Children’s responses to the *full* pairs in Experiment 1 immediately prompts an important question: are these responses unique to *full*, or can they be generalized to a larger set of maximum standard absolute GAs? Given the striking divergence in children’s responses to the *full* stimuli, we would want to know if this behavior is indicative of treatment of this term in particular or of terms that share a similar semantic representation (e.g., maximum standard absolute GAs, which have a standard that is oriented towards a maximal value, but that may not be required to hold strictly in certain contexts). The goal of Experiment 2 was to address this question.

4.5.2. Method

4.5.2.1. Participants

Thirty children representing three age groups participated in this task: 10 three-year-olds (5 boys 5 girls, range: 3;2 to 3;11, M: 3;6); 10 four-year-olds (5 boys 5 girls, range: 4;1 to 4;10, M: 4;4); and 10 five-year-olds (4 boys 6 girls, range: 5;0 to 5;7, M: 5;3). In addition, 24 adult native speakers of English (Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course) served as controls.

4.5.2.2. Materials

The materials were the same as in Experiment 1, with the exception of four target pairs. In place of the two *full* pairs, there were two pairs corresponding to another maximum standard absolute GA, *straight*. In place of the two *spotted* pairs, there were two pairs corresponding to another minimum standard absolute GA, *bumpy*. These pairs were designed similarly to those in Experiment 1, so that only one of the two pairs for each adjective would satisfy the presuppositions of the definite description. Details for these stimuli are presented in Table 7.

Table 7: Stimuli substituted for the absolute GA pairs from Experiment 1

adjective	pragmatic status of request	stimuli
straight (in place of *full*)	felicitous	2 wire rods, 9″ in length, one completely straight and the other bent into a "C" 
	infelicitous	2 wire rods, 9″ in length, one curly and the other even curlier 
bumpy (in place of *spotted*)	felicitous	two flat wooden boards, measuring 2″ x 5″ and painted orange, one with a few bumps and one with no bumps
	infelicitous	two flat wooden boards, measuring 2″ x 5″ and painted orange, one with many bumps and one with fewer bumps

4.5.2.3. Procedure

The procedure was the same as in the previous experiments.

4.5.3. Results

As in Experiment 1, children and adults responded correctly to all items with the exception of the maximum standard absolute GA stimuli. The same difference observed in Experiment 1 was found between the two *straight* pairs (abbreviated in the table as 'b/b' for 'bent/bent' and 's/b' for 'straight/bent'). The percentages of correct responses are presented in Table 8.

Table 8: Percentage of correct responses in Experiment 2a

	control items			test items				
age	color	shape	mood	*big*	*long*	*bumpy*	*straight*	(b/b, s/b)
3 yrs	83	90	93	95	95	85	55	(70, 100)
4 yrs	95	100	97	100	95	90	85	(80, 100)
5 yrs	95	100	97	100	95	90	95	(80, 100)
adults	100	100	100	100	100	100	94	(100, 100)

The same ordering effect observed in Experiment 1 was also found. Of the seven children across age groups who gave the straighter rod for the 'bent/bent' pair, six (or 85.7%) of the seven saw this pair early in the sequence, well before they were exposed to the 'straight/bent' pair.

Now, while the percentage of correct responses to the 'bent/bent' pair was in the same direction as the 'non-full/non-full' pair, the percentage of correct responses to this pair was not nearly as low as in Experiment 1. In Experiment 1a, only 33% of the children responded correctly to the 'non-full/non-full' pair and rejected the request, and in Experiment 1b, only 19% of the children did so. By contrast, in this experiment, 77% of the children rejected the request. However, participants' responses to this pair are still significantly lower than to most (but not all) of the other test pairs (two-tailed t-tests, $t(29)$: 'bent/bent' v. *big* #1/3: $t = -2.97$, $p = 0.006$; *big* #5/7: $t = -2.26$, $p = 0.03$; *long* #1/3: $t = -2.97$, $p = 0.006$; *bumpy* #5/7: $t = -2.97$, $p = 0.006$; 'straight/bent': $t = -2.97$, $p = 0.006$; but v. *bumpy* #1/4: $t = 1.99$, $p < 0.06$, marginally significant; and v. *long* #5/7: $t = -1.28$, $p = 0.211$, n.s.).

4.5.4. Discussion

The results of Experiment 2a support a generalization of the results from Experiment 1b to a larger set of absolute GAs, since participants' responses to *straight* and *bumpy* were similar to their responses for *full* and *spotted*, respectively. Both age groups rejected the infelicitous requests for the items corresponding to these GAs and did not shift standards as they did for the relative GA pairs. We observed, however, that children were more likely to reject the infelicitous request for the 'bent/bent' pair than they were for the 'non-full/non-full' pair. It is possible, though, that the higher percentage of correct responses for the 'bent/bent' pair was an effect of the stimulus design. We observed *post facto* that the objects for *straight* shown in Table 7 seem to correspond better to stimuli for a minimum standard absolute GA such as *bent* or *curly* than for the maximum standard absolute GA *straight*: while item #1 was completely straight, none of the other wires had any real degree of straightness to them. This design differed from the *full* materials, where it was the case that while only item #1 was full, the others in the PAT manifested some degree of being filled. This observation led us to conduct Experiment 2b using a minimally revised set of stimuli.

4.6. Experiment 2b

4.6.1. Introduction

The goal of this experiment was to introduce a slightly revised set of *straight* materials that more closely resembled the *full* materials used in Experiment 1 in order to determine whether participants' judgments about 'the straight one' when shown the 'bent/bent' pair would be more similar to those in Experiment 1a and 1b than they had been in Experiment 2a.

4.6.2. Method

4.6.2.1. Participants

Thirty children representing three age groups participated in this task: 10 three-year-olds (3 boys 7 girls, range: 3;2 to 3;10, M: 3;6); 10 four-year-olds (4 boys 6 girls, range: 4;1 to 4;9, M: 4;7); and 10 five-year-olds (5 boys 5 girls, range: 5;1 to 5;11, M: 5;6). In addition, 24 adult native speakers of English (Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course) served as controls.

4.6.2.2. Materials

The materials were the same as in Experiment 2a, with the exception of the *straight* pairs. In place of the three rods that were bent to varying degrees, we inserted rods that resembled more accurately the containers used for *full*. For example, a completely straight rod was paired with a rod that was straight for most of its length but which had a curl at the top, analogous to a container which is filled most, but not all, of the way. The new pairs are presented in Table 9.

Table 9: Stimuli substituted for the *straight* pairs from Experiment 2a

adjective	pragmatic status of request	stimuli
straight	felicitous	a completely straight rod and a rod with a long straight section and a curly section at the top 
	infelicitous	a rod with a long straight section and a curly section at the top and a totally curly rod 

4.6.2.3. Procedure

The procedure was the same as in the previous experiments.

4.6.3. Results

4.6.3.1. Percentages

In Experiment 2b, we observed the same trend as in the previous experiments, with children and adults responding correctly to all of the training, control, and test items, with the single exception of the 'bent/bent' stimuli. These results are presented in Table 10.

Table 10: Percentage of correct responses in Experiment 2b

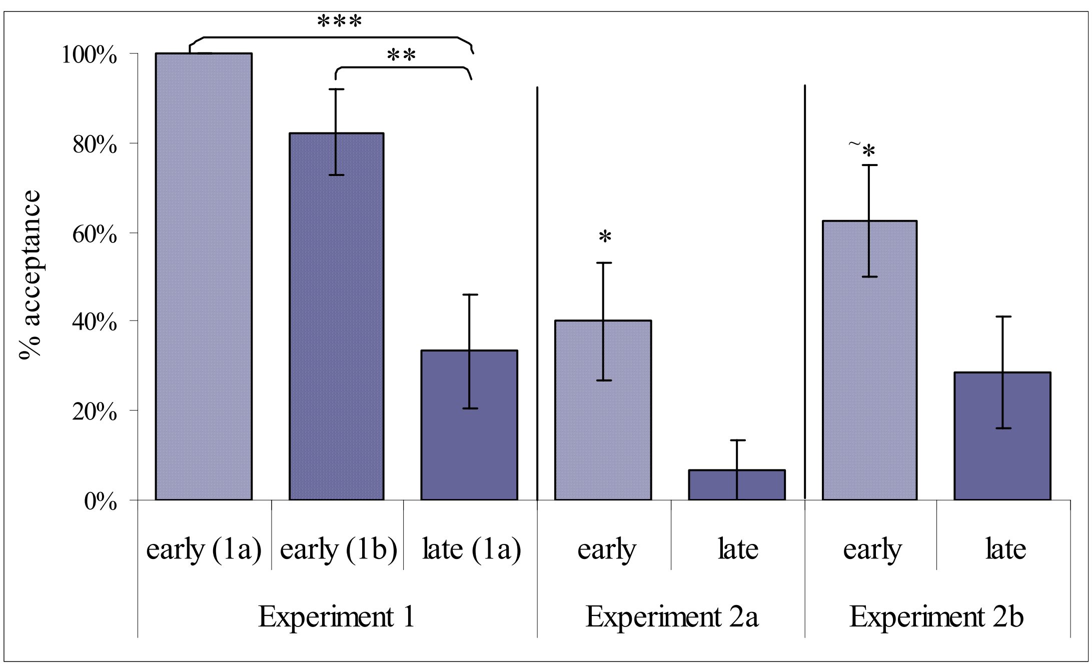
	control			test				
age	color	shape	mood	*big*	*long*	*bumpy*	*straight*	(b/b, s/b)
3 yrs	88	80	87	95	90	60	70	(30, 90)
4 yrs	90	90	97	95	90	75	80	(60, 90)
5 yrs	100	100	100	100	100	80	90	(70, 90)
adults	100	100	100	100	100	94	98	(88, 100)

This time, children's responses to the 'bent/bent' pair are more like those seen in Experiment 1 and are significantly lower than all of the other test pairs (two-tailed t-tests, t(29): 'bent/bent' v.

big #1/3: t = -5.04, p < 0.0001; *big* #5/7: t = -4.47, p = 0.0001; *long* #1/3: t = -5.39, p < 0.00001; *long* #5/7: t = -3.27, p = 0.003; *bumpy* #1/4: t = -1.88, p < 0.07; *bumpy* #5/7: t = -4.10, p = 0.0003; 'straight/bent': t = -3.53, p = 0.001).

This time, 53% of the children rejected the request for 'the straight one' when presented with the 'bent/bent' pair (compared to the 77% with the previous stimuli in Experiment 2a and 33% for the similarly-designed 'non-full/non-full' pair with the same order in Experiment 1a). Again, there was an effect of the order of presentation, with children more likely to reject the request if they saw the 'bent/bent' pair later in the sequence. Nine of the 13 children (69%) who accepted the request and gave the puppet the straighter of the two bent rods saw this pair early in the sequence. The comparison presented in Figure 16 highlights the consistency of this ordering effect across Experiments 1 and 2.

Figure 16: Rate of acceptance of the non-maximal member in Experiments 1 and 2



This between-experiment comparison allows us to draw the following conclusion: when the pair with a member exemplifying the maximal standard appears earlier in the sequence than the pair without a maximal member, children are significantly more likely to correctly reject the puppet's request for the maximal object when they are shown the non-maximal pair later in the series. A series of single-factor ANOVAs for each experiment support this conclusion (Experiment 1: $F(2, 44) = 3.21$, $p < 0.0001$; Experiment 2a: $F(1, 28) = 4.20$, $p = 0.03$; Experiment 2b: $F(1, 28) = 4.20$, $p = 0.07$, marginally significant). (See also the statistical analysis following Table 6 for differences between the two early orders and the late order in Experiment 1.)

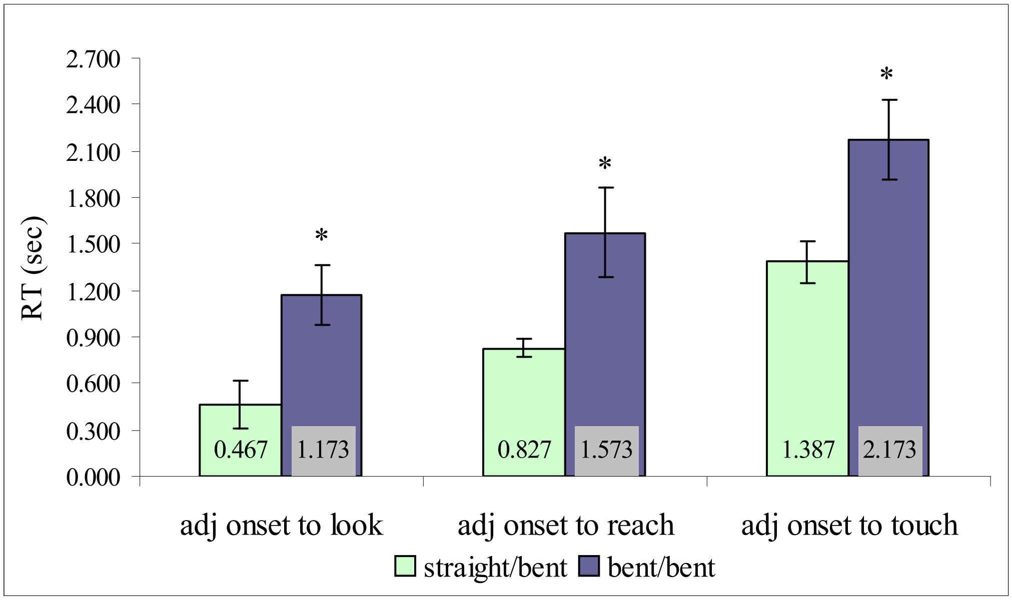
4.6.3.2. Reaction Times

In Experiment 1 we conducted an analysis of children's RTs as they responded to the puppet's request. This analysis provided us with evidence that children took longer to accept the fuller container in the 'non-full/non-full' pair than for both the full container in the 'full/non-full' pair and the bigger block in both *big* pairs. We took this increased RT to indicate that children decided to suspend their own judgments of fullness and permit the fuller container to be treated as 'the full one' in the experimental context. Although the number of subjects for which we could conduct this analysis for the *straight* stimuli was decreased,[64] we observed the same trend in RTs. There were only five children for which the comparison between the RTs for the 'straight/bent' and 'bent/bent' pairs could be made, but the differences between the adjective

[64] We targeted the responses from Experiment 2b only and did not combine responses across both experiments, as we did in Experiment 1, given that the *straight* stimuli differed in each experiment. In coding this experiment, we were also subject to the same constraints with respect to which videos we could code, as in Experiment 1. The total number of codable sessions and items was therefore less than for Experiment 1.

onset to the look, reach, and touch were all significant (one-tailed t-tests: look t(4) = -3.23, p = 0.016; reach t(4) = -3.15, p = 0.017; touch t(4) = -3.32, p = 0.015). See Figure 17.

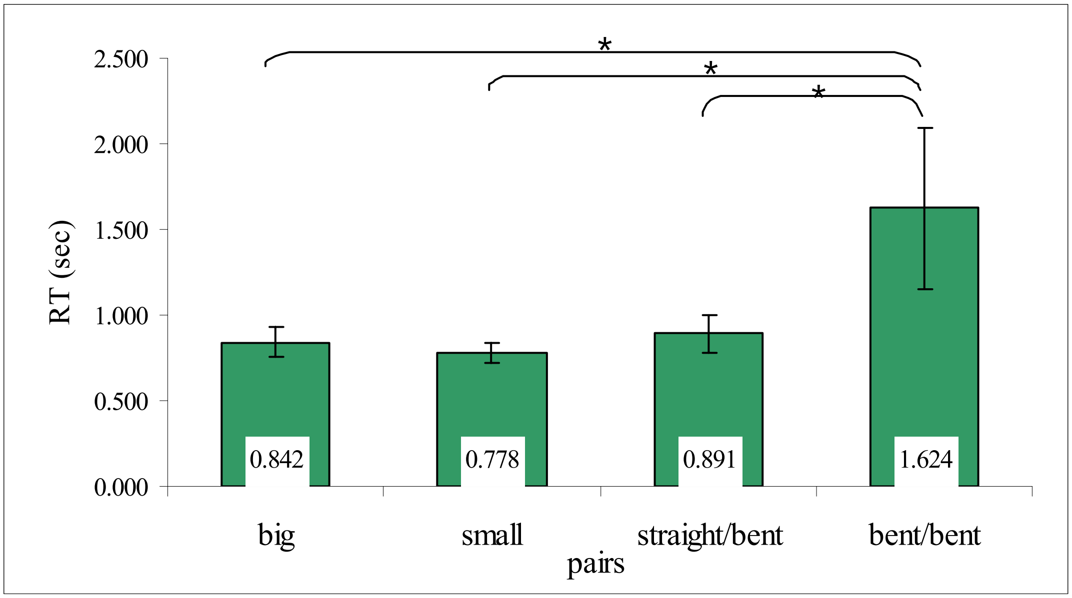
Figure 17: Reaction times for two *straight* pairs in Experiment 2b



As in Experiment 1, we also compared the difference between the look and the touch RTs for the 'bent/bent' pair to three other key pairs, the two pairs of blocks corresponding to the request for 'the big one' and the 'straight/bent' pair. We observed the same trend as we did in Experiment 1 (cf. Figure 18). Children took significantly longer to select the straighter of the two bent rods than they did for the other three pairs (two-tailed t-tests: 'bent/bent' v. 'straight/bent: t(23) = 2.15, p = 0.04; 'bent/bent' v. big: t(20) = 2.31, p = 0.03; 'bent/bent' v. small: t(22) = 2.75, p = 0.01). And, as before, the other pairs do not differ significantly from each other (two-tailed t-tests: 'straight/bent' v. big: t(31) = -0.33, p = 0.74; big v. small: t(30) = 0.64, p = 0.53; 'straight/bent' v. small: t(33) = 0.86, p = 0.40).[65]

[65] The data from the following number of children were analyzed for each pair: 15 for the big blocks, 17 for the small blocks, 18 for 'straight/bent', and 7 for 'bent/bent'.

Figure 18: Reaction times for four key pairs in Experiment 2b



4.6.4. Discussion

The results of Experiment 2 support a generalization of the results from Experiment 1 to a wider range of absolute GAs. The effect of the order of presentation and the longer RTs for the non-maximal pair are not unique to the stimuli included in Experiment 1, and therefore not unique to the lexical item *full*. Rather, they reflect a more general processing strategy on the part of the children, one in which children incorporate the discourse context into their interpretation of definite descriptions containing GAs. With the relative GAs, children consistently and automatically shift the standard of comparison. With the absolute GAs, they do not. When the presupposition of existence fails to hold with a maximum standard absolute GA, children are willing to accommodate, if the context allows it, permitting an item which they know does not meet the maximal standard to be a proxy for one that does.

4.7. General Discussion

In this chapter, we saw evidence from two very different tasks that adults and children as young as three years of age discriminate between three subclasses of GAs. In judgments along a scale, they set the standards differently for relative GAs, maximum standard absolute GAs, and minimum standard absolute GAs. Relative GAs have a standard oriented somewhere in the mid-range of the contextually relevant series. Maximum standard absolute GAs have a standard signaling the absence of a property, so only a container that has no degree of emptiness can be judged *full*. Minimum standard absolute GA have a standard signaling the presence of a property, so any disk that has spots is judged *spotted*. The responses from the scalar judgment task served two functions. First, they provided us with participants' judgments about the stimuli used in the PAT experiments, and second, they provided us with promising results concerning the differences between these adjectives, which led us to probe participants' treatment of them further.

In a set of experiments using the Presupposition Assessment Task, we showed that children consistently and easily shift the standard of comparison with relative GAs. For example, when a relative GA such as *big* is embedded in a definite description in a request (e.g., *Please give me the big one*), participants recognize that this NP picks out a unique discourse entity, and accordingly accommodate the presuppositions of the request and adjust the cutoff with the context so that, for example, the bigger member of a pair is treated as 'the big one' in that context. We saw that participants were not simply interpreting the positive form of the adjectival modifier in this NP as an implicit comparison (e.g., *bigger*), since they did not respond similarly with the minimum standard absolute GAs *spotted* or *bumpy*: with these items, they rejected the request when both items had the property, since this situation is a violation of the presupposition of uniqueness.

We found interesting results with the maximum standard absolute GAs *full* and *straight*. Although adults consistently rejected requests for e.g., 'the full one' when there were two containers only partially filled, but one was fuller than the other, children often allowed the fuller container to serve as 'the full one'. In a further investigation of this response pattern, we found that children who saw the pair with the full container before the pair with no full container were more likely to reject the request for 'the full one' when presented with the latter pair than children who were assigned to the condition where the order of the experimental stimuli was reversed. We interpreted this result of by appealing to pragmatics. Children in this order benefited from prior exposure to the maximal standard. They were given a chance to see that what counted as *full* (for the speaker in that context) was indeed full, and so their choice of what to do with the later pair was clear. Children in the reverse order (where the pair without the full container came earlier in the sequence) did not get this information, and so took extra time to evaluate the speaker's request in the discourse context and allow the fuller container to count as full. An analysis of reaction times revealed that their acceptance of the request with this pair was longer than for other key pairs, indicating that whatever was going on with their responses had to do specifically with the maximal standard, and the observed deviation from it.

Because relative GAs rely upon the context to identify the comparison class and set the standard of comparison, and this is encoded in their semantic representations, accommodation comes naturally. By contrast, maximum standard absolute GAs have a standard that is invariably maximally oriented, but there is some 'wiggle room' in how stringently this standard must be observed. It appears to be the case that participants took longer with the *full* stimuli in question because they had to evaluate what counted as *full* for the speaker in the discourse context. In other words, given two objects, the question 'How big does it have to be to be 'the big one'?' can always be answered with 'bigger than the cutoff separating the two objects by size', while the question 'How full does it have to be to be 'the full one'?' when there is no full one requires having the maximal standard in mind, and determining whether the fuller container is close enough to the maximal standard to be treated as such *and* whether this degree of fullness is acceptable given the speaker's belief and the discourse context. There is therefore a clear difference between the semantics and pragmatics here. The type of standard of comparison in question is associated with the semantics; the decision to allow a non-maximal amount to proxy for the maximal standard, given the context at hand is pragmatic.

Taken together, these results demonstrate that for children as young as three years of age, adjectives that allow for comparison of degrees along a scale are not treated as one indistinguishable class. Rather, there are nuanced distinctions among GAs based on how the standard of comparison is computed. That children responded in an experimental setting in a way that reflects these distinctions indicates that their responses must be motivated by semantic representations that encode these distinctions – semantic representations that should resemble those presented in Chapter 3. How is it that children could have such sophisticated knowledge of the semantics of these GAs by age three? The results presented in the next two chapters are attempts to identify one possible means: adverbial modification in the input serves as a cue to differences in semantic representations among GAs.

Chapter 5: Adverbial Modification Is a Cue to Scalar Structure

5.1. Introduction

In the previous chapter I reviewed experimental findings that show that by age three, children demonstrate an understanding of a distinction between the positive forms of gradable adjectives (GAs) such as *big* and *long*, which have a contextually determined standard of comparison, and those whose standard is context-independent.[66] Moreover, we saw that their understanding of the latter relies on a distinction between GAs whose standard is fixed at a minimum value (e.g., *spotted*, *bumpy*) and GAs whose standard is fixed at a maximum value (e.g., *full*, *straight*). Thus while what is considered to be *big* varies according to the objects being compared (i.e., the comparison class, or object kind), an object can be *spotted* as long as it has some amount of spots on it and can only be *full* if the contents have reached a maximal degree of fullness.[67] We concluded from the results that children's behavioral responses in these experiments must have been guided by semantic representations that distinguish between these adjectival classes. Specifically, we argued that their semantic representations of these adjectives are correlated with a difference in scalar structure, as outlined in (58).

---

[66] When I make reference to the differences among GAs, I am referring to differences among the positive form of these adjectives. See my discussion of this terminology in Chapter 3.

[67] In Chapter 4, I mentioned that while GAs such as *full* are oriented towards a maximal endpoint, which is not context *dependent*, the semantic representation does permit some 'wiggle room' around the maximum value, thereby allowing for the possibility that context plays a role in setting the standard of comparison. Thus, it is not anomalous to think of a glass as full when there is still some room for liquid, or an auditorium as full when a small number of seats remain vacant.

58. Scalar structure of GAs

a. Relative GAs: open scale, contextually-determined standard shifts

b. Maximum standard absolute GAs: maximally closed scale, standard oriented at maximal end

c. Minimum standard absolute GAs: minimally closed scale, standard oriented at minimal end

The implication, then, is that by age three, children share with adults abstract representations of the positive form of GAs that both incorporate a standard of comparison and allow for variation with respect to how this standard is set. Let's allow for the possibility that the language learner does not require positive evidence to posit the existence of a standard and furthermore, that the range of representational variation is not learned. Even so, the language learner would presumably have to learn which type of scalar structure to assign to any given GA. How could they arrive at this knowledge?

The idea I pursue in this chapter is that information that could help the language learner classify GAs according to scalar structure is available in the input. To be more explicit, I claim that regularities detectable at the surface level may serve as a reliable cue to deeper semantic differences, and that a learner could exploit such systematic differences in developing an adult state of semantic competency. I will spell out the exact details of such a learning account in Chapter 7. Before doing so, it is incumbent upon me to identify possible candidates for analysis and to demonstrate (a) that there is indeed a correlation between the distributional patterns seen on the surface and the abstract representations encoded in the grammar, and (b) that their occurrence is predictable enough to be informative. That is the task I undertake in this chapter.

The structure of this chapter is as follows. In § 5.2, I discuss two promising avenues for this learning account (resultative constructions and adverbial modification) and target the latter for investigation. In § 5.3, I present an analysis of transcripts of child-directed speech in which it appears that adverbs are few and far between in speech to children. However, I argue that learning does not occur from observation of this one source of input alone, and that a move towards a broader sample of speech is justified. In § 5.4, I present three sets of analyses of a different corpus, the BNC, in which we see statistically significant patterns of adverbial modification of GA classes. We therefore have evidence that if the language learner attends to both the speech directed to them and the ambient speech, then they are receiving evidence that GAs can be partitioned into different subclasses according to the type of adverb that modifies them.

5.2. Background

The first step is to identify surface features of the input that the language learner could exploit to learn about the scalar structure of GAs and then to determine the regularity of such features. Recall that all GAs can felicitously appear in comparative constructions, as in (1).

59. a. x is bigger than y

    b. #x is deader than y

Thus, if a learner encounters an adjective in a comparative, then she has evidence that the adjective (and the corresponding property) is gradable. However, while this syntactic construction distinguishes gradable from non-gradable adjectives, what we want to know here is if there is any evidence that a learner can use such constructions to distinguish between the different subclasses of GAs. Two such ways in which the surface string can highlight the deeper semantic differences among adjectives have been discussed in the literature: resultatives and adverbial modification. In this section, I will discuss each in turn, and ultimately select the latter as the focus of my corpus analysis.

5.2.1. Semantic Restrictions on APs in English Resultatives

Resultatives are one type of syntactic construction in English whose lexical semantic restrictions provide a means of distinguishing between different classes of adjectives. These constructions, such as the examples in (60), describe an event with a result state.

60. a. Nate hammered the metal flat.

 b. The audience booed the performer off the stage.

The verb appearing in English resultatives has been the source of a great deal of discussion. It has alternately been identified as describing a change of state and therefore allowing specification of a goal or delimitation of the action (as in hammering the metal flat, breaking the vase open or to bits, freezing solid, and so forth) (cf. Levin & Rappaport Hovav, 1991; Rappaport Hovav & Levin, 2001; Tortora, 1998), measuring out events (Vanden Wyngaerd, 2001), or labeling a durative – rather than punctual – event (cf. Beavers, 2002, 2005). The AP (*flat* in (60a)) or PP (*off the stage* in (60b)) provides the endpoint/goal/result state for the event (cf. Krifka, 1998). Green (1972) observed that while some adjectives are permissible in the AP of resultative constructions, others are not (61).

61. Adjectives in resultative constructions

 a. He wiped it clean/dry/smooth.

 b. #He wiped it dirty/damp/wet/stained.

Synthesizing a sizable literature on (1) resultatives in English, (2) GAs, and (3) event structure,[68] Wechsler (2001) argued that there is a connection between Aktionsart (lexical aspect) of the verb and the adjective head of the AP, namely that durative events pair with GAs with a maximally closed scale, and punctual events with non-gradable adjectives. Wechsler found support for this pattern in an analysis of Boas (2000)'s extremely comprehensive set of data on resultatives.[69] While the head of the AP in resultatives appears to be restricted to (maximally) closed-scale GAs, open-scale GAs routinely appear in *make*-causatives (e.g., *make x big*). The difference between these two structures is that the causative verb in a resultative encodes the manner of the event, whereas *make* does not. This pattern is illustrated in Table 11 (from Wechsler (2001), featured in Wechsler (2005)).

---

[68] (1) Beavers, 2002, 2005; Boas, 2000; Goldberg, 1995; Goldberg & Jackendoff, 2004; Levin & Rappaport Hovav, 1991, Rappaport Hovav & Levin, 2001; *inter alia*; (2) Kennedy, 1999; Rotstein & Winter, 2004; Sapir, 1944; (3) Krifka (1998).

[69] Boas used sources such as the British National Corpus, COBUILD Bank of English, dictionaries, usenet groups, and various websites on the Internet.

Table 11: Distribution of open- and closed-scale adjectives with resultatives and *make*-causatives found in Boas (2000) (adapted from Wechsler (2005))

adjectives	resultatives	*make*-causatives
famous	0	37
fat	0	5
ill	0	65
sleepy	0	19
sore	1	11
tired	0	18
clean	102	6
dry	77	8
flat	34	1
full	35	1
open	395	1
shut	207	0

Wechsler (2005)'s explanation for this descriptive generalization is that the maximal endpoint of the adjective lends its endpoint to the event. Since non-gradable adjectives and GAs with open scales lack such a scalar endpoint, they lack the potential to map the event onto their scale and bound it with an endpoint. Thus, resultatives in English, by selecting maximum standard absolute GAs and *not* relative GAs as the AP head, could therefore be informative to the child learning about the semantics of adjectives: adjectives that appear in a resultative construction must share similar representations in that they must give the event an endpoint and therefore must have a scalar structure that allows for this possibility. (See also Vanden Wyngaerd (2001) for extensive discussion of the correlation in both Dutch and English between resultatives and the scalar structure of adjectives appearing in them.)

5.2.2. Semantic Restrictions on Adverbial Modification

Adverbial modification can also serve to distinguish between different classes of adjectives. As with resultatives, adverbs such as *half*, *mostly* or *most of the way*, which have been referred to as "totality modifiers" by Paradis (1997) and "proportional modifiers" by Kennedy and McNally (2005), are restricted in terms of the adjectives they modify. And, like resultatives, these adverbs select for adjectives that describe a property with a maximal (and here, necessarily also a minimal) value. Compare (62) to (63).

62. Closed scale (total) adjectives allow modification by proportional modifiers

a. The glass is half/mostly full. (Kennedy & McNally, 2005, (19a))

b. The figure was completely visible/invisible. (Kennedy & McNally, (28c))

c. It is almost dry. (Wechsler, 2005, (3a))

63. Open scale adjectives do not allow modification by proportional modifiers

a.??The rope is half/mostly long. (Kennedy & McNally, (20a))

b.??Her brother is completely tall/short. (Kennedy & McNally, (25a))

c.??It is almost long/short. (Cruse, 1980, in Rotstein & Winter, 2004, (4))

Kennedy and McNally (pg. 353) account for this difference by saying that "if the scale associated with an adjective accepting, for example, *half* lacked either a minimal or a maximal element, it would be impossible to calculate differences relative to minimal and maximum values on the scale and thus to identify the midway point required by the degree modifier." They therefore propose the following derivation for *half full* in (64), where *half* is a function that takes the adjective *full* as its argument, and S_f represents the scale associated with *full*.

64. ⟦half⟧ (⟦full⟧) (Kennedy & McNally, (22))

$$= \lambda x.\exists d[\text{diff}(\max(S_f))(d) = \text{diff}(d)(\min(S_f)) \wedge \text{full}(x) = d]$$

This derivation gives us a property that is true of an entity iff the difference between the maximal element of the scale and the degree of fullness, and the difference between the minimal element of the scale and the degree of fullness are equal (i.e., if the degree of fullness is at the midpoint of the scale). Thus, the modificational behavior of *half* (and other such adverbs) falls out directly from the restrictions it places on the type of scalar structure of the adjective it selects, and instances of *half* in the input therefore provide evidence regarding the subclasses of GAs.[70]

In a search of deverbal adjectives in the British National Corpus, Kennedy and McNally found support for the claim that this kind of class membership can be signaled by adverbial modification. The adverbs *well*, *very*, and *much* show a strikingly different distribution in the adjectives they modify. See Table 12.

---

[70] Bolinger (1967b) notes that adverbs such as *thoroughly*, *utterly*, *completely*, and *fully* are semantically related in that they refer to "an ultimate state" (ftnt. 6, pg. 9). In terms of scalar structure, this makes sense, since these adverbs make reference to the maximal degree.

Table 12: Distribution of deverbal adjectives with three different modifiers in the BNC (adapted from Kennedy & McNally (2005, Table 1))

deverbal adjectives	*well*	*very*	*much*
acquainted	56	0	0
protected	58	2	0
documented	213	0	1
educated	78	3	0
needed	2	0	211
criticised	0	0	19
praised	1	0	17
appreciated	12	0	124
surprised	0	151	1
worried	0	192	0
frightened	0	92	0
interested	0	335	10

Faced with such a pattern, they argue that *well* combines with adjectives with totally closed scales, *very* with relative GAs, which allow the standard to be raised, and *much* with minimum-standard absolute GAs. Thus, an adjective's ability to be modified by an adverb may be an indication of how it is classified as a GA, and therefore what its semantic representation is.

While an investigation of either resultatives or adverbial modification appearing in natural language would potentially prove fruitful for determining how differences detectable in the input may be informative to the language learner about underlying differences in semantic representations, I am focusing on adverbial modification for four main reasons. First, and perhaps most importantly, analyses of resultatives that have found a correlation between the semantics of the construction and the scalar structure of the gradable adjective head of the AP have been in English. In order for an account of language learning based on surface patterns cueing deeper representational differences to really have power, one should be able to state broad universal generalizations that apply to language learners in any context. Unfortunately, the ability to do this for resultatives is questionable. It is not clear that the above-stated generalization about the semantic representation of the adjective appearing in resultatives holds cross-linguistically.

Williams, whose (2005) dissertation focused on resultatives (his 'complex causatives') appearing in English, Mandarin, and Igbo, observed variation among the three languages with respect to the acceptability of certain adjectives (cf. § 1.5.3 of his dissertation). According to Williams, both Mandarin and Igbo have constructions similar to the English resultative. The Mandarin example is in (65) and the Igbo in (66).[71] In both of these examples, the causative verb encodes manner and the adjective appears to provide a delimitation on the event.

65. tā lī duàn -le nàtiáo mùbăn. (Williams, 2005, (2))

3s kick snap -PFV that plank

'S/he made that plank snap by kicking [it].'

66. Ọ kụ wa –ra ọba ahụ. (Williams, 2005, (3))

3sS strike split –FACT gourd that

'S/he made that gourd split by striking [it].' (Williams, 2005, (3), from Hale, Ihiọnụ, & Manfredi, 1995)

Importantly, while English places restrictions on the adjective – namely that it must have a maximally closed scale – neither of these other two languages appears to have the same constraint. As we observe in (67a) and (68a), both languages have complex causative constructions that, when translated into their English resultative equivalent, would be barred (cf.

[71] Examples from Mandarin and Igbo were taken directly from Williams's (2005) dissertation. All decisions concerning orthography, glosses, and translations are his.

(67b) and (68b)). By contrast, the English *make*-causative construction permits the appearance of these adjectives, as seen in the glosses. This is to be expected, given that we saw in Table 11 that make-causatives permit a wider range of adjectives.

67. Mandarin compared to English

a. tā liǎng bǎ wǒmen dōu chǎo fán -le.

3s two BA 1p all quarrel annoyed -LE

'Those two made us all annoyed by quarreling.'

(Williams 2005, (79), from Wang 1995, (148))

b. #Those two quarreled us all angry. (Williams 2005, (80))

68. Igbo compared to English

a. Obi gba ru –ru miri

Obi move go bad –FACT water

'Obi made the water dirty by stirring.' (Williams 2005, (83))

b. ??Al stirred the water bad/dirty. (Williams 2005, (84))

Thus, if the complex causatives seen in Mandarin and Igbo are indeed the equivalent to the English resultative and are indicative of cross-linguistic variation in this construction, then the semantic restrictions on the adjective are not universal and are therefore not an unambiguous cue for the language learner. Importantly, if there are properties intrinsic to the language learner guiding her to attend to patterns in the input to map new adjectives onto scalar structure, this construction appears to be an unlikely source of information guiding the learner in the initial stages.[72] By contrast, adverbial modification seems to be a likely candidate.

---

[72] Note, however, that Lee and Naigles (2005), relating their work on verb learning in Mandarin Chinese to similar work in English, leave open the possibility that form-meaning correspondences come in two varieties – one which is

The second reason for turning to adverbs is that a number of researchers (e.g., Allerton, 1987; Kennedy & McNally, 2005; Paradis, 1997; Quirk, Greenbaum, Leech, & Svartvik, 1985) have already laid the groundwork for observing distributional differences between adverbs and the adjectives they modify. Third, adverbs are much more frequent in input than resultative constructions. Fourth and finally, (and I say this somewhat shamelessly), adverbs are easier to search for. I therefore leave investigation of how a child can learn about the semantics of adjectives from resultative constructions for future research.

Having identified a correlation between the surface form and the underlying semantic representation, the next step is to look at natural speech occurrences and determine the regularity with which this pattern of modification appears. In doing so, we are able to determine whether this source can be used by the child acquiring language to extract information concerning deeper representational differences among adjectives. In the next section I focus on an analysis of child-directed speech in particular, and then expand the search in the following section to a broader and larger corpus.

5.3. Adverbs in Child-Directed Speech

5.3.1. Corpus

Four longitudinal transcripts from two corpora available from the Child Language Data Exchange System (CHILDES) database (http://childes.psy.cmu.edu/) (MacWhinney, 2000) were targeted for analysis. These corpora were selected because they are well-known, used

---

universal and is available early in language development, and one which is language-specific and available later. Thus, although a generalization about the semantics of adjectives appearing in resultative constructions may be specific to English, that does not mean that it is not informative at all in language learning, only that it should not be one of the initial cues to a form-meaning correspondence.

frequently, and span a wide range of ages (Brown, 1973; Suppes, 1974). More specific details for each set of transcripts in both corpora are described in Table 13.

Table 13: CHILDES corpora searched

corpus	child	age range	# files	# mother utterances		# mother words
Brown	Adam	2;3-4;10	55		20,162	92,965
	Eve	1;6-2;3	20		9,788	41,478
	Sarah	2;3-5;1	139		27,642	122,794
Suppes	Nina	1;11-3;11	52		34,302	180,874
				total #	91,894	438,111

5.3.2. Method

Transcripts were searched using the CLAN program, available from the CHILDES website, with the command line in (69).

69. kwal +t*MOT +s[*adjective/adverb*] -w2 +w2 +u *.cha

This command tells the search engine to extract all of the mother's utterances in which a target adjective or adverb appears, including the two preceding and following lines, searching in all of the files with the .cha extension (the default file format in the CHILDES database) in the selected directory. A target adverb or adjective was inserted in place of the italicized portion of the command line. Preceding and following lines were viewed to provide a context for the utterances and because utterances are occasionally split over multiple lines, so that looking at one line in isolation does not always provide an accurate picture of the utterance in its entirety. Maternal imitations of child speech were permitted, but given the rarity of adverbs in child speech, this was not an issue.

Fifteen adverbs were targeted in the search. These adverbs represent a range of those that canonically restrict their modification to adjectives with closed scales (in their conventional usage) and those that do not, but which indicate gradability. Because of this split in modification strategies, these adverbs were dubbed *restricted* and *non-restricted* adverbs and are referred to as such in the remainder of this chapter. This set of fifteen adverbs searched for in the CHILDES transcripts is listed in Table 14.

Table 14: Adverbs searched for in CHILDES transcripts

adverb type	adverbs
restricted	*absolutely*
	almost †
	approximately
	completely
	half †
	hardly
	nearly
	partially
	roughly
	totally
non-restricted[73]	*really* †
	relatively
	somewhat
	too †
	very †

Once the search was complete, the search results were imported into an Excel workbook and reviewed and tallied by hand.

[73] Labeling these adverbs as 'non-restricted' is shorthand for indicating that they are able to modify both relative and absolute GAs. Clearly, though, these adverbs do have restrictions on the adjectives they are allowed to modify, since examples such as #*very wooden* or #*somewhat fatal*, where the adverb modifies a non-gradable adjective are easily constructed and unambiguously infelicitous. The pattern captured in Table 12 also demonstrates that *very* is not entirely unrestricted in its modification.

5.3.3. Analysis

Each occurrence of the adverbs listed above was reviewed, and then those that did not mark an instance of adjectival modification were filtered out of the set of results. Examples of adverb occurrences that did not involve adjectival modification are listed in (70)-(78).

70. Absolutely nothing. (Adam, file #33)

71. You should've been here the day the fish almost went down the drain. (Sarah, file #73)

72. You look like a monkey and a half. (Nina, file #33)

73. You can hardly tell because it has a big hat on. (Nina, file #42)

74. I really don't know why the farmer's running. (Adam, file #24)

75. Can I get some noodle soup too? (Eve, file #13)

76. There's too much snow. (Sarah, file #22)

77. It breaks very easily no. Doesn't it? (Nina, file #31)

78. Now put this at the very tip of the seal's nose. (Nina, file #49)

As a result of this filtering process, we see that only five of the adverbs listed above (those indicated by †) appeared in the transcripts of child-directed speech in an adjective-modifying capacity.[74] These were analyzed further for any selectional restrictions.

5.3.4. Results

The five remaining adverbs and the percentage of instances in which they modified adjectives are listed in Table 15.

---

[74] Adverbs that had the same form as an adjective in the positive form (e.g., *careful*, *far*, *fast*, *good*, *hard*, as in *be very careful*, *it won't go very far*, , *I tried very hard*, and so forth) were included as adjectives in this count, since they are presumably just as informative to the child.

Table 15: Adverbs modifying adjectives in four CHILDES transcripts

	child				
adverb	Adam	Eve	Sarah	Nina	total
almost	0/34	9/20 45.00%	2/29 6.90%	0/19	11/102 10.78%
half	0/16	0/0	2/20 10.00%	0/2	2/38 5.26%
really	0/56	0/4	11/74 14.86%	43/107 40.19%	54/241 22.41%
too	50/157 31.85%	23/125 18.40%	90/249 36.14%	122/462 26.41%	285/993 28.70%
very	104/140 74.29%	60/74 81.08%	61/80 76.25%	135/182 74.18%	360/476 75.63%

Two of these five adverbs (*almost* and *half*) place the kind of closed-scale selectional restrictions I have discussed on the adjectives they modify. The other three (*really*, *too*, *very*) do not. An initial glance at these data reveals two things. First, the majority of occurrences for these adverbs are instances that do *not* involve adjectival modification (1138 of 1850, or 61.5%). This percentage is even higher for the two restricted adverbs (127 of 140, or 90.7%).[75] Second, the number of non-restricted adverbs greatly overshadows the restricted ones (1710 versus 140 total occurrences, 699 versus 13 in an adjective-modifying capacity).

[75] All of the occurrences of *almost* in Adam's or Nina's files, and most of those excluded from analysis in Eve's and Sarah's files, referred to aspectuality of events (e.g., *almost fell*). Most of the occurrences of *half* in the transcripts were pseudo-partitive or partitive constructions (e.g., *half the people*, *half of his ear*) or measure phrases, including temporal references (e.g., *one half*, *half quart*, *half an hour*).

Let's now turn to the adjectives that are modified by these adverbs. Adjectives appearing with a frequency greater than 1 are listed in Table 16.[76]

Table 16: Adjectives modified by adverbs in CHILDES

adverb	adjectives modified
almost	*done*, *finished*, *well*[77]
really	*beautiful*, *cute*, *funny*, *good*, *hard*, *late*, *nice*, *pretty*
too	*big*, *cold*, *early*, *far*, *far away*, *fresh*, *hard*, *heavy*, *hot*, *large*, *late*, *little*, *long*, *old*, *rough*, *sharp*, *slippery*, *small*, *stiff*, *tight*, *tiny*, *tired*
very	*beautiful*, *big*, *brittle*, *busy*, *careful*, *close*, *cold*, *delicate*, *dirty*, *easy*, *far*, *funny*, *gentle*, *good*, *hard*, *heavy*, *hot*, *kind*, *little*, *long*, *nice*, *pretty*, *quiet*, *sharp*, *soft*, *soon*, *sticky*, *tired*, *uneven*, *well*

Quantity aside, there appears to be a qualitative difference in the meaning of the adjectives modified by *almost* as opposed to those modified by the other three adjectives. There is considerable overlap among the lists for unrestricted adverbs *really*, *too*, and *very*, and whereas nearly all of the adjectives listed with these adverbs are relative GAs, and therefore have a context-dependent standard, the adjectives modified by *almost* highlight progress toward a goal. Thus an *almost + adjective* phrase indicates a telic or bounded event, while a *really/too/very + adjective* phrase indicates how a property relates to a standard set by the context. Although these distributional differences are promising, the frequency of occurrence (cf. Table 15) is very low. If the language learner were going to make generalizations about the scalar structure of

[76] *Half* is not listed here because the only adjective-modifying occurrence involved an exchange between Sarah's mother and father about the coverage of color on a surface area.

[77] Instances with *well* were included if it referred to someone's state of well-being, not if it referred to how something was done; this adjective is therefore like the others in the list for *almost* in that the event being described is bounded (i.e., the person's being sick ends when s/he is well/healthy).

adjectives from adverbs such as these, more data points and a wider range of adverbs (especially restricted adverbs) would need to be observed. However, it is hard to know how best to project to a more realistic sample given the present set of data from CHILDES, which is admittedly very small and restricted to a narrow range of discourse contexts.

This is not to say that the non-adjective-modifying occurrences of these adverbs are totally uninformative. To be clear, I am saying that they are not *directly* informative about the scalar structure of a GA that is modified. Their presence in the discourse can still signal the adverb's set of selectional restrictions, of which the restriction to select for a maximum standard absolute GA is a member, and therefore be informative about the scalar structure of GAs. The meaning of the adverb should remain relatively constant, whether it is modifying an adjective, a verb, or any other category.

Let's take *almost* as an example. Across the four transcripts, there were 102 instances of this adverb, only 11 of which were instances of adjectival modification. Now, we can look at the other 91 occurrences of this adverb and ask if the child could learn anything about scalar structure from these. These instances can be divided into three types. The first are those like (71) and ((79)-(80)) below, where an event could have occurred, but ultimately failed to. There were 49 such instances.

79. You almost fell down because you weren't looking where you were going.

(Adam, file #12)

80. Remember what happened yesterday? Ya almost choked. (Sarah, file #57)

The second type is where conditions are such that they are close to the desired or standard state of affairs ((81)-(87)). There were 21 instances like these.

81. They almost fit? (Adam, file #19)

82. A: Do men have to shave, and boys, too?

M: Boys when they're almost men, not little boys like you. (Adam, file #38)

83. This is almost like tapioca. (Eve, file #7)

84. M: Tell me what color the bookcase is.

N: Red.

M: Almost. It's pink. (Nina, file #17)

85. She's almost as big as you are, isn't she? (Nina, file #37)

86. It's almost off. Can you get it the rest of the way? (Nina, file #53)

87. S: Am I as big as her?

M: Almost. It won't be long. (Sarah, file #96)[78]

The third type includes references to an event that is expected to occur in the near future but has not yet occurred ((88)-(91)). There were 21 such instances.

88. Adam almost has your umbrella ready, Diandra. (Adam, file #48)

89. E: I want some more grape juice.

M: I think not. It's almost time to have lunch. (Eve, file #17)

90. M: Now it's almost summer. So what can we do in the summer?

N: Swim in the water. (Nina, file #30)

91. It's almost time for Nina to go to sleep, isn't it? (Nina, file #51)

What these instances have in common is that *almost* flags a punctual moment, transition, or endpoint, where a state of affairs would have held, could hold, or will hold in the near future, but does not hold at present. (See discussion in § 3.5.1 and the examples of entailment patterns

---

[78] One could argue for this example being classified as the third type instead. Since the child expresses a desire to reach a certain degree of height, I have classified it as the second type.

associated with *almost* presented there.) Relatedly, Rotstein and Winter (2004) observe that 'simple sentences' with *almost* entail their negation (cf. (92a-c)).

92. a. Almost every/no student came ⇒ It is not the case that every/no student came.

b. Bill almost missed the train ⇒ Bill didn't miss the train.

c. The work is almost complete ⇒ The work is not complete.

(Rotstein & Winter, 2004, (27))

A child attending to cross-categorial instances of modification by *almost* who makes this observation about an obligatory maximal value on a scale (cf. Amaral, 2006; Hitzeman, 1992), should indeed be learning something about the scalar structure of GAs, since the adjective modified by *almost* should share something in common with these other occurrences. If children were not able to generalize across categories, then they would have to come to a similar conclusion for each set of instances, a situation which seems rather unparsimonuous.

This brief look at *almost* should also demonstrate that a focus on adverbial modification of adjectives is narrow, indeed, and that the numbers reported in this chapter are therefore conservative. It appears to be the case that a more thorough analysis of adverbial modification in the input would illustrate that the language learner is provided with considerable information about scalar structure, since that is the one aspect that can connect these occurrences across categories.

5.3.5. Discussion

The analysis of the CHILDES transcripts revealed that while distributional differences exist between the type of adjective modified by an adverb such as *almost* and adverbs such as *really*, *too*, and *very*, both the number of raw instances of adverbs in the transcripts and the ratio of adjective-modifying instances to all instances of adverbs in the corpus are very low. There are two possible conclusions from these data. We could simply say that adverbs are sparse in the input, and that children could not use these lexical items as a bootstrap for anything. Or, it is possible that the CHILDES transcripts provide us with such a small corpus that they may actually not be giving the input a fair shake. In addition, while these transcripts may provide a picture of the caregiver's speech to the child in the home environment, they may not be doing justice to the full range of discourse contexts – and therefore linguistic input – to which the child is exposed.

The analysis of the CHILDES transcripts was a good starting point, since it is the most detailed source of information we have about child-directed speech. The next step is to expand the search to a corpus that is larger and will provide a better model of a combination of child-directed speech and ambient speech. In the next section, I turn to the British National Corpus for a more comprehensive analysis of adverbial modification in the input.

5.4. Adverbs in the Exposure Language

5.4.1. Corpus

The initial search of the CHILDES transcripts yielded little evidence that adverbial modification of adjectives in caregiver speech could be used as a cue to scalar structure. However, it is highly likely that this is a result of the size of the sample and the restricted range of recording environments. In this search the corpus used was the British National Corpus (BNC) (http://info.ox.ac.uk/bnc). The BNC has approximately 100 million words. Instead of searching the entire corpus, which contains both spoken and written texts, this search is focused on the spoken language corpus, since this provides us with a better picture of the ambient speech to which children are exposed. This subset of the BNC has approximately 10,365,000 words.[79]

Despite the fact that the BNC is commonly used by both British and American linguists, one might nonetheless wonder whether its use here is justified. Its merit is that it gives us a broader picture of the exposure language. However, this corpus can be narrowed even further to a subset that is comparable to the four CHILDES transcripts used in the previous section in terms of speaker demographics. I will use this subset of the BNC, captured in Table 17, for a point of comparison to the CHILDES transcripts to show that it is reasonable to turn to the BNC for analysis.

Table 17: Filtering of spoken BNC corpus for comparison with CHILDES transcripts

BNC restrictions	number of words
entire corpus	100,106,008
spoken texts	10,365,464
demographic texts	4,206,058
female speaker	2,249,237
age 25-34	448,603

A comparison of this subset of the BNC with the previous CHILDES transcripts (see Table 18) demonstrates that these two corpora are roughly comparable in both size and composition as concerns adverbial modification.

[79] According to the online manual, the spoken texts are divided into two sets. The "context-governed" texts (approximately 60%) include scripts from lectures, speeches, news broadcasts, interviews, meetings, parliamentary proceedings, sermons, and so on. The "demographic" texts (approximately 40%) are from recordings of unscripted informal conversation by 124 volunteers (with an equal number of males and females) representing a variety of demographic characteristics (e.g., age, region, social class).

Table 18: Comparison of four CHILDES transcripts and a subset of the spoken BNC corpus

	four CHILDES transcripts	subset of spoken BNC
total # words	438,111	448,603
# adverbs	1860	2747
# adverbs modifying adjectives	713	945
% of all adverbs modifying adjectives	38.33%	34.40%
% modifying adjectives		
almost	10.78%	0.00%
half	5.26%	1.90%
really	22.41%	20.21%
too	28.70%	50.37%
very	75.63%	68.31%

The total number of adverbs and the percentage of adverbs modifying adjectives is relatively similar, which indicates that the quality of the speech in these two sources for the purpose of this analysis is similar. Whereas with the CHILDES transcripts, we were unsure how to regard the observed patterns or how to project from the smaller sample to a larger and more realistic sample, given of the similarity of this subset of the spoken BNC to the four CHILDES transcripts with respect to our research focus, we can now cast our net wider and move beyond this limited subset of the BNC to the entire spoken corpus.

5.4.2. Method

The spoken BNC was searched through The Zurich BNCweb Query System. Two sets of searches were conducted. The first focused on adverbs and the adjectives they modify, and the second on adjectives and the adverbs that modify them in the corpus. For each of these two searches, ten lexical items (two groups of five) were targeted. These items, listed in Table 19, represent canonical members of each of two different groups. This included five *restricted* and five *non-restricted* adverbs, and five relative GAs and five maximum standard absolute GAs. In searching for these particular items, I hoped to find evidence concerning whether adverbial modification of adjectives in this corpus could serve as a cue to scalar structure.

Table 19: Adverbs and adjectives targeted in search

adverbs	adjectives
restricted adverbs	maximum standard absolute GAs
almost	*clean*
completely	*dry*
entirely	*empty*
half	*full*
totally	*straight*
non-restricted adverbs	relative GAs
extremely	*big*
really	*high*
relatively	*long*
too	*tall*
very	*wide*

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 co-occurrence patterns between adverbs and adjectives. For adverbs, this meant searching for positive form adjective (labeled as "AJO") in the "+1" position. For adjectives, this meant searching for any sort of adverb that was not a particle or a *wh*-word ("AVO") in the "-1" position. This second set of results was then downloaded and compiled in a Microsoft Excel workbook. These lists were reviewed and coded by hand. For adverbs, each co-occurring adjective was coded as maximal or not. For the adjectives, each modifying adverb was coded as *restricted* or not.[80] This method

[80] Here, I use the term *restricted* to indicate those adverbs which select for a maximally closed scale (i.e., adverbs that modify maximum standard absolute GAs. However, this terminology should not be interpreted as indicating that only these adverbs are restricted in their modification. I merely intend for it to contrast with what I am calling

provided a rough measure of the distribution of adverb-adjective bigrams (collocations, or co-occurring items), even if there might be disagreement with some of the individual classifications.

The following criteria were used for classifying the lexical items. Adverbs were classified as *restricted* if they exhibited behavior similar to *almost*, *perfectly*, and *completely*, three adverbs that have been identified as restricted by Kennedy and McNally (2005) and Rotstein and Winter (2004) – that is, if the resulting phrase meant that possession of the property was almost achieved, that it applied flawlessly or without exception, or that it picked out proportions of coverage, thereby referencing a closed scale.

A number of diagnostics helped to classify adjectives as maximum standard absolute GAs. First, if the adjective could be modified by *100%* or *almost*, it was treated as maximal. Given that I am trying to establish a regularity among bigrams of adjectival modifiers and adjectives, using the reliability of a modifier's selectional restrictions to classify an adjective might seem a rather circular test. However, there are reasons why this test is valid for the present purposes. Neither of these modifiers ends in *–ly*, a canonical suffix that highlights membership in the adverbial category, or is used in the experiments presented in Chapters 5. 100% is also used by Kennedy and McNally (2005) and *almost* is used by Cruse (1980) and by Rotstein and Winter (2004) to pick out different kinds of adjectives corresponding to the distinctions laid out in this chapter.[81]

---

non-restricted adverbs, those adverbs which are relatively free in their appearance with a wide range of adjectives, but which may also favor or place restrictions on adjective meaning, and therefore on scalar structure. This point is illustrated in § 5.4.4.2.

[81] Cruse (1980) distinguishes between *complementary* and *non-complementary* pairs of *antonymous adjectives*. These two categories can be differentiated with entailment patterns. Non-complementary pairs correspond to

A second diagnostic relates to distinctions between existential and universal predicates made by Kamp and Roßdeutscher (1992), which were recruited by Yoon (1996) and Rotstein and Winter (2004) to distinguish between total and partial predicates. An existential or partial adjective can be predicated of an entity if it has *any* amount of the relevant property (e.g., moisture, dirt, sickness, etc.). A universal or total adjective can be predicated of an entity if there is *no* degree of that property (e.g., an entity can be *clean* if it has no dirt, dry if there is no moisture, healthy if it is not at all sick, and so on). An adjective in the corpus search results was therefore classified as maximal if an antonym could be posited, and the corresponding property was lacking. This distinction was behind the use of *for the most part* as a diagnostic, since this phrase indicates progress towards universal application of the property and picks out a part structure, which can only exist with a closed scale.

Finally, an adjective was classified as maximal if it would be interpreted by default as non-gradable but also had a highly frequent interpretation that allowed for the property to hold without the meaning strictly applying. For example, words like *equal* and *identical* were coded as maximal. Strictly speaking, there must be absolute correspondence between quantities for *equal* to be predicated of two entities, and a one-to-one correspondence of qualities for *identical* to be predicated of two entities. However, modification by words and phrases such as *pretty much*, *more or less*, *roughly*, *basically*, and *almost* seems to highlight the maximal treatment of these adjectives.

---

relative GAs, since if x is not short, it does not immediately follow that x is long. Complementary pairs such as *clean*/*dirty* or *wet*/*dry* roughly correspond to absolute GAs and can be modified by *almost*. Members of such pairs are negations of each other, such that given two antonyms A_1 and A_2, if an entity x can have one of these properties, it must have either one or the other.

5.4.3. Analysis

Three analyses were conducted. The first is a Chi-square analysis of the raw overall frequency of the bigrams in each of the four categories. The second uses a method modeled after Goldberg, Casenhiser, & Sethuraman (2005) and Murphy (1982), calculating the conditional probability to measure the degree to which an adverb is informative about the scalar structure of the adjective. The third analysis compares the most frequent adverb-adjective collocations for each of these 20 lexical items and incorporates a ranked t-test analysis of the results (Manning & Schütze, 1999).

5.4.4. Results

5.4.4.1. Analysis 1: Distribution of Adverb-Adjective Bigrams

To begin, 100 bigrams formed from a combination of the 20 targeted lexical items were entered into the search engine and their frequencies recorded. The complete set of frequencies is presented in Appendix G. A collapsed summary is presented here in Table 20. This first pass at a frequency analysis reveals the predicted asymmetry in adverbial modification: non-maximal adjectives do not appear with restricted adverbs, while maximal GAs appear with both restricted and non-restricted adverbs ($\chi^2 = 268.62$, $p \leq 0.001$), and just over one fifth (21.5%) of the maximal GAs appear with the restricted adverbs.

Table 20: Co-occurrence of 20 adverbs and adjectives in spoken BNC

	maximal adjectives	relative adjectives
restricted adverbs	26	0
non-restricted adverbs	95	1226

Given the promise of this initial analysis, the analysis was extended to the full set of adjectives modified by the 10 target adverbs and the full set of adverbs modifying the 10 target adjectives. For each of these two sets of lexical items, adverb-adjective bigrams with a frequency greater than one were tallied. Distributions for each set of items are presented separately in the following two tables.

The frequencies of adjectives appearing with the 10 adverbs are presented in Table 21. These results demonstrate that non-maximal adjectives are significantly more likely to appear with non-restricted adverbs than with restricted ones ($\chi^2 = 3928.58$, $p < 0.001$). And while maximum standard absolute GAs appear with both types of adverb, approximately one third of the cases of adverbial modification (32.7%) are with restricted adverbs (compared to a mere 1.3% of non-maximal adjectives).

Table 21: Adjectives appearing with adverbs in spoken BNC corpus

	maximal adjectives	other adjectives
restricted adverbs	490	257
non-restricted adverbs	1010	18957

Examples of restricted adverbs modifying non-maximal adjectives include *half racist*, *totally ridiculous*, *entirely familiar*, *half thick*, and *completely anonymous*.

Turning to the adverbs modifying the 10 target GAs, we observe the same distributional asymmetry, presented in Table 22. Non-maximal adjectives are much more likely to be modified by non-restricted adverbs, and while maximum standard absolute GAs can appear with both types of adverbs, just over one tenth (10.6%) of these adjectives are modified by restricted adverbs (compared to less than one percent (0.2%) of the non-maximal adjectives). The four instances of non-maximal adjectives modified by a restricted adverb included *exactly high* and *nearly long*.

Table 22: Adverbs modifying adjectives in spoken BNC corpus

	maximal adjectives	other adjectives
restricted adverbs	39	4
non-restricted adverbs	330	2453

A Chi-square analysis shows that the distribution of frequencies is highly significant (χ^2 = 231.85, $p < 0.001$).

5.4.4.2. Analysis 2: Conditional Probability and Scale Informativity

Whereas the previous analysis evaluated the distribution of bigrams involving the 20 target lexical items at a more global level, the second analysis looks at each lexical item and the probability that it will appear with a member of one of the other two sets, and then calculates the overall probability for each of the four sets. As above, two separate analyses were conducted, one with the adverbs as the input, and the other with the adjectives as the input. This approach gives us two tests for determining whether adverbs are informative about the scalar structure of the adjectives they modify.

The first analysis, which targets the adjectives appearing with each of the 10 adverbs, asks the following question: if an adverb appears in the corpus modifying an adjective, to what degree is it informative about the scalar structure of that adjective (i.e., signaling that the scale is closed and that the standard is maximal)? To answer this question, it is necessary to calculate the proportion of instances involving modification of a maximum standard absolute GA out of the total number of instances. Table 23 captures this information. Column A indicates the total number of appearances in the corpus. Column B indicates the number of instances in which the adverb modifies an adjective. Column C indicates the number of instances in which that adjective is a maximum standard absolute GA. Column D takes the information from Columns

B and C and returns a conditional probability, reflecting the degree to which an adverb is informative about a closed scalar structure, given that it modifies an adjective in the corpus.

Table 23: Adverb's informativity of (closed) scalar structure in spoken BNC corpus

adverbs	A # instances in spoken BNC	B # instances modifying adjective	C # instances modifying abs. maximal GA	D ℘(scale-informative| modifies an adjective) (C/B)
restricted adverbs				
almost	1225	143	90	0.63
completely	822	327	184	0.56
entirely	416	145	66	0.46
half	6094	94	24	0.26
totally	802	417	221	0.53
total	9359	1126	585	0.49
non-restricted adverbs				
extremely	480	390	10	0.03
really	17853	2925	110	0.04
relatively	282	212	33	0.16
too	6511	2598	70	0.03
very	25041	15156	855	0.06
total	50167	21281	1078	0.06

A two-tailed t-test reveals that these results are highly significant ($t(8) = 6.25$, $p = 0.00025$). The degree to which an adverb is informative about the scalar structure of the adjective it modifies – as taken from the proportion of the number of maximum standard absolute GAs out of the total number of adjectives modified – is higher if the adverb is restricted than if it is non-restricted.

Note that this analysis presumed that the only 'scale-informative' uses of the adverbs will be those that modify maximum standard absolute GAs, and therefore those that signal that the adjective's meaning corresponds to closed scale. However, if we looked at the other occurrences of the adverbs in order to determine whether these uses could also be informative about scalar structure, as we did in § 5.3.4, we would find that this list is populated by items which highlight *almost*'s meaning of approaching a point of reference or boundary, and hence it's selection of a closed scale. A cursory look at the list of 1225 utterances in which *almost* appears shows that approximately one third of these, excluding those in which *almost* modifies an adjective, are ones in which the following item has an unambiguous 'maximal' reading: e.g., *all* (22), *always* (10), *any* (17), *anything* (23), *certainly* (48), *entirely* (17), *every* (19), *everybody* (4), *everything* (9), *exactly* (10), *exclusively* (10), *immediately* (16), *inevitably* (4), *invariably* (5), *no* (3), *nothing* (3), *there* (8), *totally* (5), *unanimously* (2), *universally* (2), *without* (3), and many others. There are also at least 70 cases of *almost* followed by a number word or a verb whose meaning incorporates an amount, such as *double* or *quadruple*. Thus, attending to the range of items modified by *almost* would serve the child well in recognizing that this item modifies those items that can supply an endpoint. GAs with maximally closed scales are one such case.

The second analysis looks at the adverbs modifying each of the targeted GAs and asks the following question. Given that an adjective appears in the corpus modified by an adverb, what is the probability that the adverb will be a member of the restricted set? Or, put another way, if an adjective is modified by an adverb, what is the probability that the adverb is one that selects for an adjective with a maximally closed scale? The results of this analysis are presented in Table 24. The information presented in Columns A-C is similar to those in Table 23. As above, a proportion derived from the numbers in Columns B and C is presented in Column D. Here, this number represents the probability that an adjective appears with an adverbial modifier requiring that the adjective's scale be maximally closed when the adjective is modified by an adverb.

Table 24: Probability of being modified by a restricted adverb in spoken BNC corpus

	A	B	C	D
adjectives	# instances in spoken BNC	# instances modified by an adverb	# instances with restricted adverb	℘(modified by a restricted adverb| modified by an adverb) (C/B)
abs. max. GAs				
clean	781	93	4	0.04
dry	589	71	3	0.04
empty	386	32	4	0.13
full	2178	167	9	0.05
straight	1432	94	3	0.03
total	5366	457	23	0.05
relative GAs				
big	5704	702	0	0.00
high	2010	552	2	0.00
long	6153	1132	4	0.00
tall	233	61	1	0.02
wide	551	105	1	0.01
total	14651	2552	8	0.00

A two-tailed t-test reveals that these results are significant ($t(8) = -2.78$, $p = 0.024$), although the numbers are quite small. Given the statistical significance, though, we can conclude that when an adjective is modified by an adverb, the probability of it being modified by a restricted adverb is higher if it is a maximum standard absolute GA than if it is a relative GA. This pattern suggests that if we expanded our search to a larger set of adjectives, the trend might surface more robustly.

Given the small number of instances in which these adjectives were modified by a restricted adverb – regardless of their GA status – it would be interesting to know what other adverbial modifiers appeared with these adjectives. By far the most frequently occurring adverbs were those which highlighted the gradable status of all of these adjectives. This can be seen in Table 25, which presents the number of occurrences of these adjectives with the frequently-occurring adverbs *a bit*, *as*, *fairly*, *quite*, *really*, *so*, *too*, and *very*. These numbers are then summed and fed into equations to capture (a) the frequency of these instances out of all instances in which these adjectives were modified by an adverb in the spoken BNC (column I) and (b) the percentage of instances remaining after the analyses in the table above and below in which these adjectives were still modified by an adverb (column J).

Table 25: Frequently appearing non-restricted adverbial modifiers in spoken BNC

	A	B	C	D	E	F	G	H	I	J
adjectives	*a bit*	*as*	*fairly*	*quite*	*really*	*so*	*too*	*very*	prop'n of all adverbial instances $(\Sigma_{A\ldots H})/B_{15}$	% modified instances remaining $(B_{14}-(C_{14}+I_{15}))$
abs. max. GAs										
clean	0	3	1	1	6	10	3	0	0.55	40.86%
dry	5	5	4	0	4	2	7	0	0.51	45.07%
empty	1	0	1	1	0	0	0	0	0.09	78.13%
full	4	10	2	1	10	3	13	6	0.39	55.69%
straight	1	2	3	3	4	3	7	0	0.35	61.70%
total	11	20	11	6	24	18	30	6	0.41	53.83%
relative GAs										
big	21	72	13	10	63	42	38	143	0.83	16.81%
high	3	47	20	12	44	9	35	77	0.84	16.12%
long	17	125	12	0	21	17	209	216	0.77	22.88%
tall	2	12	4	3	14	2	1	5	0.89	9.84%
wide	1	15	4	1	9	8	7	9	0.78	20.95%
total	44	271	53	26	151	78	290	450	0.80	19.36%

As in Table 24, the proportion in column I of this table could be interpreted in terms of a conditional probability[82]: $\wp$(modified by a 'scale-selecting' adverb|modified by an adverb), where a 'scale-selecting' adverb is any adverb that requires the adjective it modifies to be

[82] The subscripts in columns I and J refer to the table numbers (14 or 15).

gradable. Considered from this perspective, this analysis yields two important findings. First, relative GAs have a much higher probability of appearing with these general 'scale-selecting' adverbs than do maximum standard absolute GAs (two-tailed t-test: $t(4) = 6.4$, $p = 0.003$). This could indicate that the input provides a great deal more information about these relative GAs mapping onto any scale, whatever its structure, *to the extent that such a conclusion can be drawn from adverbial modification*. This analysis also gives us a conservative estimate, since it only targets a handful of frequently-occurring adverbial modifiers with this feature, leaving out other, less frequent adverbs, such as *extremely*, *pretty*, *rather*, and *relatively*.

Second, relative GAs also have a much higher probability of appearing with more general 'scale-selecting' adverbs than with restricted adverbs, which select a maximally closed scale (two-tailed t-test: $t(4) = 42.3$, $p < 0.00001$). This conclusion results from comparing column D in Table 24 and column I in Table 25 above. In fact, taking this new set of adverbs into account, we see that they account for the lion's share (roughly 77-90%) of the instances in which the relative GAs appear modified by an adverb in the spoken BNC. By contrast, we are still left needing to account for a large percentage of the instances in which maximum standard absolute GAs are modified by an adverb, as seen in column J in Table 25 above.

What kind of other adverbs appear with these adjectives? To answer this question, consider the myriad perspectives a speaker might assume in any given situation (cf. Clark, 1997; Gleitman, 1990). Because adverbs can be used to express these perspectives, we should expect to find across the two subclasses of GAs that some percentage of adverbial occurrences cannot be accounted for by referring to scalar structure alone. For example, we can use adverbs such as *now* or *then* to compare the current state of affairs to the state of affairs at another point in the past; adverbs such as *always*, *ever*, *never*, or *usually* to express whether this state of affairs normally holds; and an adverb such as *abnormally* or *exceptionally* to comment on how it deviates from the normal state of affairs. We can express our attitudes by using adverbs such as *dangerously* or *disturbingly*, or our certainty (or lack thereof) with those such as *probably*. Indeed, adverbs such as these populate both lists.

What is interesting to note with respect to our two classes of GAs is that relative GAs are on the whole much more frequent than maximum standard absolute GAs (over 5.5 times in the current analysis), and they are more likely to be modified by an adverb (two-tailed t-test: $t(4) = 3.69$, $p = 0.02$), but also more likely to appear with a narrower range of adverbs (two-tailed t-test: $t(4) = -2.5$, $p < 0.07$, marginally significant). This fact is highlighted by examining the type-to-token ratio for each of these two sets of adjectives (cf. Table 26), where a 'type' is a lexical item (i.e., an individual adjective) and a 'token' is an instance of that adjective.

Table 26: Type-token analysis of adverbial modifiers

adjectives	A # instances in spoken BNC	B # instances modified by an adverb	C # types of adverbs	D type-to-token ratio (C/B)
abs. maximal GAs				
clean	781	93	34	0.37
dry	589	71	27	0.38
empty	386	32	22	0.69
full	2178	167	44	0.26
straight	1432	94	33	0.35
total	5366	457	160	0.35
relative GAs				
big	5704	702	43	0.06
high	2010	552	50	0.09
long	6153	1132	60	0.05
tall	233	61	16	0.26
wide	551	105	25	0.24
total	14651	2552	194	0.08

What are the implications of this analysis for the language learner? A child attending to adverbs in the input is exposed to a large number of relative GAs, and when these adjectives are modified by an adverb, the child is practically bombarded with evidence that these adjectives are indeed gradable. Furthermore, although they are exposed to a much smaller number of absolute GAs and a much wider range of adverbs, within this set, there is (with the exception of *empty*) a large percentage of adverbs highlighting their gradable status and a smaller but still statistically significant percentage of restricted adverbs appearing with these adjectives. The child is therefore getting consistent evidence that these adjectives map onto a maximally closed scale.

5.4.4.3. Analysis 3: Ranked Frequencies

In the third and final analysis of the BNC spoken corpus, I examine a different set of data. Instead of large sets of bigram frequencies, here the data are the most frequent collocations for each lexical item (in other words, separate sets of bigrams). As in the two previous analyses, this analysis was also conducted on each of the two sets of lexical items in turn. First, I look at the ten most frequent adjectives appearing with each of the ten adverbs. Next, I look at the ten most frequent adverbs appearing with each of the ten adjectives. The complete lists are presented in Appendix H (the most frequently modified adjectives for each of the ten adverbs) and Appendix I (the most frequent adverbial modifiers for each of the ten adjectives).

Let's first focus on the lists of adjectives appearing with the adverbs. Inspection of the list reveals that there is little overlap between the adjectives appearing with the restricted and non-restricted adverbs. The two sets have in common only the adjectives *happy* (which occurs once in each), *new* (which occurs once in the non-restricted set and three times in the restricted set), and *sure* (which occurs twice in each set). There is, however, overlap among the adjectives *within* each of the two sets. Adjectives modified by restricted adverbs are generally not relative

GAs, and have either minimal or maximal interpretations. *Different*, *new*, *separate*, and *wrong* appear in three adverb lists. By contrast, most of the adjectives modified by non-restricted adverbs are relatives GAs. *Good* and *high* appear in four adverb lists, and *big*, *important*, *nice*, and *small* appear in three adverb lists. These ten adjectives were fed into a ranked t-test analysis, which confirms this distributional split (Manning & Schütze, 1999). See Table 27.

Table 27: Ranked t-test analysis of adjectives appearing in multiple adverb lists

adjective	*t*	adjective frequency	C (restricted adverb + adjective)	C (non-restricted adverb + adjective)
separate	4.1294	531	18	0
new	4.1573	6632	28	15
wrong	6.4094	2795	46	0
different	14.5860	5165	222	0
high	16.2122	2010	0	282
big	17.7129	5704	0	367
small	18.2843	2270	0	356
important	24.5812	2894	0	632
nice	31.9099	6247	0	1078
good	48.2977	16694	0	2492

C = collocation (co-occurrence)

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{N}}}$$

numerator = (# bigram occurrences / N) – (P(adjective)*P(adverb))

N = number of words in spoken corpus

s = sample mean

Critical t value for α = 0.005 and ∞ degrees of freedom = 2.576

Turning now to the to lists of adverbs modifying the adjectives, we see that unlike the lists for the adjectives appearing with adverbs, there is quite a bit of overlap among the lists of adverbial modifiers. See Table 28 for more details on individual data points.

Table 28: Overlap among lists of frequent adverbs modifying the ten GAs

	number of adjectives modified (out of 5 GAs in each set)	
adverb	maximal GAs	relative GAs
quite	4	5
very	4	5
so	4	4
as	3	5
really	2	5
too	2	5
fairly	2	4
a bit	2	3
that	1	3

While these adverbial modifiers highlight the gradability of these adjectives, they do not partition the adjectives into sets based on scalar structure. However, some adverbs do appear in the lists of frequent modifiers of maximum standard absolute GAs but not relative GAs – *absolute*, *completely*, *half*, *mainly*, *mostly*, *perfectly*, and *virtually*. These bigrams, however, are rather infrequent and do not generate a pattern sufficiently significant to merit comment.

5.4.5. Discussion

Three sets of analyses of data collected from the British National Corpus demonstrated that there are both statistically-significant differences between restricted and non-restricted adverbs and the adjectives they are likely to modify, and between maximum standard absolute GAs and relative GAs and the adverbs that are likely to modify them. These results contrast with the analysis of caregiver speech in the CHILDES transcripts. However, if we assume that the input children are exposed to is not restricted to the caregiver speech captured in the CHILDES transcripts, and that it is more like the speech captured in the spoken BNC corpus – which contains both caregiver speech and ambient speech – then we can say that the language learner is provided with evidence of distributional differences among adverbs and the adjectives they modify, and further, that these differences are driven by the scalar structure of the adjectives.

5.5. General Discussion

I set out in this chapter to determine whether the input is rich enough for the learner to learn about distinctions among GAs. Specifically, I asked if there is enough regularity of adverbial modification to distinguish between relative GAs such as *big*, which have open scales, and maximum standard absolute GAs such as *full*, whose scales are closed. I began with an analysis of four CHILDES transcripts, and although there were differences in the distribution of adjectives occurring for a very small set of adverbs, adverbs – specifically restricted adverbs selecting maximally closed scales – were relatively rare in the child directed speech. Because it was hard to infer from this small corpus whether these results were representative of the real speech to which children are exposed, I turned to an analysis of the British National Corpus, first comparing a subset of the spoken BNC to the initial CHILDES transcripts to demonstrate that the sources are comparable with respect to adverbial modification.

The analysis of the spoken BNC revealed two main findings. First, maximum standard absolute GAs such as *full* have a higher probability of being modified by restricted adverbs such as *completely* than do non-maximal GAs such as *big*. Second, while non-maximal GAs are more likely to be modified by adverbs than are maximum standard absolute GAs, they are overwhelmingly more likely to be modified by more general 'scale-selecting' adverbs such as *very*, which highlight their gradability. Thus, taking this corpus as an indication of what the exposure language is like for the language learner, we have reason to conclude that children are provided with positive evidence of the scalar structure of adjective in the input. What implication, then, does this have for language learning?

I propose that features of the input, such as patterns of adverbial modification, play a role in how children classify adjectives with respect to scalar structure: distributions of adverb-adjective bigrams provide the learner with cues that are correlated with more abstract conceptual or semantic representations. When the child encounters a new adjective, she will need to know how to classify it. Is it relative, or absolute? Does it have an open or closed scale? She will need some evidence to make this classification. In this account adverbial modification is helping her do this work. Restricted adverbs such as *completely* act as a kind of zoom lens, helping the learner identify that the adjectives they modify must have a closed scale. If this terminology sounds familiar, it is because it has been used to describe the work that the syntactic structure of a sentence does in syntactic bootstrapping in the verb learning domain (cf. Fisher, Hall, Rakowitz, & Gleitman, 1994; Gleitman, 1990; Landau & Gleitman, 1985). My account is slightly different, however, and will be fleshed out in much more detail in Chapter 7. For now, based on the very promising results of the corpus analysis, I will offer the preliminary conclusion that the distribution of adverbs in the input, guided by their selectional restrictions on modification, is a reliable and informative cue for the child learning adjectives. I will return to the question of what children might do with such cues later in the dissertation.

Note, however, that when the selectional restrictions of an adverb are not known, the child is not completely at a loss. Assuming that the child is already equipped with representational differences for at least a small group of GAs, attention to collocations – or, frequently occurring adverb-adjective bigrams – in the input can help draw attention to distributional similarities and differences among lexical items. When overlap is detected among adjectives appearing with adverbs, the child may posit similar representations for adjectives that appear with the same adverbs, or similar characteristics for adverbs modifying the same sorts of adjectives.[83] Thus above and beyond using the selectional restrictions of adverbs to classify the representations of adjectives, the learner can also use the same sort of analyses presented in this chapter to posit a meaning for newly-encountered adverbs, and these meanings will in turn aid in the process of learning about adjectives.

If this account is on the right track, it is important to demonstrate not only that the input is rich enough to support this learning path, but that children approach learning in this way. These regularities in the input may be *useful* to the language learner, but do children *use* them? In the next chapter, I report on a set of experiments designed to investigate whether infants at 30 months of age attend to adverbial modification when assigning interpretations to novel adjectives. Based on the results presented in this chapter and the next, I offer a detailed examination in Chapter 7 of the process by which infants use the meaning and distributional patterns of adverbs to classify adjectives based on scalar structure.

[83] This process might rely on a highly frequent GA (cf. Goldberg, Casenhiser, & Sethuraman, 2004) or a clustering of highly-frequent adjectival exemplars with each adverb. See also related discussions in Ninio (1999a, b).

Chapter 6: Using Adverbs to Learn New Adjectives.

6.1. Introduction

The experiments presented in this chapter were designed to determine whether infants make use of adverbial modification when assigning an interpretation to a novel adjective. This experiment capitalizes on a distinction based on the selectional restrictions of adverbs, as illustrated in (93). While an adverb such as *completely* felicitously modifies an adjective such as *full*, it cannot felicitously modify an adjective such as *tall*.

93. completely full/#tall

The difference in semantic representations underlying the surface-level difference between these two gradable adjectives (GAs) is discussed in more detail earlier in this dissertation (see, in particular, § 3.5.1and § 5.2.2). Briefly stated here, the distinction arises from the way in which these representations reference scalar structure. While a maximum standard absolute GA such as *full* maps onto a closed scale, a relative GA such as *tall* maps onto an open scale. Because a modifier such as *completely* requires reference to a maximal endpoint of a scale,[84] it looks for a GA that can provide this element. A maximum standard absolute GA can; in some cases, even a minimum standard absolute GA can, if the object's boundaries supply the endpoint (e.g., *completely wet*/*spotted*). A relative GA simply cannot.

---

[84] Since I adopt terminology from Kennedy & McNally (2005), I refer to the maximal endpoint of the scale. However, what I say here is perfectly compatible with other taxonomies (e.g., Cruse, 1986; Rotstein & Winter, 2004; Yoon, 1996) where maximum standard absolute GAs are described instead as lacking a property (e.g., the property of being straight entails having no bend; being clean entails having no degree of dirtiness, etc.). Using this line of reasoning, a *completely*+relative GA combination is barred because although values referenced by relative GAs approach the zero value, they never actually reach it (cf. Cruse, 1986, pp. 205-206).

In the previous chapter I presented the results of a corpus search that demonstrated that statistically significant distributional differences between adverbs and the adjectives they modify can be found in the speech to which the language learner is exposed. Specifically, I showed that maximum standard absolute GAs such as *full* have a higher probability of being modified by restricted adverbs such as *completely* than GAs such as *big* do. Second, non-maximal GAs such as *big* are overwhelmingly more likely to be modified by more general 'scale-selecting' adverbs such as *very*, which highlight their gradability, than by a restricted adverb such as *completely*. Knowing that these distributional differences exist, however, does not mean that the language learner will recruit them in word learning. Here, we seek to determine whether infants use an adverbial information source as a cue to the subclass of a gradable adjective.

In the following experiment, *completely* is contrasted with the adverb *very*, which is relatively free in its modification of adjectives (cf. (94)).[85]

94. very full/tall

Infants are presented with a series of novel adjectives, modified either by *completely* or *very*. Performance in these two between-subject conditions is compared to a third, baseline condition in which infants are presented with a bare, unmodified novel adjective. Infants in each condition

---

[85] I say that *very* is 'relatively free' in its modification, because, although *very* can appear with the vast majority of GAs, there are some restrictions on its modification. *Very*'s role as an intensifier means that it is 'standard raising', in the words of Kennedy and McNally (2005). (See also von Stechow (1984).) To say, for example, that someone is *very tall* is to say that they are 'tall and then some' (Barker, 2002). Thus, it should be barred from modifying maximum standard absolute GAs, although it is not. When maximum standard absolute GAs have a relative sense (e.g., *straight*, *dry*), or are used imprecisely (e.g., *empty*, *full*), then *very* can appear as a modifier. Still, some maximum standard absolute GAs resist modification by *very*.

are presented with the same visual stimuli, a series of objects which have both absolute and relative properties. We predict that if the distributional information – and consequently, selectional restrictions – is recruited by the language learner when assigning a meaning to a novel adjective, then infants who hear *completely* will map the novel adjective onto a meaning that corresponds with an absolute property, while infants in the *very* condition should map the novel adjective onto a meaning that corresponds with a relative property, and infants in the no-adverb condition should show no preference either way, unless a default preference for attention to certain properties drives their looking patterns.

6.2. Experiment 3

6.2.1. Method

6.2.1.1. Participants

6.2.1.1.1. Children

Thirty-three infants participated. Child participants (19 girls 14 boys) with a mean age of 29;9 months (range: 28;0 to 32;2 months) were recruited from families from College Park, MD, and the surrounding area. Contact information for the families was included in a database of potential child participants in the Infant Studies Consortium at the University of Maryland, College Park. Parents were contacted by phone, given a brief description of the study, and invited to bring their child into the laboratory. 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. In return for their participation in the study, children were compensated with a reward (usually a book). An additional 19 children were excluded because of: fussiness (n = 2); experimenter error (n = 3); inattentiveness for more than 20% of the time in three or more trials (n = 9); and age and gender balance across experimental conditions (n = 5). Children were randomly assigned to one of three experimental conditions. The composition of these three conditions is as follows: *completely* (6 girls 5 boys, M: 30;2; average vocabulary production: 483, st. dev: 186.5); *very* (6 girls 5 boys, M: 29;5; average vocab. production: 522, st. dev: 111.0); and 'no-adverb' (7 girls 4 boys, M: 29;9 average vocab. production: 539, st. dev: 109.0). Paired t-tests revealed no significant difference in vocabulary production among these three conditions.

6.2.1.1.2. Adults

30 adults (27 F 3 M) participated. Adult participants were Northwestern University undergraduates from the Department of Linguistics subject pool, who were fulfilling an experimental requirement for a Linguistics course. They were recruited online using the Experimetrix software. The data from three additional adults (2 F 1 M) were excluded, because the participants indicated that a language other than English was their native language.

6.2.1.2. Materials

This experiment employed the intermodal preferential-looking paradigm, originally pioneered by Spelke (1979) and further developed by Booth and Waxman (2003), Golinkoff, Hirsh-Pasek, Cauley, and Gordon (1987), Hollich, Rocroi, Hirsh-Pasek, and Golinkoff (1999), Naigles (1990), Waxman and Booth (2001), *inter alia*. In this paradigm, participants are shown two images on a video screen. During a familiarization phase, visual stimuli are presented on the screen and accompanied by auditory stimuli labeling the objects or events. A contrast phase then follows. Finally, during a test phase, two images related to those seen during familiarization are displayed simultaneously on either side of a video screen, and participants are directed to choose which of the two images better corresponds to the auditory stimulus. Infants' eye gazes towards one image or the other are recorded and taken as an indirect measurement of their choice.

6.2.1.2.1. Visual Stimuli

Images of objects were either identified through an online search of images or created using the Microsoft Office drawing function, and selected for their ability to display two key properties simultaneously. Some photographic images were further edited to best capture the relevant property. All images were in color. A complete list of stimuli is included in Table 29.

Table 29: Novel adjectives and visual stimuli used in Experiments 1 and 2

		experimental phase		
trial	adjective	familiarization	contrast	test
1	*pelgy* [pɛlgi]	2 orange objects, both long and straight	orange object that is short and curly	2 orange objects, 1 long & curly, 1 short & straight
2	*keetel* [kitəl]	2 windows, both high and closed	window that is low and open	2 windows, 1 high & open, 1 low & closed
3	*vickel* [vɪkəl]	2 toy blocks, both large and splotched	block that is small and solid	2 blocks, 1 large & solid, 1 small & splotched
4	*zaipin* [zepɪn]	2 balls, both wide and smooth	ball that is round and bumpy	2 balls, 1 wide & bumpy, 1 round & smooth
5	*wuggin* [wʌgɪn]	2 containers, both tall and transparent[86]	container that is short and opaque	2 containers, 1 tall & opaque, 1 short & transparent

[86] Note that *transparent*/*opaque* are unlike the other absolute properties in this experiment and are like *full*/*empty* in that they map onto two ends of the same scale, so both can be used interchangeably as the maximal or minimal terms. Both members of the pair are felicitously modified by *completely* and 100% (cf. Kennedy, 2007).

6.2.1.2.2. Auditory Stimuli

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. Three sets of each stimulus item were made. These were then transferred to a computer and 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 then burned onto a DVD and presented to participants on a computer at a rate of 30 frames per second. The novel adjectives corresponding to the properties in question are listed in Table 29 above. The entire set of auditory stimuli is described in more detail in the following section.

6.2.1.2.3. Trial design

The experimental session consisted of five distinct trials, each corresponding to two different properties and having three separate phases: familiarization, contrast, and test. The trial design was modeled after Waxman and Booth (2001). See Table 30 for a complete description of the auditory stimuli accompanying each trial phase, the times for each phase, and the visual stimuli for one novel adjective (*wuggin*).

Table 30: Example of a trial design with the novel adjective *wuggin*

	phase				
	pre-trial	familiarization	contrast (1)	contrast (2)	test
	4 sec	18 sec (6 s, 6 s, 6 s)	7 sec	7 sec	12 sec (4 s, 8 s)
properties	relative: tall absolute: transparent				
condition					
completely	Let's look at some things that are *wuggin*!	Look! These are both completely wuggin! This one is *completely wuggin*… And this one is *completely wuggin*!	Uh oh! This one is NOT *wuggin*!	Yay! This one is *COMPLETELY wuggin*!	Look! They're different! Which one is *wuggin*?
very	Let's look at some things that are *wuggin*!	Look! These are both *very wuggin*! This one is very *wuggin*… And this one is *very wuggin*!	Uh oh! This one is NOT *wuggin*!	Yay! This one is *VERY wuggin*!	Look! They're different! Which one is *wuggin*?
no adverb	Let's look at some things that are *wuggin*!	Look! These are both *wuggin*! This one is *wuggin*… And this one is *wuggin*!	Uh oh! This one is NOT *wuggin*!	Yay! This one IS *wuggin*!	Look! They're different! Which one is *wuggin*?

To capture participants' attention at the beginning of each trial, a still black-and-white photograph of a smiling infant appeared at the center of the video screen for four seconds and was accompanied by an audio track of an infant giggling. Following this image, the participant viewed a blank white screen and heard the female speaker say, "Let's look at some things that are A!" *A* was the corresponding novel adjective for each trial. This pre-trial presentation lasted four seconds and was intended to help participants segment the adverb and adjective separately from the speech stream in the *completely* and *very* conditions. Trials were presented in one of two orders, balanced across conditions and counterbalanced across subjects. Subjects were randomly assigned to one of the two orders and one of the three modifier conditions (*completely*, *very*, and no adverb). The visual presentation was consistent across participants; what varied was the auditory presentation.

Familiarization Phase (18 s)

During the familiarization phase, participants saw two objects, presented simultaneously, then one at a time on alternating sides of the screen. Both objects had a relative property (e.g., both were tall) and an absolute property (e.g., both were transparent). The accompanying video varied as a function of the condition. For example, in the *completely* condition, participants heard the objects described as *completely wuggin* (or whatever novel adjective corresponded to the trial). In the *very* condition, the objects were described as *very wuggin*. In the 'no-adverb' baseline condition, the objects were simply described as *wuggin*.

Contrast Phase (14 s)

Following familiarization, participants proceeded on to a contrast phase in which they were shown two objects, presented one after the other in succession in the center of the video screen. The first was a contrast object that instantiated the opposite properties of those seen in the familiarization phase (e.g., the object was short and opaque). In all three conditions, the speaker somewhat disappointedly indicated that this object was not, e.g., *wuggin*, using no adverbial modifier. This object helped participants narrow the range of possible discourse entities. Since neither of the two properties was seen during familiarization, explicitly labeling this object as *not wuggin* provided the participant with evidence that there are limits to what counts as wuggin. The second image was one of the two familiar objects seen previously. The choice of the familiar object was counterbalanced across trials. In the accompanying audio, the speaker cheerfully labeled the object as *completely wuggin*, *very wuggin*, or *wuggin*. The intonation was controlled so that contrastive focus (L+H* L-H% pitch accenting) was placed on the adverb, when present (e.g., "This one is COMPLETELY wuggin!") or on the copula when there was no adverb (e.g., "This one IS wuggin!"). This was done because in the first half of the contrast phase, the emphasis was on negation (e.g., "This one is NOT wuggin.").

Test Phase (12 s)

In the last phase of the trial, the test phase, participants were shown two objects simultaneously, one on either side of the video screen. Here, the two properties were teased apart, so that one object had the relative property from familiarization but not the absolute property (e.g., it was tall but not transparent), and the other had the absolute property from familiarization but not the relative property (e.g., it was transparent, but not tall). The test phase was divided into two distinct periods, a salience window and a response window. The first (*salience*) was designed to assess infants' baseline attention. In this four-second period of the test phase in each condition, the speaker said the same thing (e.g., "Look! They're different!"). The screen then went blank momentarily (0.33 seconds) before the second period of the test phase began. In this period (*response*), the two objects reappeared, and the speaker asked the participants to turn their attention to the object that fit the description (e.g., “Which one is wuggin?”). The audio was the same in each condition; no adverb was present, since the goal was to determine how the presence of the adverb earlier in the experiment constrained the reference for the novel adjective. The position of the anticipated match was counterbalanced across trials with respect to the left or right side of the screen and the side of the screen on which the familiar object for the second half of the contrast phase had first appeared.

6.2.1.3. Apparatus and Procedure

6.2.1.3.1. Children

Infants and their caregiver(s) were welcomed to our laboratory waiting area. While the infant played freely with toys and interacted with research assistants and the experimenter, the caregiver was given a brief description of the experimental procedure and asked to complete a consent form and the MacArthur Communicative Development Inventory: Words and Sentences. After some time (usually 10-15 minutes), once the child and caregiver had become comfortable with the environment, the experimenter escorted the infant and caregiver into an adjoining testing room measuring 14' x 7', which was quiet and dimly lit. The infant was tested individually, seated either in a highchair or on the caregiver's lap directly facing a television screen approximately 6' away. If seated in the highchair, the infant was accompanied by a caregiver who was seated behind the child (n = 21). If the child was seated on the caregiver’s lap, the caregiver wore a visor and was asked to refrain from talking or offering any form of encouragement while in the testing room with the child (n = 12). During the experimental session, if infants solicited their caregiver’s attention, the caregiver was permitted to direct the child’s attention back to the general direction of the screen.

Once the infant and caregiver were settled, the experimenter moved to a control room to start the experimental procedure. Stimuli were presented via a wall-mounted plasma television, 44" x 24.5". Visual stimuli were centered in the screen and filled a sizable portion of the screen. Objects ranged from 6" to14" in height and were 23.5" apart from center to center. Audio stimuli were presented from two speakers located along each side of the screen. The experimenter remained in the control room, out of view from the participants. Infants' looks to the 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 for later coding.

6.2.1.3.2. Adults

Adults viewed the same video as the children. After completing a consent form and a background information form (see Appendix J), they were invited into our video room in our laboratory, a 9' x 10.5' dimly-lit room. 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. At the start of each experimental session, the experimenter read from a script, preparing the adults for the experimental task. This script is included in Appendix K. A maximum of three (and usually one to two) adult participants were in the room at a time. Each participant was given small response packets to complete and a clipboard. Each packet had a cover sheet, where participant information was recorded. Each of the five remaining pages of the packet corresponded to one of the five trials, with “LEFT” and “RIGHT” noted in large, bold capital letters in the center of the page (see Appendix L). 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.

6.2.2. Analysis

6.2.2.1. Children

Videos of the infants' eye gazes were transferred as .mov files to a Macintosh computer and were then coded off-line by an experimenter using the SuperCoder software (Hollich, 2003). The sound was removed to ensure that the coder, who was blind to the experimental condition, was only coding the direction of visual fixation (left image, right image, or neither) during the test phase. Videos were coded frame by frame (30 frames per second, or 1/30 sec per frame). One experimenter was the primary coder for all of the infant videos. A second experimenter independently coded 5 of the videos across all conditions in both experiments. For the baseline and response windows, there was 96% agreement between coders.

For each infant, counting each frame across the five trials, the proportion of looks directed towards the object with the relative property during the test phase was calculated. This information was then averaged across infants in each condition to produce a detailed record of the time-course of infants' looking behavior throughout the entire test phase of the three conditions. Trials in which the infant was inattentive (i.e., looking at neither the left nor the right image) for 30% or more of the test phase were excluded. Infants who reached this criterion for three or more trials were replaced (n = 9).

Two windows of the test phase were then selected for the purpose of comparison and statistical analysis. The first, salience, provided a baseline of children's looking before the speaker asked them to direct their attention to an object. The second, response, provided a measure of the meaning infants assigned to the novel adjective during the experimental session. Each window was the same length, 1.5 seconds (45 frames). The salience window began 10 frames (approx. 300 ms) from the onset of "Look! They're different!" The response window began 10 frames (approx. 300 ms) from the onset of the adjective. Within each window, the relevant point of comparison was the proportion of attention to the object with the relative property out of the total attention to both objects. It is assumed that infants will look longer at the object whose salient property best matches the interpretation they have assigned to the novel adjective. Or, put another way, what they think the novel adjective means will guide their looks to one object or the other during the response window.

6.2.2.2. Adults

Adult responses from the forced-choice paper-and-pencil task were reviewed and entered by the experimenter. The relevant measurement was the percentage of the selection of the object with the relative property out of all selections. No responses were left blank.

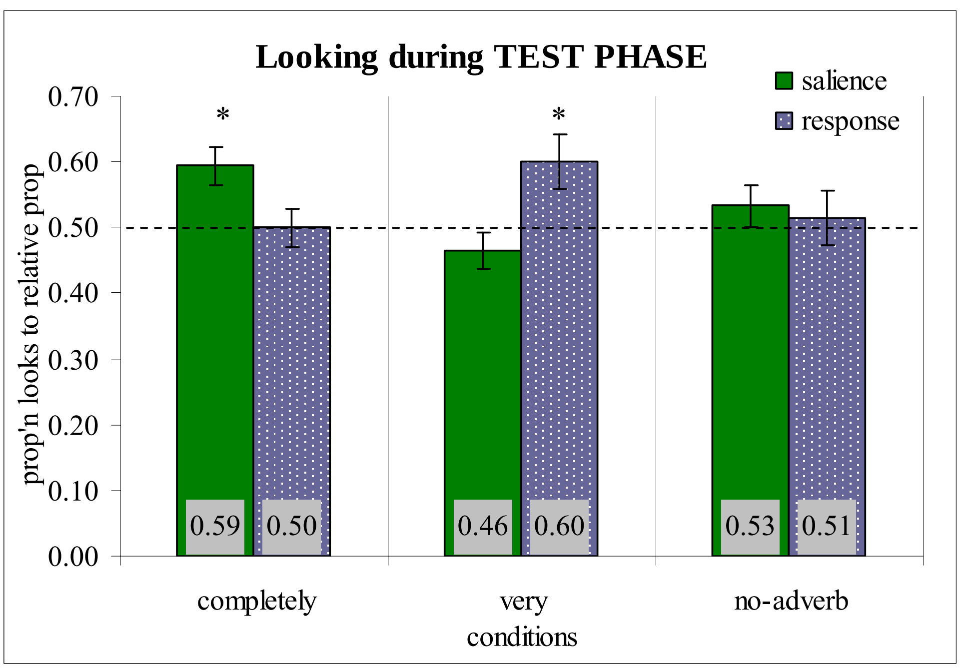
6.2.3. Results

Recall our predictions. Infants in the no-adverb condition should not pattern differently between the salience and response windows, unless there is a default preference to attend to one of the two object properties. If infants are guided by the selectional restrictions of the adverb when assigning a meaning to the novel adjective, infants in the *completely* condition will map the novel adjective onto a meaning that corresponds with an absolute property, since this adverb requires that the adjective it modifies corresponds to a closed scale. Infants in the *very* condition may pattern with the infants in the no-adverb condition, if this adverb can freely appear with any adjective. If infants are sensitive to distributional information from the input and its standard-raising function, however, they should be inclined to map the novel adjective modified by *very* onto an open scale and assign it a relative interpretation.

6.2.3.1. Children

The pattern of children's looking behavior during the salience and response windows of the test phase for all three conditions is presented in Figure 19. It is immediately apparent from this figure that there is a difference between the two adverbial conditions – both from each other and from the no-adverb baseline condition. While infants in the *completely* condition begin by looking toward the object with the relative property during salience (e.g., the tall, opaque object in the *wuggin* trial), they end by looking away from this object during the response window. Conversely, infants in the *very* condition begin by looking away from this object and end by looking toward it. Infants in the no-adverb condition show no difference in looking behavior between the two windows. These observable differences are supported by statistical analyses.

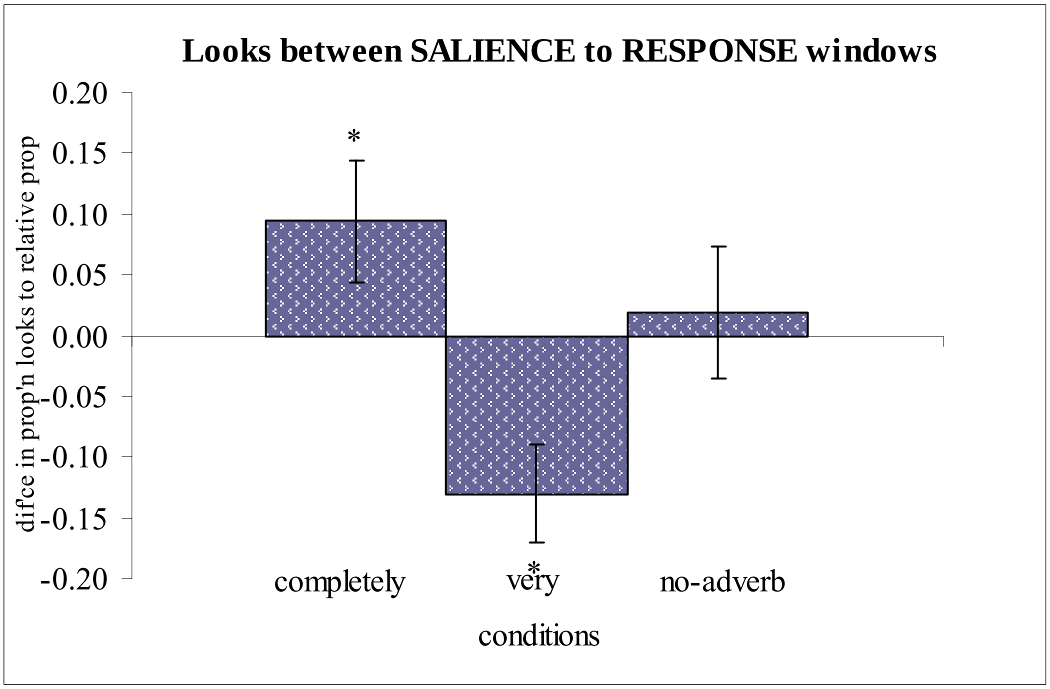
Figure 19: Proportion of looking time to the object with the relative property by infants during two test windows



A 2 x 3 ANOVA was conducted, with the three adverbial conditions as between-subject factors and the salience and response windows as within-subject factors. This analysis revealed a main effect of salience/response window $F(1, 33) = 6.196$, $p < 0.02$, a marginally significant interaction $F(2, 30) = 2.653$, $p < 0.08$, but no main effect of condition $F(2, 32) = 0.272$, $p = 0.763$. To investigate the source of this pattern, two sets of t-tests were then conducted on the data. First, one-tailed t-tests compared the difference between the salience and response windows for participants in each of the three conditions. The degree of freedom for each test was 10. This analysis revealed a statistically significant difference between the two windows for both of the adverbial conditions (*completely*: $t = 1.9$, $p = 0.043$; *very*: $t = -3.25$, $p = 0.0044$), but not for the no-adverb condition ($t = 0.34$, $p = 0.3692$).

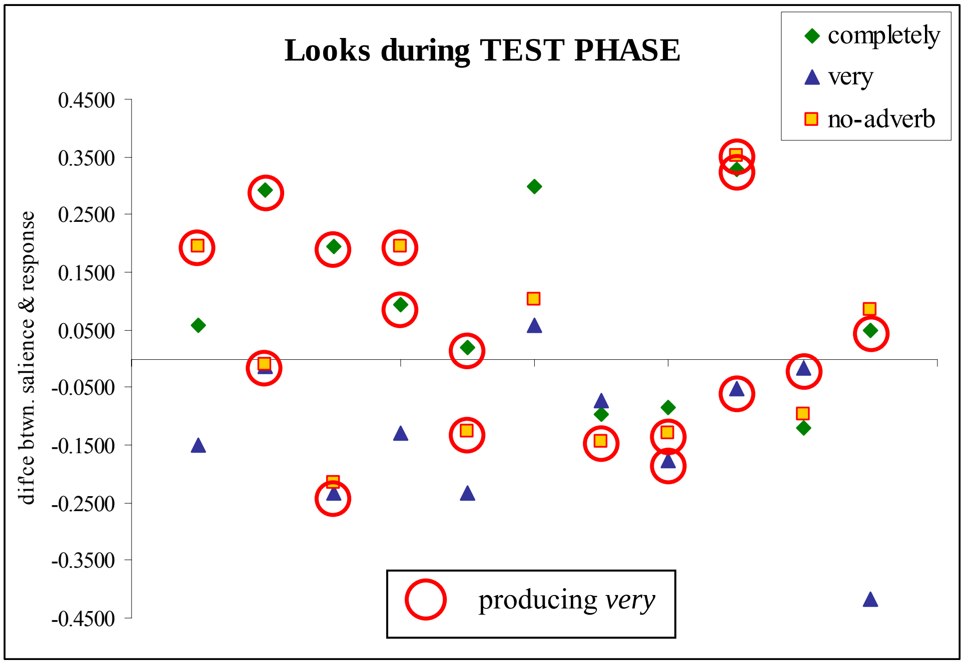
A second set of two-tailed t-tests analyzed the difference from chance-level performance for participants in each of the three conditions and in each of the two test windows. The degree of freedom for each test was 10. In the salience window only infants in the *completely* condition behaved significantly different from chance: *completely*: $t = 3.47$, $p = 0.006$; *very*: $t = -0.772$, $p = 0.458$; no-adverb: $t = 1.028$, $p = 0.3282$. By contrast, during the response window, only infants in the *very* condition behaved significantly different from chance: *completely*: $t = -0.012$, $p = 0.9907$; *very*: $t = 2.93$, $p = 0.015$; no-adverb: $t = 0.34$, $p = 0.7409$. 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*) appears to pull the attention toward the object with the relative property during the response window, the other adverb (*completely*) pulls attention away from it. This difference is seen even more clearly by analyzing the proportion of looks in terms of the difference between the two windows. See Figure 20.

Figure 20: Difference in proportion of infants' looks to object with relative property between two test windows



Independent two-tailed t-tests were conducted to determine whether the difference between the salience and response windows was significant between any two of the three conditions. The degree of freedom for each test was 20. Unsurprisingly, given the pattern shown in Figure 19 and the results of previous t-tests, the difference between the *completely* and *very* conditions easily reached significance ($t = 3.52$, $p = 0.0022$), the difference between the no-adverb and *very* conditions was significant ($t = 2.21$, $p = 0.0389$), and the difference between the no-adverb and *completely* conditions was not significant ($t = 1.03$, $p = 0.3153$). An evaluation of individual differences reveals that for eight of the eleven children in the *completely* condition, the difference was positive, while for ten of the eleven children in the *very* condition, it was negative. For children in the no-adverb condition, six displayed a negative difference and five a positive difference. Differences for individual children are plotted in Figure 21.

Figure 21: Difference between two test phase windows for individual children



Thus, most of the children in the *completely* condition began by looking at the object with the relative property and switched when asked to make a choice between the objects based on the meaning they assigned to the novel adjective, whereas all but one of the children in the *very* condition began by looking at the object with the absolute property and switched to the object with the relative property in the second window, while children in the no-adverb condition were divided.

These difference scores were also used to compare child responses based on linguistic production. A two-tailed t-test analysis of vocabulary production as measured by parents' responses on the MacArthur Inventory did not reveal any difference between infants whose production was above the mean (or above 500) and those whose production was below ($t(31) = 1.29$, $p = 0.2069$). This held for each of the conditions (*completely*: $t(9) = 0.17$, $p = 0.8688$; *very*: $t(8) = 1.58$, $p = 0.1528$; no-adverb: $t(9) = 0.93$, $p = 0.3766$).

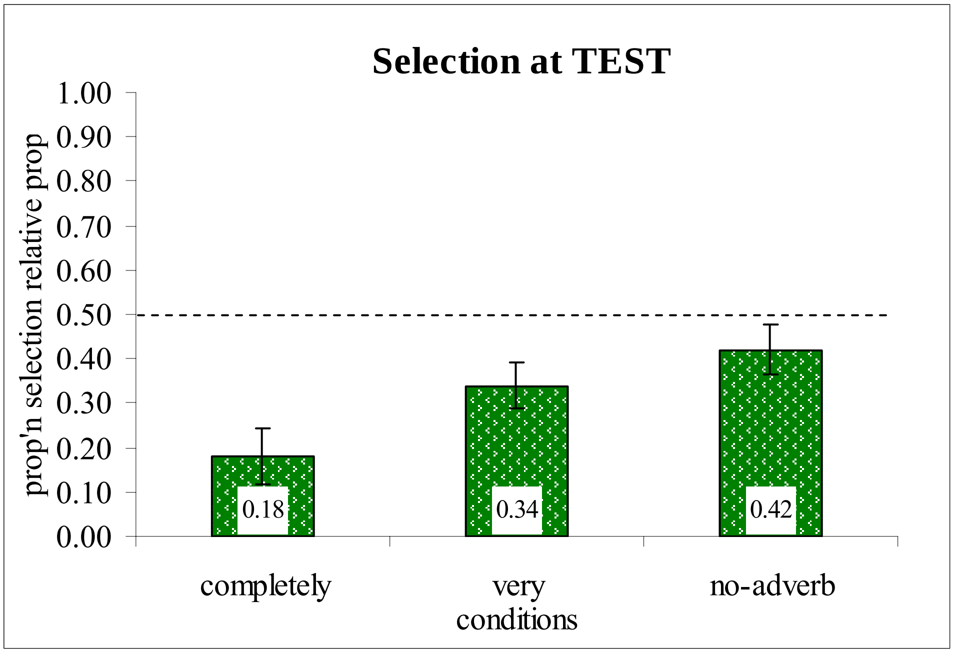
However, parents were given a supplementary set of questions and asked if their child was producing *very* or *completely*. 16 children across conditions were producing *very*; these children are indicated with circles in Figure 21. Only one child was reported to be producing both adverbs. No child was reported as producing *completely* but not *very*. The difference between infants who produce *very* and those who produce neither adverb is not significant in a two-tailed t-test ($t(30) = 1.33$, $p = 0.1935$, but marginally significant in a one-tailed t-test ($p = 0.0968$). This difference stems from the fact that children who produce *very* on average exhibit a positive difference (mean = 0.03) while those producing neither adverb generally exhibit a negative difference (mean = -0.06).

Closer inspection of the data reveals that the trend towards an effect is driven by marginal statistical significance in the *completely* condition (one-tailed t-tests: *completely*: $t(9) = 1.66$, $p = 0.0656$; *very*: $t(8) = 0.42$, $p = 0.3428$; no-adverb: $t(9) = 0.68$, $p = 0.2568$): the six children in the *completely* condition who are reported to be producing *very* exhibit a positive difference between the salience and response windows. Thus six of the eight children in the *completely* condition who show this positive difference are producing *very*. These results may suggest that although these children are not producing *completely*, they may already comprehend it, since they are already producing *very* and therefore know about its selectional restrictions. However, further investigation would need to be conducted to provide support for this idea.

6.2.3.2. Adults

Adults generally resisted selecting the object with the relative property (a point that we will re-visit), but like the children, they were influenced by the presence of an adverbial modifier. Presented with a bare novel adjective, their selection of an object at test was no different from guessing. This pattern is captured in Figure 22.

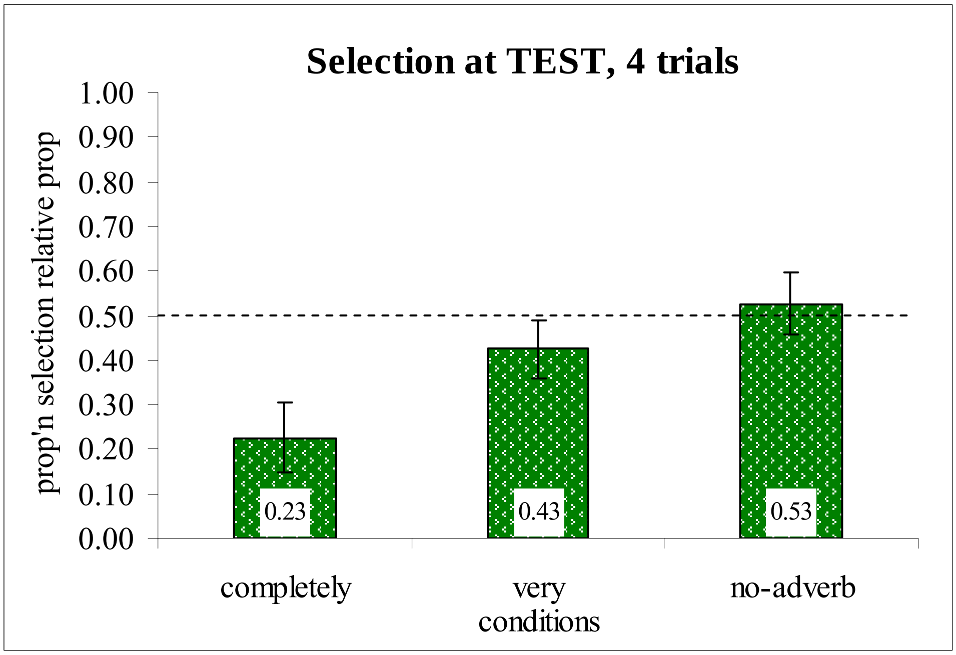
Figure 22: Percentage selection of object with relative property by adults



Two sets of two-tailed independent t-tests were conducted. The first analyzed the difference of each condition from chance. The second analyzed the difference between any two of the three conditions from each other. Both adverbial conditions, but not the no-adverb condition, were significantly different from chance (df = 9; *completely*: t = -5.079, p = 0.0007; *very*: t = -3.077, p = 0.0132; no-adverb: t = -1.455, p = 0.1796). The only pair of conditions that differs significantly from each other, though, is no-adverb/*completely* (df = 18 for all three comparisons, t = 2.86, p = 0.0104). The *very* condition is not significantly different from either of the other conditions (v. *completely*: t = -1.7, p = 0.1063; v. no-adverb: t = -1.05, p = 0.3076).

At the beginning of this section, it was noted that adults were generally disinclined to select the object with the relative property. For one trial, in fact, adults never selected this object, no matter what condition they were in. The novel adjective in this trial, *pelgy*, had the potential to map to either length or straightness. It seems that adults did not attend to the absolute length of the object at all, and only noticed the straightness or curliness of the object. Excluding this data point from the analysis, we are left with a similar picture. See Figure 23.

Figure 23: Percentage selection of object with relative property by adults in 4 of 5 trials



With the *pelgy* trial excluded from each participant's experimental data, the no-adverb condition is still no different from chance level (t = -.362, p = 0.7257). However, now the *very* condition is also no different from chance level (t = 1.154, p = 0.2782), while the *completely* condition is well below (t = 3.481, p = 0.0069). For children as well as adults, then, the role of *completely* was to pull participants' attention away from the object with the relative property.

6.2.4. Discussion

The purpose of Experiment 3 was to determine whether infants make use of adverbial modification when assigning a meaning to a novel adjective. The results demonstrated not only that the presence of an adverb matters, but that the kind of adverb matters. While infants in the no-adverb baseline condition showed no difference in performance between the salience and response windows during the experiment's test phase, infants in both of the other two adverb conditions pulled away from this baseline, in opposite directions, as predicted.

Infants in the *completely* condition began by looking at the object with the relative property during the salience window, and attended more to the object with an absolute property during the response window. This pattern appears to reflect an initial novelty preference, and then word extension constrained by the adverb's selectional restrictions to modify a property and an adjective that map to a closed scale. Specifically, infants who know about the selectional restrictions of *completely* understand that any adjective it modifies and the corresponding property must be compatible with a maximal endpoint. When they heard a novel adjective modified by *completely* during the initial phases of the trial, they assigned it a maximal interpretation. When the object properties were teased apart at test, infants were surprised to see the relative property disassociated from the absolute one, since a relative interpretation is inconsistent with the interpretation they assigned the adjective. Thus, they displayed a novelty preference specific to the meaning of *completely* during the salience window, then switched their attention to the object with the absolute property when asked about the interpretation they assigned to the novel adjective.

By contrast, infants in the *very* condition began by looking at both objects at chance during the salience window, but spent more time looking at the object with the relative property during the response window. Again, this pattern is expected if infants are sensitive to both the meaning of the adverb as an intensifier and its distribution in the input. In the analysis of corpus data in the previous chapter, we saw that maximal GAs are more likely to be modified by restricted adverbs than non-maximal GAs are, and that non-maximal GAs are much more likely to be modified by adverbs such as *very*. *Very* is also highly frequent, and the vast majority of its adjective-modifying occurrences are instances where *very* modifies an adjective with a scale that is not maximally closed. In the spoken BNC, *very* has a frequency of 25,041 (or 2421. 36 instances per million words), while *completely* has a frequency of only 822 (79.48 instances per million words). In addition, *very* is much less likely to modify a maximal GA: maximal GAs account for only 5.6% of all adjective-modifying instances of *very*, while they account for 45.0% of the adjective-modifying instances for *completely*. Thus it makes sense that although *very* can modify a wide range of adjectives, infants in the *very* condition who were sensitive to its distribution in the exposure language would have assigned a non-maximal (relative) interpretation to the novel adjectives in this experiment.

These differences among the three conditions indicate that the restrictional and distributional characteristics of adverbs may be recruited by the language learner acquiring adjectives. However, these results' affirmative answer to the question of whether infants use adverbs to learn about adjectives also raises a number of questions about the exact nature of this process. Most importantly, what, precisely, is the role of the adverb? Is it the *meaning* of the adverb that matters, or something else much more basic, such as relative frequency? Experiment 4 was designed to answer these questions.

6.3. Experiment 4

The purpose of Experiment 4 was to follow up on the results of Experiment 3 and determine whether it is the *meaning* of *completely* that drew infants' attention away from the object with the relative property, or whether there is another reason for this effect, such as relative frequency or novelty preference. Our null hypothesis is that the meanings of the adverbs in Experiment 3 were recruited by the infants when assigning an interpretation to the novel adjective. The alternative hypothesis is that there was another factor at work leading infants in the two different adverb conditions to pattern differently from each other. Let's consider how a factor other than the meaning of the adverbs could account for the pattern of results seen in Experiment 3.

In the Discussion section for Experiment 3, I noted that *very* is highly frequent, but *completely* is not. (Indeed, this is an understatement: *very* is approximately 30 times more frequent than *completely* in the spoken BNC.) The fact that *very* is more frequent in the input is mirrored in child production. Only one child from the previous experiment was reported to be producing *completely*, while 16 of the 33 infants were reported to be producing *very*. Could the effects observed in Experiment 3, then, have been driven by the relative frequency of these adverbs? It is possible that infants at the age of those tested in these experiments do not know the meaning of *completely* but are still influenced by the mere presence of an adverb. When they encounter a novel adverbial modifier – one which is relatively infrequent compared to a modifier such as *very*, whose meaning they know – their attention may be pulled away from the relative property (which may be the default when an adverb is present) and towards the absolute property. Experiment 4 was designed to probe this possibility.

In this experiment, two new conditions are introduced. The first condition involves the use of *extremely*, an adverb which has in common with *very* a meaning of intensification and generally non-restricted modification, but which has in common with *completely* a very low frequency. In the spoken BNC, *extremely* has a frequency of only 480 (46.41 instances per million words), and maximal GAs only account for only 2.0% of all adjective-modifying instances of *extremely*. The second condition involves the use of a novel adverb *penticly*. This novel word has a phonological structure similar to that of *completely*: they have the same number of syllables and similar pronunciation (i.e., consonants).

The reason behind the choice of these two conditions is the following. If the pattern of responses we observed in the *completely* condition in Experiment 3 can be attributed to a sheer novelty effect (i.e., responses are triggered by exposure to a novel adverb that draws attention away from a baseline preference), then we should expect to see similarities between responses in the *completely* condition in Experiment 3 and the two new conditions in Experiment 4. *Extremely* is also very infrequent, and *penticly*, having 0 frequency in the input, should elicit the 'ultimate' novelty effect. If, however, the pattern of responses we observed in infants in the *completely* condition was due to its *meaning*, we should see a difference between the *completely* condition and both of these two new conditions, since neither shares the same selectional restrictions as *completely*. Infants in the *penticly* condition should perform at chance, along with the infants that were in the no-adverb condition. Infants in the *extremely* condition may perform the same if they are unaware of the meaning of *extremely*, or pattern in a manner similar to the infants in the *very* condition, if they recognize its meaning.

6.3.1. Method

6.3.1.1. Participants

6.3.1.1.1. Children

Twenty-two infants participated. Child participants (11 girls 11 boys) with a mean age of 29;8 months (range: 28;2 to 31;3 months) were recruited from families from both the North Shore and greater Chicagoland area, and from College Park MD, and the surrounding area. As in Experiment 3, contact information for the families was included as part of a database of potential child participants in the Project on Child Development at Northwestern University and the Infant Studies Consortium at the University of Maryland, College Park. Parents were contacted by phone, given a brief description of the study, and invited to bring their child into the laboratory.

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. In return for their participation in the study, children were compensated with a reward (a book or small toy). An additional 12 children were excluded because of: fussiness (n = 4); equipment error (n = 2); and inattentiveness for more than 20% of the time in three or more trials (n = 6). Children were randomly assigned to one of two experimental conditions. The composition of these three conditions is as follows: *penticly* (5 girls 6 boys, M: 29;8; average vocab. production: 439); and *extremely* (6 girls 5 boys, M: 30;0; average vocab. production: 556).

6.3.1.1.2. Adults

20 adults (16 F 4 M) participated. Adult participants were Northwestern University undergraduates from the Department of Linguistics subject pool fulfilling an experimental requirement for a Linguistics course. All participants were native speakers of English.

6.3.1.2. Materials

The experimental methodology was the same as in Experiment 3, a split-screen intermodal preferential looking paradigm with five trials, each with three phases. The visual stimuli from Experiment 3 were used. The auditory stimuli were the same in structure as Experiment 3, but re-recorded by another female speaker of American English using the same guidelines. The difference between the auditory stimuli in Experiments 1 and 2 was the adverbs. In this experiment, in one condition, all novel adjectives were modified by *extremely*; in another, the novel adjectives were modified by the novel adverb *penticly*.[87] The text for these two new conditions is presented in Table 29.

Table 31: Auditory stimuli used in Experiment 4

condition	pre-trial	familiarization	contrast (1)	contrast (2)	test
extremely	Let's look at some things that are *wuggin*!	Look! These are both *extremely wuggin*! This one is *extremely wuggin*… And this one is *extremely wuggin*!	Uh oh! This one is NOT *wuggin*!	Yay! This one is *EXTREMELY wuggin*!	Look! They're different! Which one is *wuggin*?
penticly	Let's look at some things that are *wuggin*!	Look! These are both *penticly wuggin*! This one is *penticly wuggin*… And this one is *penticly wuggin*!	Uh oh! This one is NOT *wuggin*!	Yay! This one is *PENTICLY wuggin*!	Look! They're different! Which one is *wuggin*?

6.3.1.3. Apparatus and Procedure

6.3.1.3.1. Children

For infant participants run at the University of Maryland, College Park, the apparatus and procedure was the same as in Experiment 3. For participants run at Northwestern University, the details were only slightly different. Infants and their caregiver(s) were welcomed to the

[87] There were two other minor differences in the auditory stimuli. The novel adjective *zaipin* was pronounced [zepɪn] in Experiment 1, but in Experiment 2 was pronounced [zɑipɪn]. This was not predicted to lead to any differences for that trial. The quality of the recordings was also somewhat degraded, because of technical issues with the recording equipment. However, experimenters were positive that this did not contribute to any distraction or confusion on the part of the participants.

laboratory waiting area. While the infant played freely with toys and interacted with research assistants and the experimenter, the caregiver was given a brief description of the experimental procedure and asked to complete a consent form and the MacArthur Communicative Development Inventory: Words and Sentences. The parental responses from the MacArthur vocabulary inventories were later tabulated and reviewed by an experimenter.

After some time (usually 15-20 minutes), once the child and caregiver had become comfortable with the environment, the experimenter escorted the infant and caregiver into an adjoining testing room measuring 14' x 10' that was quiet and dimly lit. The infant was tested individually, seated in an infant-seat directly facing a projection screen approximately 6' away. In most cases, the infant was accompanied by a caregiver who was seated either directly behind the child or behind and to the left of the child. The caregiver wore a visor and was asked to refrain from talking or offering any form of encouragement while in the testing room with the child. During the experimental session, if infants solicited their caregiver’s attention, the caregiver attempted to direct the child’s attention back to the general direction of the screen. Data from children whose caregiver did not adhere to this request would have been excluded from the results, but we did not encounter this situation.

Once the infant and caregiver were settled, the experimenter moved behind the screen to control the experimental procedure. Stimuli were presented via a ceiling-mounted projector onto a 4.5' by 4.5' projection screen. Visual stimuli were centered in the screen and as in Experiment 3 filled a sizable portion of the screen. Audio stimuli were presented from two speakers located directly below the screen. The experimenter remained behind a dark curtain behind the screen, blocking any apparatus from the view of the participants. Infants’ looks to the stimuli were recorded with a Sony Digital8 Handycam centered inconspicuously above the screen. These videos were captured digitally onto a laptop computer using iMovie for coding later.

6.3.1.3.2. Adults

The procedure for the adults was identical to the procedure in Experiment 3.

6.3.2. Analysis

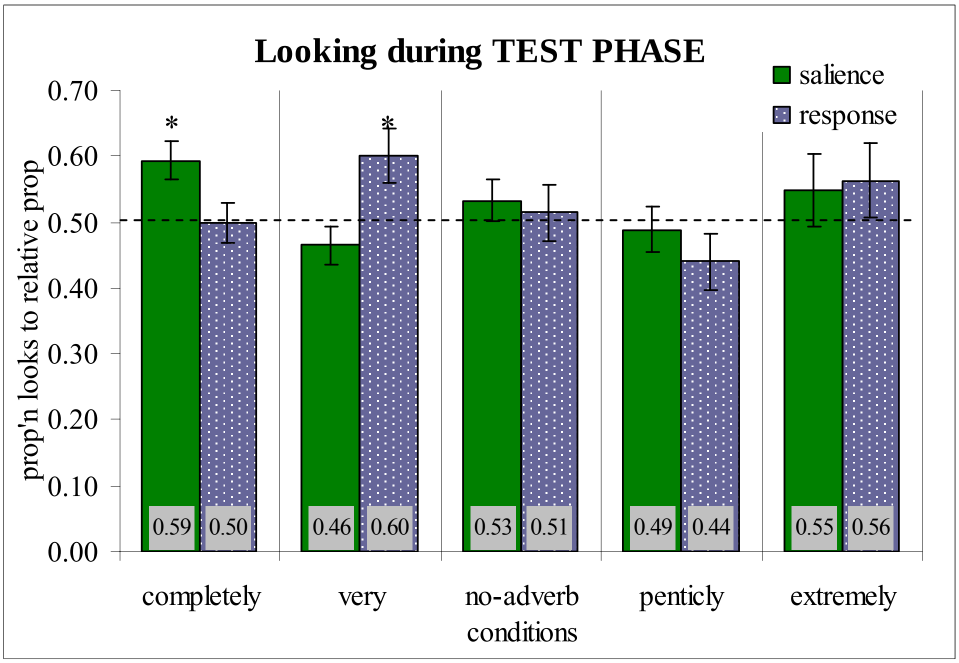
The method of analysis for both infants and adults was the same as the analysis in Experiment 3.

6.3.3. Results

6.3.3.1. Children

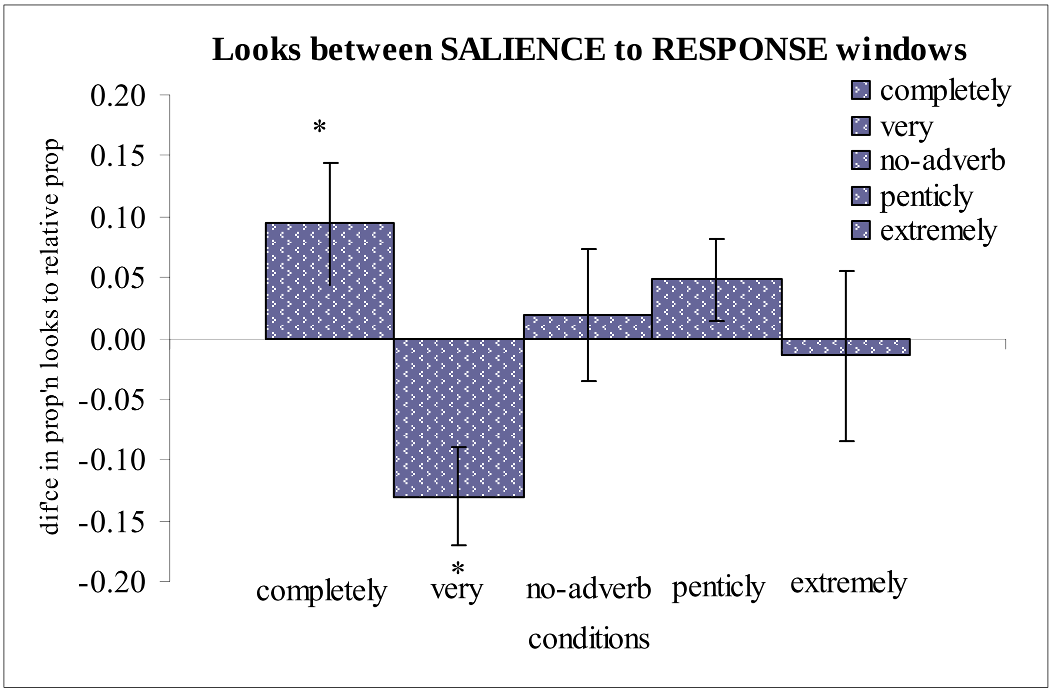
Infants' looking patterns in the salience and response windows in all five conditions across the two experiments are captured in Figure 24. This figure highlights the fact that the infants in the *completely* and the *very* conditions from Experiment 3 were the only infants who demonstrated a difference in looking between the two test phase windows. All other infants from the three other conditions patterned at chance level in both windows of the test phase.

Figure 24: Proportion of looking time to the object with the relative property by infants during two test windows



This pattern is reinforced by examining the difference between the two windows for each condition, as in Figure 25.

Figure 25: Difference in proportion of infants' looks to object with relative property between two test windows in all conditions



A 2 x 2 ANOVA was conducted, with the two adverbial conditions as between-subject factors and the salience and response windows as within-subject factors. This analysis revealed no within-subjects main effect $F(1, 21) = 0$, $p = 1.0$, no interaction $F(1, 21) = 0.5$, $p = 0.4877$, and only a marginally-significant interaction $F(1, 21) = 3$, $p = 0.0986$. Two-tailed t-tests support these results. The first set of t-tests compared the difference between the salience and response windows for participants in both conditions. The degree of freedom for each test was 10. In neither condition did infants show a difference in looking pattern between the two windows (*extremely*: $t = -0.21$, $p = 0.8379$; *penticly*: $t = 1.43$, $p = 0.1832$ (two-tailed), 0.0916 (one-tailed), marginally significant).

As in Experiment 3, a second set of two-tailed t-tests determined whether responses in either of these conditions were different from chance. In neither condition and in neither window were responses different from chance. This is the same pattern we observed with infants in the no-adverb condition in Experiment 3, but not the same pattern as the two adverbial conditions from Experiment 3. In the salience window, the results were as follows: *extremely*: t = 0.871, p = 0.4042; *penticly*: t = -0.329, p = 0.7489. In the response window, the results were as follows: *extremely*: t = 1.13, p = 0.2849; *penticly*: t = -1.386, p = 0.1959.

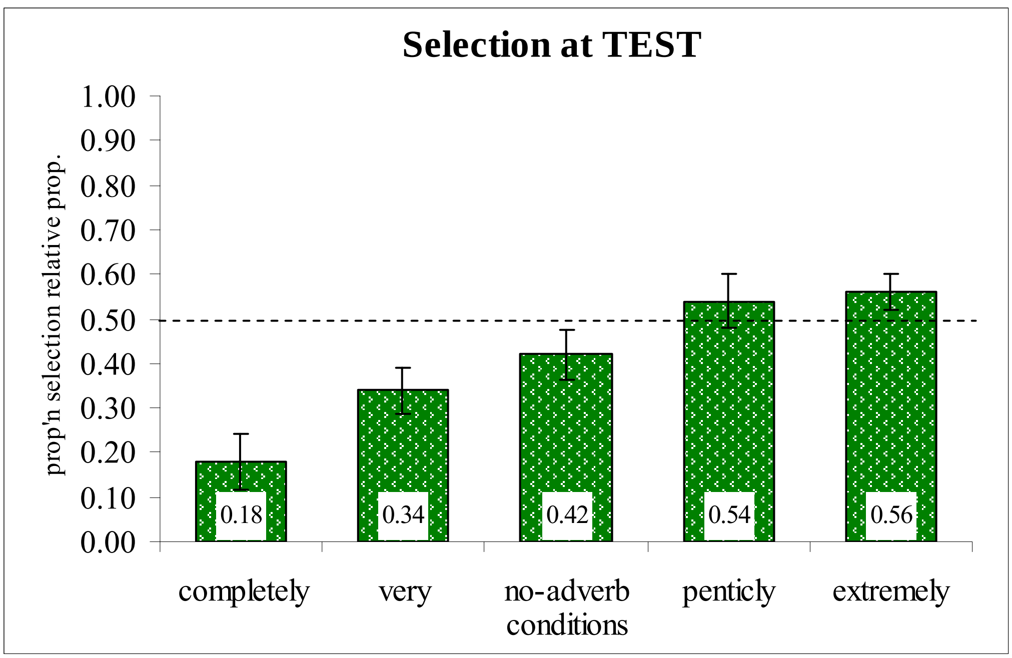
To compare performance across the conditions in the two experiments, an additional series of two-tailed t-tests (df = 20) were conducted between pairs of conditions, evaluating the difference between the salience and response windows. There was no difference between the two conditions in Experiment 4 (*extremely*/*penticly*: t = -0.81, p = 0.4275) or between these two conditions and the no-adverb condition (no-adverb/*extremely*: t = 0.38, p = 0.708; no-adverb/*penticly*: t = 0.46, p = 0.6505). While two-tailed comparisons with the *completely* condition indicate no significant difference, one-tailed comparisons are marginally significant (*completely*/*extremely*: t = 1.66, p two-tailed = 0.1125, p one-tailed = 0.0563; *completely*/*penticly*: t = 1.33, p two-tailed = 0.1985, p one-tailed = 0.0992).

Finally, comparisons between each of the two conditions and the *very* condition revealed no difference between the *very* and *extremely* conditions (*very*/*extremely*: t =-1.45, p = 0.1626) but a highly significant difference between the *very* and *penticly* conditions (*very*/*penticly*: t = 3.4, p = 0.0028). This last result is not surprising, since in the *penticly* condition, the difference between the two windows was only marginally significant, with the salience window not differing from chance (as in the *very* condition). The response window, though, was marginally significant from chance in the opposite direction from the response window in the *very* condition. This difference is supported by a two-tailed t-test (response windows: t(20) = 2.94, p = 0.004; salience windows: t(20) = -.25, p = 0.8051).

6.3.3.2. Adults

Curiously, for adults, the rate of selection of the object with the relative property rose for both of these two conditions in comparison to the three earlier conditions. This pattern is illustrated in Figure 26. With this increase in selection of the object with the relative property, adults demonstrated that these two adverbial modifiers are not like *completely*, which drew their attention to the object with the absolute property.

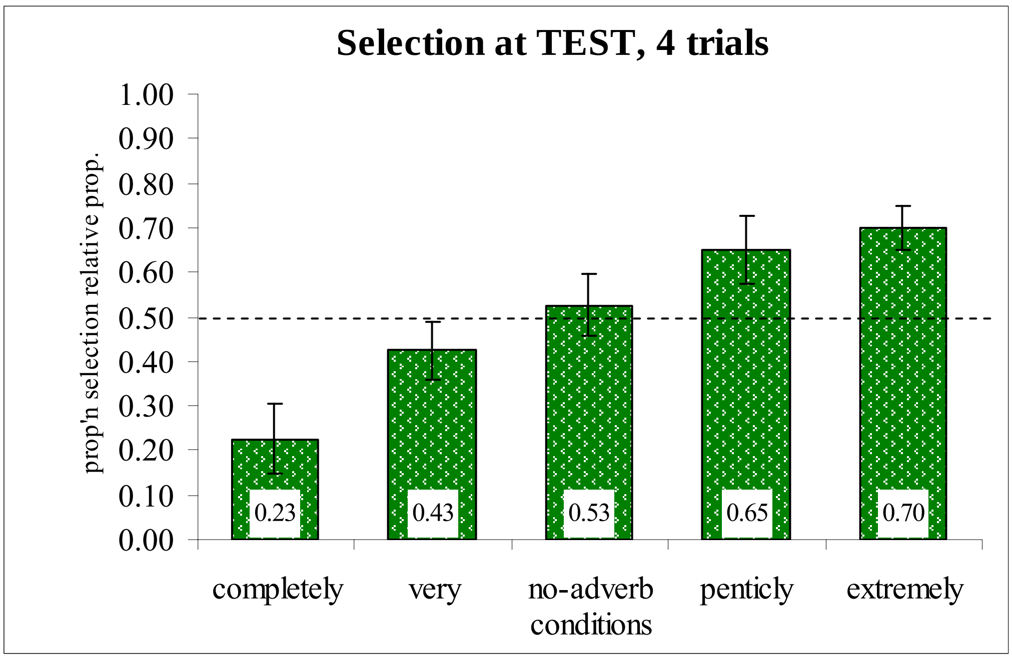
Figure 26: Percentage selection of object with relative property by adults



Independent two-tailed t-tests revealed a highly significant difference between both of the two conditions and the *completely* condition (*extremely* v. *completely*: df = 18 for all tests, $t = 5.1$, $p < 0.0001$; *penticly* v. *completely*: $t = 4.14$, $p < 0.0001$). The difference between the *extremely* and no-adverb conditions is only marginally significant ($t = 2.05$, $p = 0.0552$), and there is no difference between the *penticly* and no-adverb conditions ($t = 1.47$, $p = 0.1588$). In addition, neither condition differs significantly from chance (*extremely*: t = 0.667, p = 0.5215; *penticly*: t = 0.667, p = 0. 5215).

As in Experiment 3, participants were reluctant to allow *pelgy* to map to the relative property. Only one participant in the *penticly* condition responded that the long, curly object at test was the *pelgy* one (versus the short, straight one). An additional analysis was conducted, excluding this data point. The revised percentages are presented in Figure 23.

Figure 27: Percentage selection of object with relative property by adults in 4 of 5 trials



The *extremely* condition is now significant from chance (df = 9 for all tests, t = -4, p = 0.0031), but the difference between the *penticly* condition and chance is only marginally significant (t = -1.974, p = 0. 0798). The same pattern holds for differences between both conditions and the no-adverb condition (*extremely* v. no-adverb: t = 2.05, p = 0.0552; *penticly* v. no-adverb: t = 1.21, p = 0.2419). However, both conditions are significantly different from the *very* condition (*extremely* v. *very*: t = 3.35, p = 0.0036; *penticly* v. *very*: t = 2.24, p = 0.0379) and therefore *mutatis mutandis* for the *completely* condition. Adult participants' reluctance to map a novel adjective onto a relative interpretation in both the 'no-adverb' and novel adverb conditions may reflect the fact that they are driven by 'interpretative economy' (Kennedy, 2007) and minimize the role of context in assigning an interpretation until they encounter a reason for it. Thus, when given a choice between a relative GA and absolute GA interpretation for the adjective, they opt for the latter, since the standard of comparison does not depend on the context. This intuition would need to be followed up with further research. For now I leave it as an open possibility.

6.3.4. Discussion

The results from Experiment 4 elucidate the results from Experiment 3. Recall that the children in the first experiment who were assigned to an adverbial condition (*completely* or *very*) patterned differently from those assigned to the no-adverb condition. While infants who did not hear an adverbial modifier showed no difference between the salience and response segments of the test phase, infants in the *completely* condition looked more at the object with the relative property during the salience window and switched their looks to the object with the absolute property during the response window. By contrast, infants in the *very* condition began with a looking pattern at chance level, then switched rather dramatically to the object with the relative property during the response window.

In the second experiment I sought to determine precisely what the role of the adverb is. I introduced two new conditions, one with a novel adverb (*penticly*) and one with *extremely*, an adverb that was matched in meaning with *very* but more or less matched in frequency with *completely*. These follow-up conditions allowed me to determine whether the novelty of a low-frequency adverb such as *completely* was sufficient to direct children's attention away from the relative property. This hypothesis was paired with the observation from the corpus data presented in Chapter 5 that *very* appears significantly more often with non-maximal adjectives in the input and is therefore more likely to modify the adjective corresponding to the relative property in this set of experiments.

I predicted that infants in these two new conditions would pattern similarly with infants in the *completely* condition and differently from those in the *very* condition if they were simply being influenced by a novelty effect. If, however, children in the earlier adverbial conditions were responding to the meaning of these adverbs, then the pattern of results would be different. Children in the *penticly* condition would pattern with the children in the earlier no-adverb condition. For children in the *extremely* condition, there were two possibilities: either they would also pattern at chance with the children in the no-adverb condition, since *extremely* is a low-frequency adverb, or they would respond in a similar manner to the children in the *very* condition, since these words are both intensifiers. In this second set of experiments, infants in both conditions patterned at chance, indicating that in the first experiment, they were indeed being guided by the meaning of *completely* and *very* and were not just influenced by novelty.

6.4. General Discussion

Taken together, the results of Experiments 1 and 2 demonstrate that children as young as 30 months of age make use of adverbial modifiers when learning the meaning of adjectives. Their choice of which object property to attend to during the test phase of the trials appears to have been guided by their knowledge of the selectional restrictions of the adverbs. That is, given that infants were presented with novel adjectives in these experiments, their 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. This information is reflected in distributional differences in the exposure language: a restricted adverb such as *completely* selects for an adjective that allows for reference to a maximal endpoint and is significantly more likely to appear modifying such adjectives than non-maximal ones, and an adverb such as *very*, though able to modify a much wider variety of adjectives, is significantly more likely to appear modifying a non-maximal adjective. Even infants who are not yet producing these adverbs appear to have attended to their appearance in the input.

Despite the relatively low frequency with which *completely* appears in the input, the correlation between this lexical item and the adjectives that appear in bigrams with it is robust enough to be informative. *Penticly* is a novel adverb with 0 frequency in the input and no meaning associated with it, so it is not surprising that infants are not guided in either direction by this adverb. The results from the *extremely* condition leave room for speculation, however, since here, infants were also at chance, as they were in the *penticly* and no-adverb conditions. Recall that *extremely* has an intensifier (or standard-raising) meaning similar to that of *very*, but unlike *very*, it is extremely infrequent. In fact, it is half as frequent as *completely* (25041 vs. 822 vs. 480, respectively, in the spoken BNC). However, even with this relatively small number of occurrences, it is overwhelmingly more likely to modify a non-maximal adjective (390 vs. 10 instances modifying an adjective). So why did infants in this condition not pattern with those in the *very* condition? There are two possibilities.

The first is a simple one: *extremely* is too infrequent for infants to have generalized anything about its meaning by tracking its occurrence, and they may therefore treat it basically as a novel adverb. The second is pragmatic in nature. Infants may have deduced something about the meaning of *extremely* from its distribution in the input, but given its relatively low frequency, this generalization may still be fragile. When they hear it in the experiment, they might be confused: its meaning overlaps with *very*, but *very* was not the modifier they heard (and might have expected to hear modifying the adjective, given its high frequency). So although they might have been inclined to associate *extremely* with the relative property, they instead look more to the object with the absolute property during the response window. Further research is necessary to decide this matter.

Chapter 7: Bringing It All Together: Conclusions and a Proposal

7.1. Taking Stock: Preliminary Conclusions

Before proceeding with this chapter, let me briefly review the content of the five previous chapters. In Chapter 2, I reviewed previous research on children's acquisition of adjectives, focusing specifically on gradable adjectives (GAs), adjectives that admit of degrees and therefore appear felicitously in comparative constructions. Though this research pointed towards preschool-age children's ability to set a standard of comparison based on information from the discourse context and to shift this standard when contextual variables change, I pointed out that because these experiments focused on adjectives related to physical dimensions, we are still left with an incomplete picture of what children know about the full range of GAs. While some GAs (relative GAs) do rely on the context in this way, others (absolute GAs) do not. In contrast to adjectives like *big* and *tall*, adjectives like *full* and *spotted* are gradable, but their standard does not depend on a contextually-relevant comparison class. Instead, it is oriented either towards the maximal absence of a property or the presence of a minimal value. Consequently, they are not susceptible to the same vagueness issues as relative GAs. In Chapter 3, I reviewed previous work in semantics and philosophy accounting for the differences among these three subclasses of GAs, concluding that a distinction between an open or closed scalar structure captures these differences.

Having described the difference in semantic representation between these three subclasses of GAs (relative, maximum standard absolute, and minimum standard absolute), I returned to a question about child language: do children have correct semantic representations for each of the three subclasses of GAs? That is, does their performance with dimensional adjectives reflect their knowledge of relative GAs alone (in which case we need to fill out our picture of how they represent other GAs), or is it indicative of their understanding of all GAs (in which case they erroneously think that all gradable predicates are context-dependent)?

In Chapter 4, I addressed this question. Presenting experimental evidence from two very different kinds of tasks (a Scalar Judgment Task and a Presupposition Task), I showed that in their behavioral responses, children as young as age three distinguish between these three subclasses of GAs. They think that relative GAs such as *big* and *long* do depend on the context for the standard of comparison: they set the standard somewhere near the midpoint of a series, and allow this standard to shift when the context changes so that a cutoff can always be established to differentiate between two different-size objects. By contrast, minimum standard absolute GAs have a standard oriented towards the minimal value of a scale. They signal the presence of a property (i.e., *spotted* means 'possessing some degree of spots'), and what is means to be e.g., *spotted* or *bumpy* is based on this value. Similarly, maximum standard absolute GAs have a standard oriented towards the maximal value of a scale. They signal the absence of a property (i.e., *full* means 'maximally full' or 'having no degree of emptiness'), which determines what it means to be e.g., *full* or *straight*. An interesting semantics/pragmatics distinction (seen in children's willingness to accommodate a speaker's request) also arose in the course of these experiments, allowing us to distinguish between the vagueness that is part and parcel of relative GAs and the imprecision arising from leniency in the degree to which the maximum standard must hold for this class of absolute GAs (i.e., for something to be considered *full* it may not have to be 'maximally' full, but close enough to this value to count as such).

Based on these results, I asked how children could have arrived at such sophisticated knowledge by age three. Following patterns we observed in Chapter 3, I suggested that one possibility is that infants pay attention to patterns of adverbial modification. How is adverbial modification informative about the scalar structure of GAs? In Chapter 3, I showed that some adverbs select for closed scales (e.g., *completely*, *perfectly*, *almost*, and so forth, selected for a maximal value, while *slightly* selects for a minimal value), while others (e.g., *very*) are intensifiers that modify any adjective that has a relative interpretation (either as a default or as a secondary, coercible interpretation). These differences give rise to the variation in entailment patterns discussed in that chapter.

In Chapter 5, I presented the results of a corpus analysis that demonstrated that distributional patterns of adverb-adjective bigrams in the exposure language are robust enough to be used by the language learner to classify GAs by their scalar structure. This corpus analysis revealed that maximum standard absolute GAs such as *full* have a higher probability of being modified by restricted adverbs such as *completely* than do non-maximal GAs such as *big*, and that non-maximal GAs are overwhelmingly more likely to be modified by general 'scale-selecting' adverbs such as *very*. Given that this information is reliably detectable in the exposure language, I then asked whether infants make use of such information when learning about adjectives.

In Chapter 6, I provided evidence from a word learning experiment with 30-month-olds that infants do use adverbs to assign either a relative or absolute interpretation to a novel adjective. When they hear a novel adjective (e.g., *wuggin*) modified by *completely*, which requires a closed scale, they assign an interpretation to the adjective that is consistent with a maximum standard absolute gradable status (i.e., *transparent*, not *tall*); but when the adjective is modified by *very*, which boosts the standard, so to speak, they assign an interpretation consistent with a relative gradable status (i.e., *tall*, not *transparent*). This pattern is consistent with the distribution of adverb-adjective bigrams identified in Chapter 5. What is also important is that infants show no default interpretation when the novel adjective is not modified by any adverb, when it is modified by a novel adverb (*penticly*), or when it is modified by an adverb with lower frequency than *completely* but a meaning like that of *very*. Infants therefore appear to be guided in their interpretation of new adjectives they encounter in the input by adverbs whose meaning they know. Adverbial modification thus appears to be one way to classify adjectives as relative or absolute gradable, mapping to either an open or closed scale. In this chapter, I describe in detail what the role of adverbial modifiers is in helping the language learner classify a given adjective with respect to its scalar structure.

7.2. A Remaining Question and a Proposal

We've seen that patterns of adverbial modification can be used to classify GAs by their scalar structure, and that infants attend to an adverb whose meaning (i.e., selectional restrictions on scalar structure) they know when assigning an interpretation to a novel adjective. However, there is an important question that remains to be answered, which concerns the following question. How does an infant move from groupings of lexical items based on relative frequency or probability of occurrence to generalizations about semantic representation? Even if an infant has discovered a correlation between an adverb (or set of adverbs) and a set of adjectives in the input, how does she know that this correlation is based on scalar structure? Furthermore, how does she use this aspect of meaning to infer something about the meaning of newly-encountered adjectives that also appear with that kind of adverb (or perhaps never does)? Learning about adjectives from adverbs requires that children would first have to know something about the meaning of the adverbs, but how would they have learned the meaning of the adverbs without first knowing something about the meaning of the adjectives? The idea I will pursue here is that children enter the acquisition process with an expectation that adjectives can map onto scales.

This chapter proceeds as follows. In § 7.3, I discuss how tracking form-meaning correspondences in the input can be informative to the language learner. Here, I make a specific connection between previous accounts of verb learning and the present account of adjective learning, the point of commonality being that a probabilistic analysis of syntactic environments in the input and the lexical items appearing in them is used to infer something about the meaning of a word and the category to which it belongs. In § 7.4, I describe in detail how an analysis of adverbial modification in the input can be informative about scalar structure. In § 7.5, I describe how this account of the way that infants learn about the meaning of adjectives explains the experimental results from Chapter 6. In § 7.6, I present in detail the aspects of scalar structure that need to be learned and make a connection to a previous account of learning from the input. In § 7.7, I make a connection between paths in the representation of events and scales in the representation of properties, which suggests that constraints arising from similar conceptual representations are at work in both domains. In § 7.8, I finish by showing that evidence from studies in cognitive development indicates that infants attend to the structure of paths when parsing events, which makes it more likely that they would also be attending to the structure of scales.

7.3. Form and Meaning in Language Learning

The goal in this chapter is to develop an account of how children learn something about the meaning of adjectives by attending to patterns of adverbial modification in the input. This type of proposal – that children can abstract something about word meaning from patterns detectable at the surface level in the input – closely resembles one that has been proposed in verb learning. A number of researchers over the last two decades have pursued the hypothesis that robust correspondences between form and meaning can play a central role in how infants learn about verbs. Here I will review the main approaches in this line of research, focusing on those accounts that are best-suited to elucidate the question of how adverbial modifiers can be informative to the child learning about adjectives.

7.3.1. Verb Learning

Pursuing the idea that the surface form of language can be a window into more abstract differences in meaning (Fillmore, 1968a, 1968b; Grimshaw, 1982; Jackendoff, 1978, 1983; McCawley, 1968; Vendler, 1972), a number of researchers in language acquisition have focused on how such surface-level cues can play a role in children's developing understanding of a number of linguistic distinctions such as the count/mass distinction in the nominal domain (e.g, *an apple*/*two apples* vs. *some rice*) or the difference between the types of words appearing in predicate position (e.g., *This is hairy*/*Harry*) (cf. Brown, 1957; Bunger, 2006; Katz, Baker, & Macnamara, 1974; Lee & Naigles, 2005; Naigles, 1990; Wagner, 2006 for a range of interests). Within the verbal domain, a predominant view has been that infants can recruit information about the range of frames in which a verb appears, paired with experiences in the world, to learn about verb meaning. The hypothesis, generally stated, is that verbs with similar meanings share similar subcategorization patterns or argument structure (i.e., whether it requires a direct object, a locative prepositional phrase, etc.), which are captured in the syntax. The phrase structure therefore acts as a 'zoom lens' for infants' acquisition of verb meaning by helping them narrow down the range of possible verb meanings for any given verb (Fisher, Hall, Rakowitz, Gleitman, 1994; Gleitman, 1990; Landau & Gleitman, 1985).

In an analysis of maternal input to children, Naigles and Hoff-Ginsberg (1995) argue that it is indeed the *range* of frames, and not singular occurrences, that are informative about verb meaning. That is, what the language learner needs to track is the relative frequency with which a given verb appears across a range of frames, and the relative frequency with which these frames appear across a range of verbs. Using these tendencies, infants can make generalizations that they can then extend to new occurrences of verbs in order to make predictions about meaning. Their analysis demonstrates that these frequencies are robust enough in the input to be informative to the language learner. (See also Goldberg, Casenhiser, and Sethuraman (2004, 2005) for similar claims appealing to constructions, and a thorough statistical analysis of distributional patterns in the input.)

Alishashi and Stevenson (2005a, b) argue in a similar vein that groups of argument structure frames (their *constructions*)[88] are a key component of a probabilistic account of verb meaning. Like Naigles and Hoff-Ginsberg, they claim that the language learner makes generalizations on two levels, taking into account the verbs that appear with different frame classes and the groups of frames that appear with individual verbs. In their Bayesian model, the frequency of occurrence of verbs and frames was weighted based on occurrence in the input. The language learner is then supposed to use this knowledge to make predictions about missing information (e.g., features encoded in a frame).

---

[88] Alishashi and Stevenson (2005a) describe a frame as containing the syntactic pattern of verb usage (roughly, a linear identification of the order of the verb and its argument(s)), accompanied by the semantic properties of a verb and its arguments in that pattern (i.e., thematic roles). Each frame has a head verb and semantic primitives. A group of such frames sharing a common syntax is a construction. They say that they diverge from Goldberg (1995) in thinking that it is not the more simple or general meaning of core *light* verbs (e.g., *go*, *make*) as such that give rise to constructions, but entrenchment of primitives through probabilistic usage, which may be correlated with such verbs.

7.3.2. Adjective Learning

Let's now connect these claims about infant verb learning[89] with the account I have proposed for adjective learning in the previous chapters (specifically, Chapter 5). In the accounts of verb learning outlined above, the syntactic frame is said to be informative about the number and type of arguments a verb takes, which are in turn correlated with the number and type of participants in the type of event the verb denotes. Syntactic distribution is therefore informative about the lexical representation of verbs and the semantics of events. Learners track the verbs appearing in frames (or classes of frames) and the range of frames appearing with individual verbs, and then use this information to make inferences about the lexical representation of the verb and also make predictions about new instances.

In the account of adjective leaning presented here, the role of frames is played by adverbs. The choice of adverbial modifier is correlated with the selectional requirements of the adverb, which are based on the scalar structure of the property denoted by the adjective. Distributions of adverb-adjective bigrams are therefore informative about the lexical representation of adverbs and adjectives because they are informative about differences in scalar structure. In both cases, distributional differences that are detectable at the surface level are a cue to more abstract differences that partition these lexical items according to their semantics. Just as children track the types of verbs appearing in different frames and the range of frames

[89] The ability to track surface-level patterns in the input and infer something about the similarities or differences in meaning that account for these patterns is not particular to verb learning; rather, these patterns in the input are a resource the child recruits in their acquisition of language in general. However, because the claims for verb learning have been made most explicitly, and are accompanied by analyses of the input that address statistical probabilities, I use this domain as the basis of comparison.

appearing with a particular verb, they also track the types of adjectives that are modified by different adverb classes and the range of adverbs that modify individual adjectives.

The analysis I conducted in Chapter 5 demonstrated that this account of adjective learning is possible, by showing that there are differences among restricted and non-restricted adverbs (i.e., adverbs with or without restrictions about the scalar structure of the adjectives they modify) in their modification of maximal and non-maximal adjectives. The general finding was that maximum standard absolute GAs such as *full* have a higher probability of being modified by restricted adverbs such as *completely* than do non-maximal GAs such as *big*, and that these non-maximal GAs are overwhelmingly more likely to be modified by general 'scale-selecting' adverbs such as *very*. The trends were significant in analyses of both adverbs and adjectives.

7.4. The Role of Adverbs in the Scalar Classification of Adjectives

I have said that if language learners perform a statistical analysis of the input and discover distributional differences among adverb-adjective bigrams, they must also know what it is that gives rise to these differences. This point is crucial, because the power of these groups of adverbs and adjectives is that the language learner can use them to assign the right abstract representations to the surface forms or deduce word meaning. It is also crucial is that in this account, the child is learning about the meaning of adjectives from adverbs *and* learning about the meaning of adverbs from adjectives. Why is this so? Let's consider what the task looks like both when infants *know* the meaning of a given adverb and when they *do not know* its meaning.[90]

---

[90] See Xu and Tenenbaum (2007) for an account of word learning that uses a Bayesian framework and that is compatible with the account offered here. In such an account, expectations about scalar structure are part of the *prior*.

To begin, if infants know the meaning of an adverb, then they know something about its selectional restrictions with respect to the scalar structure of the adjectives it modifies (along with similar features of non-adjectival items, although I am leaving those aside for now and restricting our discussion to adjective-modifying instances). Thus, when this adverb appears with an adjective infants know (that is, comprehend, but not necessarily produce), this appearance strengthens their hypothesis about what kind of adjectives this adverb modifies. This strengthened hypothesis in turn helps them when they observe this adverb modifying an adjective whose meaning they do not know. When this happens, infants are able to use their knowledge of the selectional restrictions of the adverb and the probabilities they have calculated concerning the known adjectives it modifies to assign a probability of word meaning to this new adjective. If the adverb is more likely to modify relative GAs, then it is likely that this new adjective is a relative GA and maps onto an open scale. If the adverb is more likely to modify maximal adjectives, then it is likely that this new adjective is an absolute GA and maps onto a closed scale. Note, though, that this is not the only resource at the infant's fingertips to classify the adjective. An infant can also use, for example, overlap of other features with the other adjectives in this group.

Now, what happens when infants do not know the meaning of an adverb? If they have attended to distributional differences in the input, they should be able to use the meanings of the adjectives that they know that appear with this adverb to deduce something about its meaning. By paying attention to which adjectives appear with this adverb and the frequency with which they appear, and the other adverbs with which they (do or do not) appear and the frequency with which they appear with these, then infants should be able to make a generalization concerning the selectional restrictions of this adverb, assigning it a probability of having one or another kind of selection restriction with respect to scalar structure. (Here, too, I am focusing only on what infants can learn from adverb-adjective bigram distributions. Of course, the full range of items a given adverb modifies should be taken into account when identifying how it interacts with scalar structure. (See, for example, the analysis of occurrences of *almost* in transcripts from CHILDES included in Chapter 5.) The meaning of the adverb can then play a role in adjective learning. However, when infants do not know the meaning of an adverb, and do not know the meaning of the adjective(s) appearing with it, they can only track frequencies, and must use any number of other cues – e.g., other syntactic environments in which these lexical items appear, real-world experiences, etc. – to jumpstart the process.

7.5. Explaining the Experimental Results

Let's assume, then, that the infants in the word learning experiments presented in Chapter 6 expected that the adverbial modifiers they heard would be informative about the scalar structure of the novel adjectives that were modified. How can we explain the experimental results under this account? Recall that the infants in the no-adverb condition performed at chance level in both the salience and response windows of the test phase, thereby indicating that there was no apparent default scalar interpretation of a novel adjective in the absence of input cues. By contrast, in the conditions involving an adverb with robust distributional patterns in the exposure language, infants exhibited looking patterns that were in line with these distributions. During the response window, when asked which object had the relevant property, infants in the *completely* condition looked longer to the object that manifested the absolute, but not the relative property, while infants in the *very* condition looked longer to the object that manifested the relative, but not the absolute property. This is to be expected: *completely* selects for a closed scale and is more likely to modify a maximum standard absolute GA (e.g., *full*) than a relative

GA (e.g., *big*), while *very* is an intensifier that is compatible with a much wider range of adjectives, but in the input is more likely to appear with relative GAs. We can conclude from these results that infants recruited what they knew about the distributions of these modifiers with real adjectives in the exposure language when assigning a meaning to novel adjectives that appeared with them in the experiment.

The distribution of adverb-adjective bigrams in the exposure language arises from compatibility between the selectional restrictions that are part of the meaning of the adverbial modifiers with the scalar structure in lexical representation of the adjectives. Infants were able to use adverbs with robust distributional tendencies to assign either an absolute or relative interpretation to the novel adjectives in these experiments. We therefore have reason to think that infants knew that differences in scalar structure underlie the distributional patterns in the exposure language, and used this information when assigning a meaning to new adjectives appearing with these adverbs. It is thus the *meaning* of these adverbs that matters.

It should not be surprising, then, that infants in the condition in which the novel adjectives were modified by a novel adverb, *penticly*, exhibited no such bias. There is no distributional information from the exposure language on which they can rely. The only information they have what they are given in the experiment – an experiment in which the novel adverb is paired with five different sets of stimuli, described by five different novel adjectives, in less than five minutes total time. Even if infants do have a default bias concerning the selectional restrictions of a novel adverb and can use it as a hypothesis guiding their interpretation of novel adjectives, they will need some kind of information that bears on their hypothesis. In the absence of such information, they suspend their judgment about the scalar structure of the adjective. An experiment that remains to be done, then, is pairing a novel adverb such as *penticly* with adjectives 30-month-old infants *do* know to see if they deduce something about its selectional restrictions, and can then use this information when assigning an interpretation to novel adjectives.

A second possibility exists, however, that is quite different from the scenario described above. It could be that the infants did not interpret *penticly* as an adverb. Perhaps, infants ignored the salient morphological cue (the -*ly* ending) and interpreted this word as another novel adjective. Some adjectives do end this way (e.g., *silly*). If this was the interpretation infants assigned to *penticly*, then their performance in the experiment could be explained in the following way. Upon hearing, "Look! These are both penticly wuggin!," infants understood the speaker to say that both objects had two properties, one denoted by *penticly* and another denoted by *wuggin*. In fact, the objects in the experiment did have two relevant properties, a relative and an absolute one. At test, when children were asked, "Which one is wuggin?," their task would not be to use the supposed selectional restrictions of the adverb to learn about the meaning of the adjective, but rather to align the second adjective with the correct choice of adjective-property mapping.

It appears to be the case that relative GAs generally precede absolute GAs when there is more than one adjectival modifier ((95)-(96)), an observation which may have to do with the type of properties denoted by these two types of GAs.[91]

95. a. tall transparent containers

 b. #transparent tall containers

96. a. long straight rod

[91] See Bever (1970), Cinque (1994), Dixon (1977), Prasada (1992), Sproat and Shih (1991) and Teodorescu (2006) for discussions of adjective ordering in prenominal position.

b. #straight long rod

Infants who did not recognize *penticly* as an adverb may have interpreted it as an adjective corresponding to a relative property, and the second adjective as corresponding to an absolute property.

If this were the case, we might expect them to consistently map the novel adjective to an absolute property. However, this may not have happened for the same reasons cited in my explanation for the no-adverb condition: infants were being asked to learn the meaning of five different adjectives in a very short amount of time, without multiple instances for each.[92] If, according to this account, infants were trying to keep track of the five novel adjectives and their possible property correlates, along with *penticly* – whose meaning either would have kept changing from trial to trial or would have to be general enough to subsume five very different properties – they may have suffered from cognitive overload and therefore patterned at chance.

However, this take on the experimental results also assumes that children's expectations fly in the face of the syntactic structure licensed in English. It requires that infants interpreted "These are both penticly wuggin" as "These are both A_1 A_2." However, in English, this sentence would be ungrammatical; the adjectives would have to be conjoined (compare (97a) to (97b)).

97. a. *These are both tall transparent.

b. These are both tall <u>and</u> transparent.

In addition, the restrictions on adjective ordering hold for adjectives in prenominal, not predicate, position (cf. (98)).

---

[92] See Liu, Golinkoff, and Sak (2001), Mintz and Gleitman (2002), Waxman (1999), Waxman and Klibanoff (2000), and Xu and Tenenbaum (2007) for discussions of how infants benefit from exposure to multiple exemplars in word learning experiments.

98. These are both transparent and tall.

Infants would need either to not know about this grammatical restriction at 30 months of age, suspend their expectations for the sake of this experiment, or posit a null nominal element so that prenominal adjective ordering restrictions hold. (A follow-up experiment in which these words do appear in prenominal position would help to determine whether this was a factor.) For all of the reasons outlined here, saying that *penticly* was not interpreted as an adverb but rather as another novel adjective seems a bit far-fetched (especially given the emphatic prosodic contour *penticly* was assigned), but this possibility cannot be ruled out at present.

Finally, we need to say how the account I have proposed for using adverbs to classify adjectives according to their scalar structure can be reconciled with the results from the *extremely* condition. Recall that as in the no-adverb and *penticly* conditions, infants patterned at chance in the *extremely* condition. That is, when they heard, e.g., *extremely A*, they were no more likely to map the novel adjective onto a relative or absolute property. This should be puzzling given that *extremely* shares a meaning similar to that of *very* and like *very*, is highly unlikely to modify a maximal lexical item.[93] However, there is a key difference between *very* and *extremely*: frequency.

Infants are practically inundated with evidence in the input concerning *very*'s behavior as a modifier. This is not the case with *extremely*. In fact, in Chapter 5 (where a detailed analysis of frequency patterns in the spoken BNC corpus is presented), I pointed out that *extremely* is only half as frequent as *completely*, which was already tens of times less frequent than *very* in the spoken BNC corpus. The explanation I offered at the end of Chapter 5 was that *extremely* may

---

[93] In Chapter 5, I also showed that *very* is more likely to modify a relative GA when it modifies an adjective in the spoken BNC. The same holds for *extremely*.

be too infrequent for infants to have generalized anything about its meaning, and as a result, infants may simply treat it as a novel adverb. I also offered a second possibility, which is pragmatic in nature. Infants may expect to hear the more highly frequent adverb *very* to be used as a modifier, so when they hear *extremely*, they are confused. Although they may have a budding hypothesis that *extremely* is an intensifier, they are reluctant to proceed confidently with this kind of meaning and so end up patterning at chance.

It appears, then, that infants at 30 months of age are performing a probabilistic analysis of the input, guided by constraints arising from their conceptual representations, to deduce the meaning of these lexical items (both adjectives and adverbs). When they pattern at chance, there is good reason – either they are not given the information they require, or they are uncertain about what to feed forward. There are a number of follow-up experiments, including those suggested above, that would help us determine more precisely what the learning process looks like.

7.6. Learning about Scales

I have explained how infants can use the distribution of adverbs and adjectives to learn how to classify adjectives according to their scalar structure and adverbs according to their selectional restrictions concerning scalar structure. Other evidence strongly suggests that the type of questions children hear can help distinguish between adjectives and nouns (Gasser & Smith, 1998) or between color and size terms (Sandhofer, 2002). For example, parents pose questions such as “How big is that?,” “Which is the big one?,” and “What color is that?,” and “What size is that” – questions that evoke very different kinds of answers and draw attention to different properties or entities. These questions, then, may also serve to partition adjectives by their meaning. But even if, after a thorough analysis of distributions in the input, the learner ends up with different classes of adverbs and adjectives, the child must still know what it is that unites these different groups. I have proposed that infants are guided by conceptual representations to think that the differences in distribution may arise from differences in scalar structure.

Directly relevant to this claim is work by Bloom and Wynn (1997), in which the authors address how syntactic and semantic cues in the input could help the language learner classify and learn the meaning of number words. I do not wish to evaluate whether or not the account they propose is indeed the path by which children acquire the meaning of number words. What is relevant here is their analysis of the language input, in which they identify the inability of number words to be modified by adverbs such as *too* and *very*[94] as one of the main cues distinguishing number words, along with their appearance with count vs. mass nouns, the order relative to adjectives, and their ability to appear in partitive constructions. Number words therefore pattern with quantifiers (e.g., *every*, *most*, *some*, *few*), but are different from adjectives such as *much* and *more*[95] and GAs (e.g., *big*, *little*). The full set of distinctions among the four categories according to the features analyzed by Bloom and Wynn (1997) in the speech to and by children is captured in Figure 28.

[94] They simply call these adverbs 'modifiers'.

[95] See Schwarzschild (2002) for reasons why these terms are considered adjectives and a discussion of their semantics.

Figure 28: Categorization according to four features of the input in Bloom & Wynn (1997)

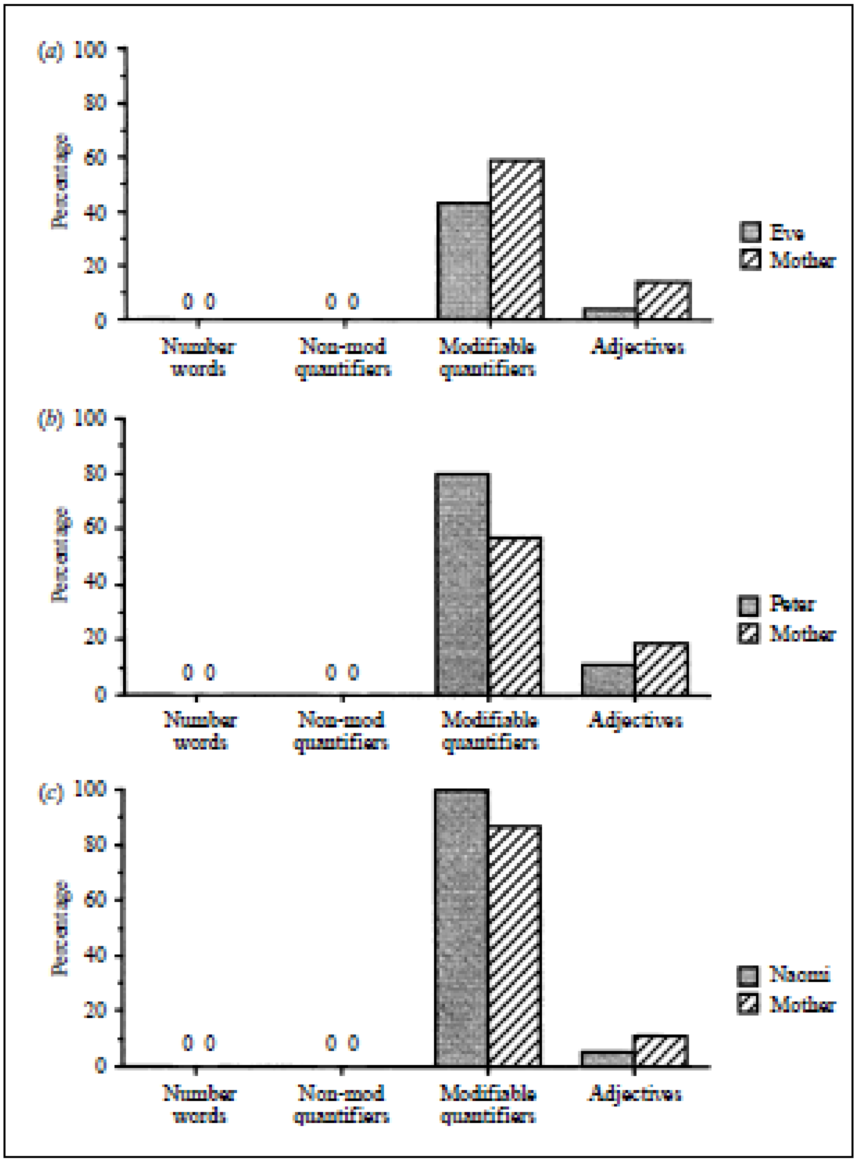
target words		cue			
category	examples (from B&W)	1: Appears with a mass noun in input?	2: Can be modified by *too*, *very*?	3: Can follow adjectives?	4: Can appear in partitive constructions?
number words	*two*, *three*	N	N	N	Y
'non-modifiable quantifiers'	*some*, *both*, *another*, *any*, *all*, *several*, *only*	N	N	N	Y
'modifiable quantifiers'	*much*, *many*, *most*[96], *few*	Y	Y	N	Y
'adjectives'	*big*, *little*	N	Y	Y	N

Bloom and Wynn argue that the meaning of these modifiers, which seem to select for "continuous" properties, is incompatible with number words, which refer to "discrete" properties, and that given that some adjectives (e.g., *even*, *odd*, *prime*) are also barred from being modified by *too* and *very*, this distinction crosscuts the syntax.

In an analysis of mother's and child's speech in three CHILDES transcripts (Eve, Peter, Naomi) (Bloom, Hood, & Lightbown, 1974; Bloom, Lightbown, & Hood, 1975; Brown, 1973; MacWhinney, 2000; Sachs, 1983), Bloom and Wynn demonstrate that children's production reflects these distinctions in adult speech (cf. Figure 29, which captures the percentage of utterances in which the target category was modified by *too* or *very*).

[96] *Most* is included in this list, because Bloom and Wynn (1997) include it, but it clearly cannot appear with either of the two modifiers in question (e.g., **too most*, **very most*). *More* was deliberately excluded from their analysis.

Figure 29: Percentage of category modified by *too* or *very*, results from Analysis 2 of Bloom & Wynn (1997) (figure reproduced from source)



Based on this distributional pattern, they concluded that both adults and young children recognize a "semantic distinction between words denoting properties that fall on a continuum (e.g. size or magnitude) versus words denoting discrete properties (e.g. being of a certain numerosity)" (pg. 525). That children's production mirrors the mothers' also appears to be an indication that they attended to the distribution of the target words in the input and made inferences about the semantic classes based on these differences.

Now, since number words never appear with these adverbial modifiers, and children are not told that *very two* and similar phrases are barred, how do children know that number words cannot be modified by these terms? Further, how do they know that number words can be differentiated from other words based on something like a discrete/continuous distinction? An initial answer offered by Bloom and Wynn is that children may make use of "indirect negative evidence": if number words never appear with mass nouns, which are highly frequent, then they may have evidence that they are restricted to count nouns, which are individuated and call for discrete enumeration. However, given what Bloom and Wynn claim is a relatively low frequency of these modifiers, they argue that the same inference may not be possible in this case and that there may be an alternate possibility.

The second possibility is that children begin with the assumption that "all predicates cannot be used with modifiers, that they denote absolute properties. Under this account, only upon hearing a word used with a modifier will a child infer that it could refer to a continuous property…In other words, the child's null hypothesis, in the absence of any syntactic and semantic evidence to the contrary, is that all predicates are like number words in that they cannot be modified" (pg. 518).

The proposal I am making is slightly different. I propose that children look for evidence in the input about the semantic status of adjectives (predicates). Scales are part of their conceptual repertoire, so when they are provided with evidence that an adjective has a gradable interpretation, they recruit this conceptual information to assign the correct interpretation to this new word. The null hypothesis is not that predicates cannot be modified, but that patterns of modification will vary and will be informative about scalar structure.

Let's revisit the claim that it is the presence or absence of modifiers that matters. In Chapters 3 and 5, I discussed a range of adverbial modifiers, demonstrating that such terms can be partitioned into groups based on their restrictions on the semantic representation of the words they modify.[97] While modifiers such as *too* and *very* select for a relative interpretation, modifiers such as *almost* and *completely* require a maximal value and modifiers like *slightly* a minimal value. Thus, children must not just attend to *whether* predicates appear with modifiers. They must also observe the *kind* of modifiers with which they appear.

Now, it may be that when Bloom and Wynn refer to modifiers, they are simply using shorthand for the kind of modifiers that are the focus of their analysis. If this is the case, lets it is also possible to revisit their claim that these modifiers are infrequent, at least relative to mass nouns. Recall that Bloom and Wynn reason that since number words never appear with mass nouns but do appear with count nouns, children receive indirect negative evidence that number words are restricted to appearing with terms that denote entities that can be individuated. However, in the case of modifiers, they argue that since these terms are infrequent, children would not be provided with the same kind of negative evidence about the discrete/continuous distinction.

[97] Although I focused on cases in which adverbs modified adjectives, any selectional restrictions that may arise from the representation of an adverb will place requirements not only on adjectives it modifies, but on any term in general. Thus, an adverb such as *almost* requires that adjectives can provide a maximal value (e.g., *full*) and that events are telic or bounded (*ran five miles*).

It is not clear, though, that such modifiers are actually more infrequent than mass nouns. Turning again to the spoken subset of the British National Corpus, we can see that a comparison of these two modifiers with highly frequent mass nouns (drawn from a list provided in the Wikipedia entry on *mass nouns*) reveals that *very* is much more frequent than and *too* nearly as frequent as the most frequent of these mass nouns (*money*). If children can learn something about the lack of appearance of an item with mass nouns, they can certainly learn something about its lack of appearance with these modifiers as well.

Figure 30: Frequency of selected modifiers and mass nouns in the BNC spoken texts

target items	number of matches in 10,341,729 words	instances per million
modifiers		
very	25,041	2421.36
too	6511	629.59
mass nouns		
money	6599	638.09
water	2897	280.13
information	2080	201.13
traffic	1181	114.2
equipment	393	38
furniture	345	33.36
rice	155	14.99
laughter	21	2.03

Second, it is not just the raw frequency that matters, but relative frequency. Two aspects of frequency matter in particular: the frequency with which items appear in different classes of environments and the range of environments in which a particular item can appear. Bloom and Wynn demonstrated in their analysis of the CHILDES transcripts that number words do not appear with *too* and *very*. However, it is not the case that number words cannot appear with any modifiers. In fact, for any predicate, the question is not *whether or not* modifiers appear with these items, but *which* modifiers appear with them and what the probability of cooccurrence is.

In Chapter 5, I pointed out that adverbs such as *too* and *very* signal gradability on the part of the item modified. If this class of modifiers does not appear with number words, what other class of modifiers does?

Number words cannot be modified by *slightly* (99), which appears to indicate that they are not like minimum standard absolute GAs (e.g., *spotted*, *wet*) in requiring a minimal degree to hold. They also cannot be modified by *completely*, *100%*, or *half* (100), which appears to indicate that they are not like maximum standard absolute GAs (e.g., *full*, *dry*) in requiring a maximal degree of the relevant property's absence. But, as Bloom and Wynn note, number words can appear in partitive constructions (101).

99. *slightly three

100. a. *completely three

b. *100% three

c. *half three

101. a. 100% of three

b. half of three

Thus, numbers admit of part structure, which in and of itself is not surprising. However, consider the fact that numbers are ordered along a number line. This linear ordering is itself a scale, and as such, should admit of part structure (cf. Schwarzschild & Wilkinson, 2002). If numbers are scalar predicates, we would expect them to appear with a class of modifiers that highlights this aspect of their meaning. In fact, numbers routinely appear with modifiers such as those in (102).

102. a. almost three

b. nearly three

c. exactly three

d. at least three

Recall from Chapters 3 and 5 that modification by *almost* is only permissible when a predicate admits of a maximal value (cf. Amaral, 2006; Rotstein & Winter, 2004; Winter, to appear; *inter alia*). Thus, number words pattern here with maximum standard absolute GAs. And, like maximum standard absolute GAs, numbers allow for imprecision (103).

103. a. Marie arrived at three o'clock. (Lasersohn, 1999, (1), (55))

(uttered some time after Marie arrived at a social event at 3:02)

b. The drive from Chicago to Washington, D.C. is 700 miles.

(uttered in preparation for a roadtrip, which actually measures 711 miles)

Bloom and Wynn say that numbers are absolute properties that do not admit of variation. There is a rich literature centering around scalar implicatures[98], including those associated with number words. On a very basic level, numbers have two interpretations, as illustrated in ((104)-(105)).

104. a. She has 3 children.

b. = She has *exactly* 3 children.

105. a. I have $5.

b. =I have *at least* $5.

[98] Cf. Carson, 1998; Chierchia, 2001; Fauconier, 1975; Gazdar, 1979; Geurts, 1998; Grice, 1975; Hirschberg, 1985; Horn, 1972; Sperber & Wilson, 1986; *inter alia*.

Numbers, then, do indeed admit of variation. It may be that in the beginning, children are not sensitive to this type of variation, but that this may be due to pragmatic and not semantic knowledge.[99]

The point is that with numbers, too, it is possible for children to expect that their appearance in syntactic environments – specifically, patterns of modification – will be informative about the type of scalar structure that is part of their meaning. And while numbers may diverge from other gradable predicates in terms of a discrete/continuous distinction, this feature may be associated with the way that number words highlight intervals on the scale, and does not necessarily indicate a difference in the composition of the scale *per se*. (See Fox & Hackl (2006) for related discussion.) For both number words and adjectives, modifiers are a cue to the semantic classification of these words with respect to their scalar structure.

In talking about the representation of GAs, I have focused on the structure of their scales (i.e., whether the scale is open or closed, and if it is closed, whether it is closed at the minimal and/or maximal ends). Given that adjectives vary along this aspect, the task for the word learner is to figure out how to classify a new adjective in the input. It follows that infants will need evidence to guide them in this classification. They must be presented with evidence of some kind that the property can admit of degrees and therefore that the corresponding adjective is gradable. Setting aside all of the interesting conceptual and real-world cues that may exist,

---

[99] For work on children's performance with scalar implicatures, cf. Chierchia *et al*., 2001; Gualmini *et al*., 2001; Huang, Snedeker, & Spelke, submitted; Miller, Schmitt, Chang, & Munn, 2005; Musolino, 2004; Noveck, 2001; Papafragou & Musolino, 2003; Papafragou & Schwarz, 2005-6; and Papafragou & Tantalou, 2004. For an explicit comparison of numbers and quantifiers in children's understanding of implicatures, see Hurewitz, Papafragou, Gleitman, and Gelman (2006).

linguistically, if a child hears a word in a comparative construction, observes it in both attributive and predicate positions, notices that it can be modified by a degree or proportional modifier or by measure phrases, and so on – all of the diagnostics for gradability I mentioned in Chapter 3 – then they have good reason to think that this adjective is gradable. Otherwise, they must assume that it is not.

An adjective classified as gradable maps onto a scale, and the infant must also decide what kind of scalar structure to assign – whether it is open or closed. Note that infants do not need to 'discover' endpoints. They need to figure out when it is appropriate to mark a scale as open or closed. Now, as I discussed in Chapters and 4, this open/closed status has implications for how the standard of comparison is set (i.e., whether it is context-dependent and variable, or oriented towards a minimal or maximal value). It may seem a bit odd to think of infants using perceptual cues to learn about endpoints of scales, given that there is so much imprecision in language usage. Infants would seem to need to infer that something was the endpoint without a stringent maximal value actually holding. However, there is considerable evidence that children do have an expectation for the goal of an action, and can infer one even when the agent did not achieve it. (See § 7.8.) Knowing that the standard is context-dependent requires infants to know that taxonomic groupings of objects sharing similar properties can be informative about the standard that holds within that kind. The fact that young children's early ability to perform comparisons within basic-level kinds contrasts with their difficulty performing comparisons across kinds (cf. Waxman & Klibanoff, 2000; Waxman & Markow, 1995, 1998) could be seen as their learning how to identify or restrict the relevant domain that serves as the comparison class.

There is certainly more to knowing the meaning of GAs, though, than knowing about the structure of the scales onto which they map. A child will also have to learn about aspects concerning the composition of scales, such as the relevant dimension of the scale (i.e., SIZE, WEIGHT, TRANSPARENCY, etc.) or whether the adjective is a member of an antonym pair. If so, they will have to learn if the adjective is the unmarked/marked (or positive/negative) member, what its opposite is, and so forth. That children are still mastering some of these aspects into their fifth year is therefore not surprising.

7.7. Scales and Paths

Positing that infants have expectations about scalar structure has allowed us to explain how they can use distributional information from the input as cues to the lexical representations of adverbs and adjectives. Assigning an important conceptual and semantic role to scales is not only relevant to learning about words that denote properties, though. In this section I show that a very similar conceptual representation, paths, is found in the domain of events and plays a key role in semantic explanations of descriptions of events. By making a connection between scales and paths, we arrive at a better understanding of how this core conceptual component may more generally constrain infants' expectations about the possible meanings of words.

7.7.1. Paths in Event Representation

Let's begin by laying out some of the basic assumptions about what paths are and how they capture the structure of events. (Unfortunately, I will be forced to gloss over some very interesting aspects of path structure and focus only on what motivates my account of adjective learning and scalar classification.) Paths have been appealed to by a wide variety of researchers to describe our conceptual representations of events and the way our spatial language encodes relations in the world.[100] The structure of paths bears a striking resemblance to the structure of

[100] Cf. Gruber, 1965; Jackendoff, 1983, 1991, 1996; Johnson, 1987; Krifka, 1998; Lakoff, 1987; Landau & Jackendoff, 1993; Langacker, 1987; Regier, 1996; and Talmy 1983, 1985a, b, 1988.

scales. Like scales, paths are conceived of as one-dimensional and directional. And, like scales, paths are also composed of a sequence of contiguous (adjacent and connected) elements, which allows for the measurement of time, space, or motion along the path (e.g., *in*/*for five hours*, *for 5 miles*, *a long way*, *gradually*, etc.). Paths therefore have a part structure similar to that of scales.[101] Events such as those in (106a-c) can therefore be captured by paths.

106. a. Bill ran down the road. (Jackendoff, 1996)

b. The mouse skittered toward the clock. (Jackendoff, 1983)

c. I ran out of the house. (Talmy, 1985a, (62a))

Paths may also be bounded by a SOURCE and/or a GOAL, as in (107a-d), hereafter in lowercase.[102]

107. a. John ran from the house. (SOURCE) (Jackendoff, 1983, (9.9c))

b. The car drove into the valley. (GOAL) (Zwarts, (11a))

c. Mary walked from the university to the capitol. (SOURCE, GOAL) (Krifka, 1998, (74))

---

[101] Following Krifka (1998) and Jackendoff (1983, 1991, 1996), I will assume paths are linear, convex, non-branching, non-circular, and non-crossing. Here, I treat scales and paths are interchangeable concepts; however, I leave the question of whether these concepts should ultimately be collapsed open for future investigation. It may be that there is simply a tight connection between these two conceptual representations, whereby a scale used to measure progress along a path, and what distinguishes paths is that they can be used to track not only progress of an event, but also motion or spatial movement, and therefore can take various shapes, points can overlap, segments may be traversed repeatedly, etc. See Johnson (1987) for a discussion of possible differences between scale and path schema, and Eschenbach (2000), Eschenbach *et al*. (2000), and Zwarts (2005) for alternative perspectives.

[102] Cf. Dowty, 1991; Gruber, 1965; Jackendoff, 1983, 1991, 1996; Johnson, 1987; Krifka, 1998; Regier, 1996; Zwarts, 2005; *inter alia*.

d. I drove the Blue Ridge Skyway from beginning to end. (SOURCE, GOAL)

(adapted from Dowty, 1991)

Bounded events necessarily have a goal. Such events can be identified by Vendler (1957)'s *in an hour*/*for an hour* test ((108a) vs. (108b)): telic predicates can be modified by *in an hour*, but not *for an hour*, while atelic predicates manifest the opposite pattern. (See Gruber, 1965; Morzycki, 2006; and Zwarts, 2005 for discussion of the contribution of APs and PPs.)

108. a. The cart rolled to NY/into the house/over the bridge (in/*for an hour). (telic)

(Jackendoff, 1996, (26a))

b. The cart rolled along the road/toward the house (for/*in an hour). (atelic)

(Jackendoff, 1996, (27a))

Paths may also describe situations other than events. They may therefore have a spatial rather than a temporal dimension (109a-c), and they may even be metaphorical (110a-b) (cf. Gruber, 1965; Jackendoff, 1983, 1991; Johnson, 1987; Lakoff, 1987).

109. a. The firehouse is two miles from my house. (Jackendoff, 1983)

b. The road goes from New York to Boston. (Jackendoff, 1996, discussed in Krifka, 1998)

c. The signs point to New York. (Landau & Jackendoff, 1993)

110. a. He can go quickly from one mood to another. (Langacker, 1987)

b. Our clients range from psychiatrists to psychopaths.

(Jackendoff, 1983, after Gruber, 1965)

In these instances, as with events proper, a path captures directionality, permits measurement, and may or may not be bounded.

7.7.2. Homomorphism of Part Structures: How Paths and Scales Align

Because events, objects and pluralities (or collections of individuals), and scales and paths all have part structure,[103] a connection can be established between combinations of these elements. For example, the progress of an event over time can be measured by the way in which a participant of the event, or the theme, is affected.[104, 105] For example, in (111a-e), tracking the state of either the direct object of a verb of creation (111a) or consumption (111b), the subject of a sentence (111c-d), or multiple entities (111e) reveals how the event unfolds over time. The beginning of the event in (111a) can be marked by the apple being whole, but it is (completely) eaten by the end. And in (111d), John's position along the trail tracks the hiking event.

111. a. The man built a new oxcart.

b. John ate the apple. (Beavers, 2006, (15))

c. The train traveled to New York. (Jackendoff, 1983, (10.17a))

d. John hiked the Barton Springs trail. (Beavers, 2006, (15c))

e. It took Hurricane Archibald 3 ½ hours to cross the Florida peninsula. (Dowty, 1991, (26c))

Although these predicates are telic, they need not be. For example, it is possible to track the progress of an unbounded event.

---

[103] For discussion on the part structure of objects and pluralities, see Link (1983) and Landman (2000), and for the part structure of events, see Bach (1986) and Link (1987).

[104] Beavers (2006) distinguishes between a participant being *affected* or *totally affected*, a difference based on discussion by Dowty (1991). As this difference is outside the scope of the current discussion, I will leave it aside.

[105] Cf. Beavers, 2006, in press; Dowty, 1991; Hay, Kennedy, & Levin, 1999; Krifka, 1998; Morzycki, 2006; Ramchand, 1997; Tenny, 1987, 1992, 1994; and Winter, to appear.

112. a. The balloon slowly descended. (adapted from Hay, Kennedy, & Levin, 1999)

b. John ate (some) apples. (adapted from Krifka, 1998)

A homomorphism between the part structure of an object or plurality and time (113a) or space (113b) likewise sets up an incremental relationship such that the boundaries of the object or plurality delimit the path. For example, in (113a), the troops that compose the army make it across the river within an hour's time.

113. a. The army crossed the river in an hour. (Krifka, 1998)

b. The line of trees extended down the road. (Jackendoff, 1996)

c. The crowd exited the auditorium (in 21 minutes). (Dowty, 1991, (25b))

When a property changes over time, as in (114a-d), the initial and final states serve to delimit the path. For example, in (114a), the temperature of the water can be used to track the event of Mary heating it. At the beginning of the event, the water is at 30º C. The event ends when the water reaches 90º C.

114. a. Mary heated the water from 30º C to 90º C (in an hour). (Krifka, 1998, (76a))

b. Mary whipped the cream stiff (in an hour/*for an hour). (Krifka, 1998, (77b))

c. The weather turned from bad to worse. (Gruber, 1965, 7.7 (16))

d. John changed from happy into sad [in an instant]. (Gruber, 1965, 7.7 (24))

Finally, an explicit connection has been made between events and adjectival scales with respect to changes of state.[106] This connection comes in three main forms: resultatives, changes of state expressed by the positive and comparative forms of adjectives, and deadjectival verbs.

---

[106] Cf. Beavers, 2006, in press; Hay, Kennedy, & Levin, 1999; Jackendoff, 1983, 1996; Kearns, to appear; Kennedy & Levin, to appear; Schwarzschild & Wilkinson, 2002; Vanden Wyngaerd, 2001; Wechsler, 2005; and Winter, to appear.

Accounts of these three phenomena converge to suggest that they are indicative of a more general phenomenon involving measurement and the structure of paths or scales. I will simply briefly describe each of these areas so that a clear picture emerges of the connection between paths of events and the scalar structure of adjectives.

7.7.2.1. Resultatives

As I discussed in Chapter 5, resultative constructions in English (Beavers, 2006, in press; Wechsler, 2001, 2005) and in Dutch (Vanden Wyngaerd, 2001) place restrictions on the scalar structure of the adjective that heads the AP.[107] Specifically, the adjective must make reference to a maximal value, which provides an endpoint that bounds the event. Thus a maximum standard absolute GA, but not a relative GA, can appear in a resultative such as the one in (115).

115. Nate hammered the metal flat/*wide.

We can therefore think of the adjectival scale as allowing us to measure the change of state of the property, which in turn tracks the progress of an event, as in (111) and (114).

7.7.2.2. Positive Forms and Comparatives

Sentences such as those in (116a-b) describe a change of state involving the positive or comparative form of a gradable adjective, respectively. The size of the balloon decreases in both instances.

116. a. The balloon became small. (Jackendoff, 1983, (10.18a))

b. The balloon became smaller. (Jackendoff, 1983, (10.18b))

---

[107] Though I address only those resultatives in which the result XP is an AP, resultatives can also appear with result PPs. In both cases, the lexical semantics of the verb – which are scalar, or express durative events – constrain the possibilities for the result XP (cf. Beavers, 2006, in press; Filip, 2007; Rappaport Hovav & Levin, 2001; *inter alia*).

Jackendoff (1991, 1996) argues that predicates such as *is red* or *is small* express states, which do not have a temporal dimension as the unfolding of an event does. However, as he notes (1983), the process of changing states does have such a dimension, so a path structure can capture such situations. A sentence such as (116a) is therefore like those in (117a-b) in that they all express a change of state, where being small or being a cook or a pumpkin is the result state.

117. a. John became a cook. (Gruber, 1965, 7.7 (5))

b. The coach became a pumpkin. (Gruber, 1965, 7.7 (6))

Jackendoff (1983) says that in (116b), the theme approaches the goal without reaching it. The balloon may still be large, but it is smaller than at the beginning of the event. (This follows from the semantics laid out in Chapter 3.) This sentence is therefore similar to one such as (108b), repeated here as (118), where the house is the goal, and the cart rolled towards it, never reaching it, but is closer than when it first started.

118. The cart rolled toward the house.

Appealing to these connections with events, Jackendoff proposes the following structures in (119a-b) for the sentences in (116a-b).

119. a. [GO_{Ident} ([BALLOON], [$_{Path}$ TO_{Ident} ([$_{Property}$ SMALL])])]

b. [GO_{Ident} ([BALLOON], [$_{Path}$ $TOWARD_{Ident}$ ([$_{Property}$ SMALL])])]

(Jackendoff, 1983, (10.19))

These structures indicate a change in the 'identificational parameter' of the balloon such that it moves along a path in the direction of the goal of being small. In (119a), which encodes the identificational TO, the balloon reaches this goal, so the point of reference is the endpoint or result state. In (119b), which encodes TOWARD, it does not. (See Zwarts (2005) for discussion of the prepositions *to* and *toward*.) The fact that the positive form in these cases involves a bounded path, but the comparative form does not, also accounts for the pattern in (120), as in (108) above.

120. The water got hotter for/*in two hours. (Jackendoff, 1996, (38a))

Jackendoff (1983) also appeals to path structure in the representation of the comparative in (121).

121. Sally is three inches shorter than Bill. (Jackendoff, 1983, (10.20a))

Similar to the cases above, this sentence says something about the state of Sally, specifically how her height (the reference value) compares to the standard value, Bill's height. In his representation (Figure 31), the state expressed by identificational BE takes an Identificational Place,[108] a path on which the distance between Sally and Bill is 3 inches and she is oriented toward being short.

Figure 31: Jackendoff (1983)'s representation for *Sally is three inches shorter than Bill* (figure adapted from source, (10.21))

$$[_{\text{State}}\ \text{BE}_{\text{Ident}}\ ([\text{SALLY}],\ \left[_{\text{Place}}\ \text{ON}_{\text{Ident}} \left(\left[_{\text{Path}} \begin{matrix} \text{FROM}\ ([\text{BILL}]) \\ \text{TOWARD}_{\text{Ident}}\ ([\text{SHORT}]) \\ [_{\text{Amount}}\ 3\ \text{INCHES}] \end{matrix} \right] \right) \right])]$$

Jackendoff observes that this situation is similar to the one expressed in (122).

122. He lives two miles down the road from here. (Jackendoff, 1983)

[108] Jackendoff (1991) described his Places and Paths as Spaces. However, Paths are necessarily 1-dimensional and directional. All other Spaces are Places. Paths may or may not be bounded, an aspect that is captured with the [±b] feature for boundedness.

This representation is consistent with semantic accounts offered by Hay, Kennedy, and Levin (1999), Kennedy and Levin (2007), and Winter (2006), in that it allows us to treat differentials, intervals, and comparison between a reference and a standard value the same way for both paths (of events) and adjectival scales.

7.7.2.3. Deadjectival Verbs

Deadjectival verbs such as *flatten*, *straighten*, *lengthen* are a subset of Bertinetto and Squartini (1995)'s 'gradual completion verbs' (after Vendler) and Dowty (1979)'s 'degree achievement' verbs (also called 'extent predicates' by Gawron (2006a, b)). As their name indicates, deadjectival verbs are formed from an adjective.[109] Bertinetto and Squartini (1995) say that these verbs indicate the attainment of a further degree in the given dimension. Thus, these verbs, too, denote events that involve a change of state that can be measured, and so can be modified by measure phrase such as *(by) a lot* and *gradually* (Bertinetto & Squartini, 1995). As with the events described earlier, the progress of the event denoted by such verbs can be tracked by observing the state of an event participant. For example, over the course of the event described in (123a), Kim lengthens the rope until at the end, the rope is five inches longer. In (123a), the pages of the volume gradually turn from white to yellow over a number of years.

123. a. Kim lengthened the rope (five inches). (Hay, Kennedy, & Levin, 1999)

b. [Over the years,] the pages yellowed. (adapted from Jackendoff, 1983)

In addition to 'event' readings, deadjectival verbs also give rise to non-event readings. As an illustration, take (124a) from (Gawron, 2006a, a).

124. a. The crack widened at the north gate. (Gawron, 2006a, (6a))

---

[109] These verbs are formed from the root form, not the positive form of the adjective.

b. The crack widened from the tower to the north gate. (Gawron, 2006a, (6b))

This sentence has two readings. In the event reading, the crack widened over time. In the state ('extent') reading, the width of the crack changes over space so that it is wider at the north gate than it is elsewhere. Gawron (2006a, b) proposes that, as with other extent predicates such as *cover*, *extend*, and *fill*, verbs such as *widen*, *lengthen*, *redden*, *cool*, *narrow*, and so forth take path arguments. This aspect of the meaning is highlighted in (124b), where the path is given boundaries. Gawron introduces a path function into the semantics of extent predicates, which tracks change over space or time. Scalar predicates, recall, denote measurement functions, and so will track the change in property. For Gawron (2006b), the path function constrains the domain over which the measurement of a change takes place.

We can further observe an interesting difference between different types of deadjectival verbs regarding telicity, which allows us to relate paths and scales. While the sentences in (125) have both a telic and an atelic interpretation, and so can be modified by both *in* and *for* measure phrases, those in (126) have only one interpretation. (126a) is telic, while (126b) is atelic.

125. a. The water cooled down in/for two minutes. (Jackendoff, 1996, (39a))

b. The chemical reddened in/(continuously) for half a minute. (Jackendoff, 1996, (39b))

126. a. The amoeba flattened out in/*for two minutes. (Jackendoff, 1996, (37b))

b. The gap between the boats widened ??in/for a few minutes.

(Kennedy & Levin, to appear, (6))

Here, too, Jackendoff (1996) appeals to path structure and argues that differences in the boundedness of the path accounts for the telicity of the predicate. Abusch (1986) describes the difference by saying that the sense of a verb like *widen* is to *become wider*, while the sense of a verb like *flatten* is to *become flat*. (See Kearns (2007) for a similar claim in which the positive form of the adjective provides the 'standard endstate'.) As in the discussion in the previous section, we again see a difference between actually arriving at a result state expressed by the positive form of the adjective and merely approaching it.

However, neither one of these descriptions of the difference between verbs like *widen* and *flatten* actually accounts for why we observe a difference between these two types of deadjectival verbs in the telicity of the default interpretation. In addition, there are differences in entailment patterns between different types of deadjectival verbs related to telicity. The difference observed in (127) is reminiscent of the difference between telic and atelic descriptions illustrated in (128).

127. a. The soup is cooling, but it hasn't cooled. (Kennedy & Levin, to appear, (8a))

b. #The gap is widening, but it hasn't widened. (Kennedy & Levin, to appear, (7a))

128. a. John is building a house, but the house isn't (completely) built yet.

b. #John is writing in his notebook, there's no writing on the pages.

More recent semantic accounts have focused on the question of why verbs like *flatten* and *cool* differ from those like *widen* and *lengthen*.

By appealing to the scalar structure of the adjective at the core of these verbs, these accounts have been able to offer an explanation for the difference behavior observed in (126)-(127) (cf. Hay, Kennedy, & Levin, 1999; Kearns (2007); Kennedy & Levin, to appear; Winter, to appear). These accounts differ with respect to how the scalar structure is referenced, and also how the possible and default interpretations arise. However, the generalization seems to be that deadjectival verbs formed from GAs mapping to closed scales (e.g., *flat*) have default telic interpretations (i.e., to *flatten* is to 'become flat'), while deadjectival verbs formed from GAs mapping to open scales (e.g., *wide*, *short*) have default atelic interpretations (i.e., to *widen* is to 'become wider') (cf. Hay, Kennedy, & Levin, 1999). Crucially, as Kennedy and Levin (to appear) point out, these verbs entail a change of state, which always requires a minimal starting value, so these verbs should always permit a minimum standard interpretation as long as there was a non-zero amount of change. Thus, even deadjectival verbs formed from maximum standard absolute GAs, which have a default telic interpretation, should permit atelic interpretations in certain circumstances.

Here, though, the maximal value on the scale plays a key role. Recall from the discussion in Chapter 4, that imprecise uses with maximum standard absolute GAs are licensed pragmatically. That is, a container may be judged *full* or a theatre *empty* without either being maximally so. What is required is that the entity be *close enough* to the maximal standard to count as such. For example, the interpretation of *full* in (129a), is clearly close to maximally *full*. When the maximal interpretation of *full* is forced, as in (129b), a negative continuation seems odd.

129. a. I'm full, but those pots de crème look so good, I'm sure I still have room left for a little dessert.

b. #I'm completely full, but I still have room for dessert.

Likewise, with deadjectival verbs, the participant or theme may be affected to a degree that approaches the maximal degree (and must be affected to some degree), without being totally affected (130), a fact which relies upon the homomorphism of part structures described above in § 7.7.2.

130. a. The sky darkened in an hour but it wasn't completely dark.

b. #The sky darkened in an hour but it wasn't dark. (Kearns, 2007, (37a-b))

However, when a maximal interpretation is forced and the theme must be totally affected, as in (131), the maximal value on the scale must hold.

131. #All of the sky darkened in an hour, but it wasn't completely dark.

(Kennedy & Levin, to appear, (14a))

I began this section by sketching out the basic characteristics of paths and demonstrated how including paths in our conceptual representations allows us to capture key features of events: directionality, the unfolding of a sequence of subevents, and measurement of time, space, and motion. I pointed out that paths may have a source and/or a goal bounding the path, with the presence or absence of the latter accounting for the telic/atelic distinction among predicates. I provided examples that demonstrated that path structures are not restricted to events, and can be used to capture spatial but not temporally-bounded situations and metaphors. I then outlined a specific connection between paths and scales, appealing to the features of both, which allow for a homomorphism of their part structures. Events in which a change of state takes place most clearly illustrate this connection: incremental changes in the degree of a property, or the way in which a theme is affected, can track the course of an event. Finally, I discussed three cases in particular involving the interaction between adjective and verb meaning – resultatives, changes of state expressed by positive and comparative forms, and deadjectival verbs – which illustrate the importance of appealing to scalar structure to track properties in an event. These cases therefore also illustrate the intimate connection between scales and paths in our conceptual representations and in the linguistic representations used to talk about these situations.

7.8. Paths and Children's Encoding of Events

In this final section, I pursue what I see is a promising connection, reviewing evidence that path structures play an important role in the way infants and young children parse events. This is an important direction to take, because if the shared aspects of scales and paths that I have discussed are constraints on conceptual representations (Fodor, 1981; Jackendoff, 1992),[110] and if word learning builds on cognitive development (cf. Clark, 2004; Mandler, 2004 for reviews of related work), then they should also constrain the inferences and generalizations that children make from the patterns they observe in language.[111] If children are attending to the structure of paths of events at an early age, then it is highly likely that they are guided by the same kind of conceptual constraints that play an important role in scalar structure: change across a dimension, directionality, open/closed status (boundedness of paths), and so on.

Evidence from a wide range of studies seems to indicate that young children are paying attention to the structure of paths – specifically the source or goal of an event[112] – very early

---

[110] A view that may be compatible with this story is that these conceptual structures arise through a redescription of perceptual information into image schemata (Mandler, 1992). What I have said is that language acquisition is guided early on by conceptual representations, so as long as the structure of paths and scales are part of the infant's conceptual representations, infants should be guided by these structures as they learn the meaning of words. There are a number of differences between this type of account and one that relies on innate primitives, which I will not address here.

[111] See Casasola, Bhaqwat, and Ferguson (2006) for a related discussion of how nonlinguistic and linguistic inputs may interact in children's developing understanding of motion events.

[112] Because this section is about the structure of paths, I am focusing my attention on studies related to those aspects of paths most closely related to scales, i.e., directionality and boundedness. However, other research has shown that infants also encode other aspects of paths. For example, Pulverman and Golinkoff (2004) and Pulverman (2005)

on.[113] For example, in an investigation probing a possible asymmetry in the encoding of goals and sources of paths in normally-developing preschoolers, children with Williams Syndrome[114], and normal adults, Lakusta and Landau (2005) found that all three groups displayed a goal bias. Participants were shown a series of 5-second-long videotapes of people and objects engaged in a variety of events, including 'manner of motion' (e.g., crawling, spinning), 'change of possession' (e.g., throwing/catching, giving/getting), 'change of state' (e.g., color change), and 'attachment/detachment' (e.g., sticking on, ripping off). After each tape, the participants were asked to describe what happened. The dependent measure was the proportion of responses for each event type that included reference to either the goal or the source. The results from the normally-developing preschool participants indicated a reliable effect of source vs. goal across all event types, as illustrated in below (their Figures 7-10).

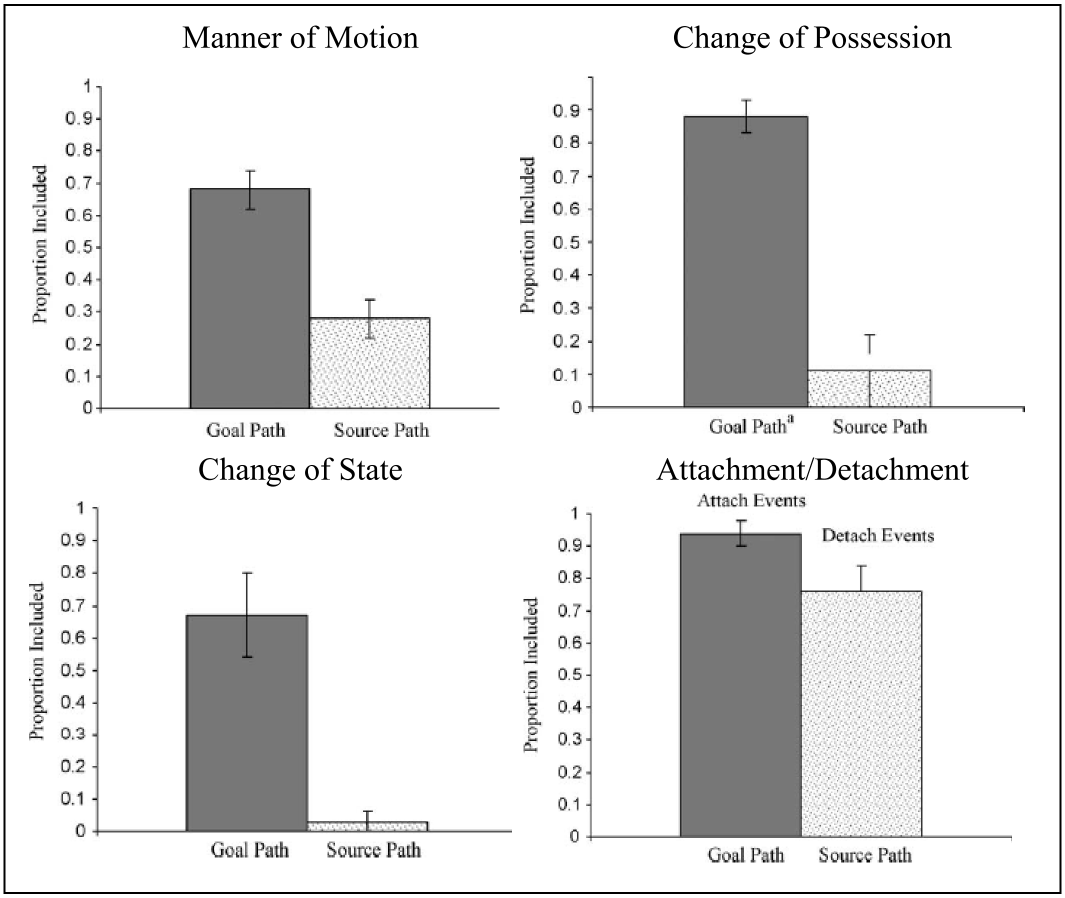
---

demonstrated that infants seven to nine months of age attend to an object's manner of motion and its path in relation to another object (i.e., how a figure moves relative to a ground).

[113] Though I focus on how young children appeal to conceptual structures of paths in their understanding of events of motion here, there is also evidence of young children's knowledge of path structure from early linguistic production. However, since children's production is influenced by their caregiver's input, it is hard to tease apart frequency effects from the influence of other factors. See, however, Bloom (1973), Choi and Bowerman (1991), Clark and Carpenter (1989, 1994), and Clark, Carpenter, and Deutsch (1995) for evidence from child language.

[114] Williams Syndrome is a rare genetic condition involving, among other symptoms, spatial impairment.

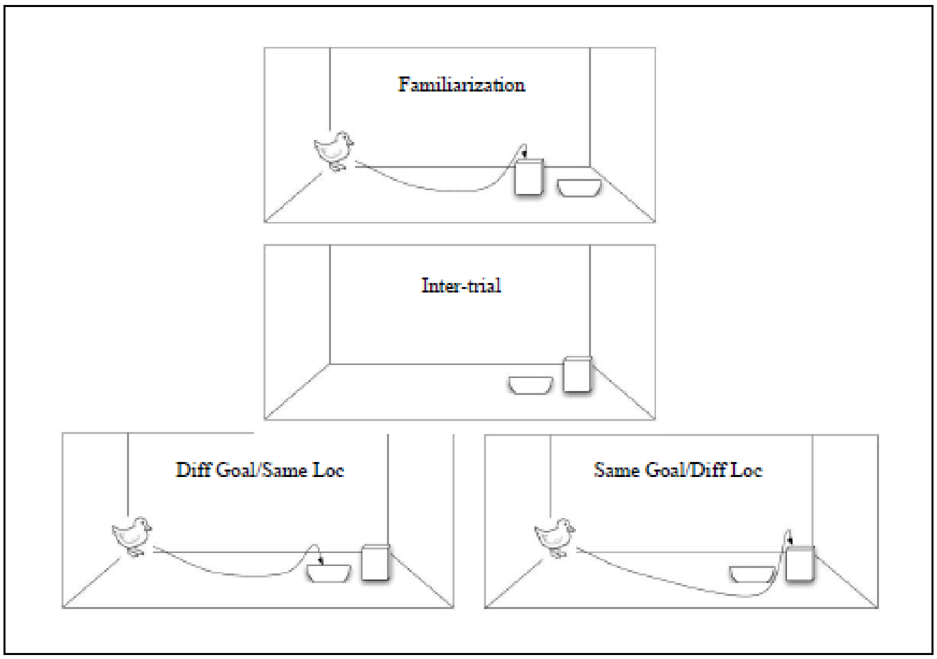
Figure 32: Goal/Source asymmetry in preschool participants across four event types in Experiments 3A and 3B of Lakusta & Landau (2005) (figures adapted from source)



Freeman, Sinha, & Stedmon (1980) have also argued for what they termed an 'allative bias' in 3- and 4-year-olds.

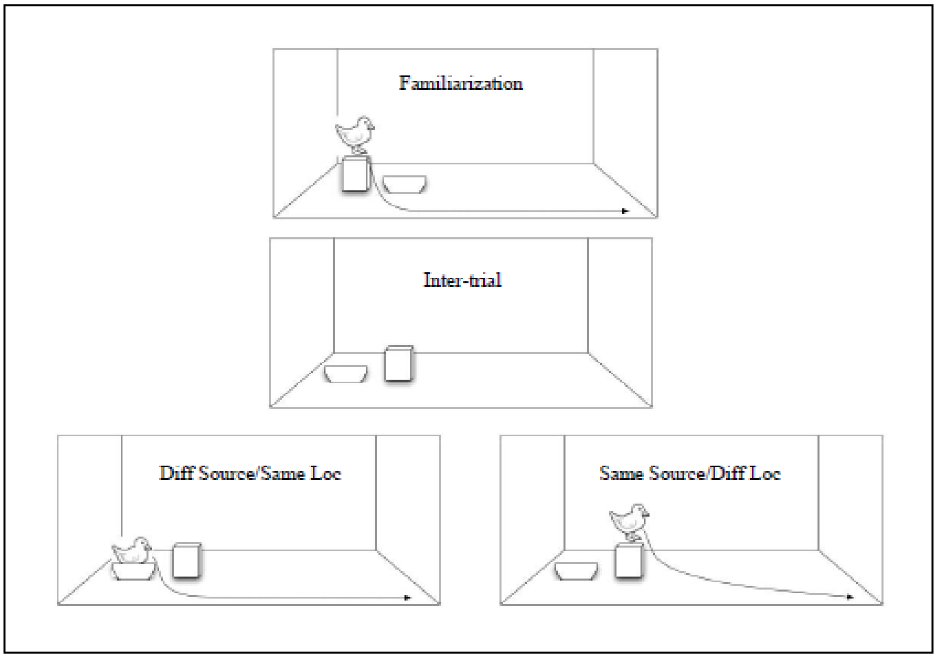
Lakusta (2005) found a similar goal/source asymmetry with 12-month-olds. (See also Lakusta, Wagner, O'Hearn, & Landau (2007).) In her studies, infants were familiarized to a toy duck moving to one of two 'goal' objects (onto a red block or into a green bowl). After familiarization, the curtain was lowered, and the experimenter switched the locations of the two objects. The infants were then shown the two objects in their new locations. Then, during a test trial, infants saw the duck move from the original location to either a different goal in the same location as before or the same goal in a different location. (See Figure 33, her Figure 2.1.) Infants looked longer (by an average of approximately 2 seconds) during the ‘different goal/same location’ trials, indicating a violation of expectation.

Figure 33: Experimental design for Experiment 1 of Lakusta (2005) (figure reproduced from source)



In a control experiment that did not involve motion, but just presentation of the duck with one of the two objects, looking time did not differ significantly. In a second control experiment, in which the duck began on one of the two objects (now ‘source’ objects) and moved to a different location (cf. Figure 34, her Figure 2.5), infants also did not show a difference in looking time between two test conditions.

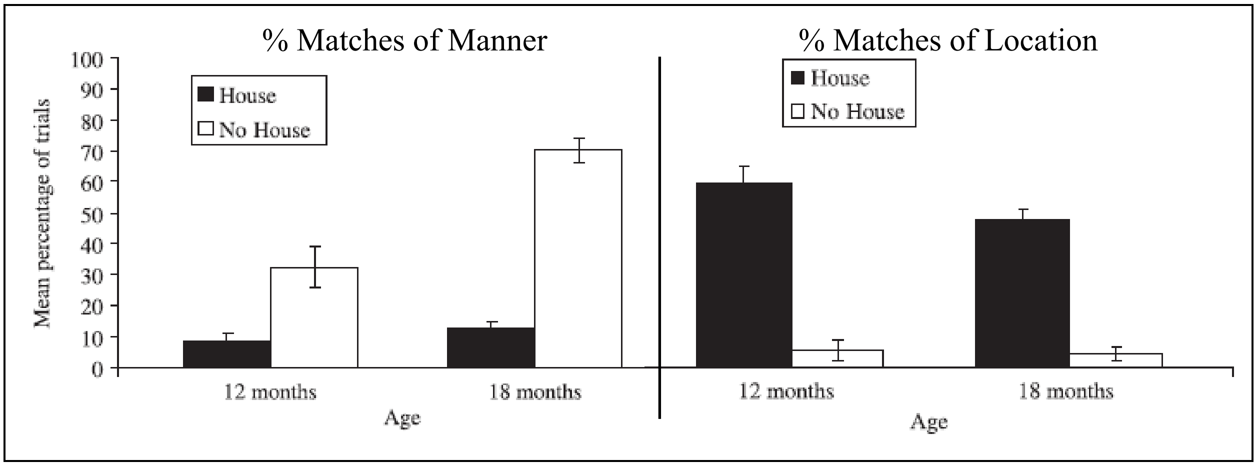
Figure 34: Experimental design for Experiment 2 of Lakusta (2005) (figure reproduced from source)



In Experiment 3, where the sources were made more salient, infants did show a difference in looking time, indicating that they had encoded the source. But when these so-called ‘super’ sources were pitted against the goals from Experiment 1, and infants were shown either a ‘same source/different goal’ sequence or a ‘different source/same goal’ sequence at test, they looked longer in the ‘same source/different goal’ condition by an average of approximately 4.5 seconds. Thus, infants encode both source and goal, but appear to have a bias to attend to the latter.

Carpenter, Call, and Tomasello (2005), following work by Bekkering, Wohlschläger, and Gattis (2000), demonstrated that 12- and 18-month-old infants take into account the presence of a salient goal object when they copy an adult’s actions. In their study, an adult moved a toy mouse across a surface, hopping or sliding it, either into a toy house, or onto a bare mat. When the house was present, infants appeared to interpret the action as ‘putting the mouse in the house’ and were less likely to imitate the manner of motion (i.e., the hopping or sliding action) (cf. Figure 35, their Figures 2 and 5).

Figure 35: Percentage of trials in which infants imitated an adult's manner of motion (hopping or sliding) in Carpenter, Call, & Tomasello (2005) (figure adapted from source)



These findings appear to show that children are more likely to encode the goal of a path than an actor's manner of motion. However, the fact that the difference between the two conditions was the presence or absence of a house may mean that the results say something more about infants' ability to take contextually salient information into account at a very early age, and not necessarily that goals have a privileged status.

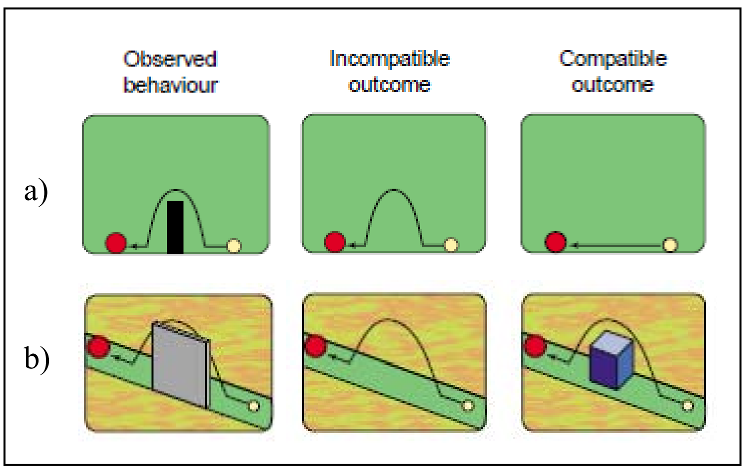
Finally, a number of studies have shown that infants appear to have a goal-biased perspective on events,[115] interpreting an action as a means to an end, and that they infer the presence of a goal even when it is not reached.[116] For example, Gergely, Nádasdy, Csibra, &

[115] Cf. Baldwin, Baird, Saylor, & Clark, 2001; Buresh & Woodward, 2007; Csibra & Gergely, 2007; Csibra, Gergely, Bíró, Koós, & Brockbank, 1999; Gergely, Nádasdy, Csibra, & Bíró 1995; Onishi, 2005; Premack & Premack, 1997; Sodian, 2004; and Woodward, 1998, 2005.

[116] Cf. Csibra , Biro, Koós, & Gergely, 2003; Meltzoff, 1995; and Wagner & Carey, 2005.

Bíró (1995) found that when 12-month-olds observed an initial behavior such as the ones in the far left column in Figure 36 below, they looked longer when they were later shown an ‘incompatible’ outcome rather than one that is compatible with achieving a goal, thereby indicating a violation of expectation. In both cases, the jumping action observed during familiarization is presumably interpreted as the most efficient means to an end, given the presence of a visible (a) or inferred (because it is apparently occluded) (b) obstacle.

Figure 36: Experimental design from Gergely *et al*. (1995) (figure adapted from Gergely & Csibra, 2003)



Woodward (1998) has also shown that six- and nine-month olds have expectations about the goal of a grasping event, looking longer when a human hand takes the same path to grasp a new toy than when it takes a new path to grasp the same toy. However, while these findings are highly relevant to discussions of infants’ theory of mind and intentionality, it is not clear how much they relate to infants’ knowledge of the structure of events (and therefore paths) *per se*. Further evidence is needed to examine just how the conceptual representation of paths influences event representation.

In this section, I have presented evidence from a variety of studies that infants and preschoolers are sensitive to the structure of paths. In the previous section, I reviewed evidence that scales and paths are intimately connected in conceptual and linguistic representations. If infants are attending to aspects of paths in events early on, it seems reasonable to draw the analogy and say that they are also doing the same with scales. This hypothesis should lead to some very interesting investigations in future research.

References

Abney, S. (1987). *The English noun phrase in its sentential aspect*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Abusch, D. (1986). Verbs of change, causation, and time. Technical report 86-50, Stanford University. Stanford, CA: Center for the Study of Language and Information.

Alishashi, A., & Stevenson, S. (2005a). The acquisition and use of argument structure constructions: A Bayesian model. *Proceedings of the ACL 2005 Workshop on Psychocomputational Models of Human Language*

Alishashi, A., & Stevenson, S. (2005b). A probabilistic model of early argument structure acquisition. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*.

Allerton, D. J. (1987). English intensifiers and their idiosyncracies. In R. Steele & T. Threadgold (Eds.), *Language topics. Essays in honour of Michael Halliday* (pp. 15-31). Amsterdam: John Benjamins.

Amaral, P. (2006). On the semantics of *almost*. LSA Annual Meeting, January 5-7, Albuquerque, NM.

Andrick, G., & Tager-Flusberg, H. (1986). The acquisition of colour terms. *Journal of Child Language*, *13*, 119-134.

Aristotle. (350 B.C.E.). *Categories*. Trans. by E. M. Edghill. Text version, provided by The Internet Classics Archive. Downloaded from http://classics.mit.edu//Aristotle/categories.html.

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, *9*, 5-16.

Backsheider, A., & Shatz, M. (1993). Children's acquisition of the lexical domain of color. In J. Beals, G. Cooke, D. Kathman, K. E. McCullough, S. Kita, & D. Testen (Eds.), *What We*

Think, What We Mean, and How We Say it: Papers from the Parasession on the Correspondence of Conceptual, Semantic, and Grammatical Representations, *29th Regional Meeting of the Chicago Linguistic Society* (pp. 11-22). Chicago, IL: Chicago Linguistics Society.

Baldwin, D., Baird, J., Saylor, M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*, 708-717.

Barker, C. (2002). Dynamics of vagueness. *Linguistics and Philosophy*, *25*, 1-36.

Barner, D., & Snedeker, J. (2007). Compositionality and statistics in adjective acquisition: 4-year-olds interpret *tall* and *short* according to the typical sizes of novel noun referents. In H. Caunt-Nulton, S. Kulatilake, & I. Woo (Eds.), *Proceedings of the 31st Annual Boston University Conference on Language Development* (pp. 81-92). Somerville, MA: Cascadilla Press.

Bartlett, E. (1976). Sizing things up: The acquisition of the meaning of dimensional adjectives. *Journal of Child Language*, *3*, 205-219.

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

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

Beavers, J. (2002). Aspect and the distribution of prepositional resultative phrases in English. LinGO working paper no. 2002-07. Stanford, CA: CSLI Publications.

Beavers, J. (2006). *Argument/oblique alternations and the structure of lexical meaning*. Doctoral dissertation, Stanford University, Stanford, CA.

Beavers, J. (in press). Scalar complexity and the structure of events. In J. Dölling & T. Heyde-Zybatow (Eds.), *Event structures in linguistic form and interpretation*. Berlin: Mouton de Gruyter.

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology, 53A*, 153-164.

Bertinetto, P. M., & Squartini, M. (1995). An attempt at defining the class of 'gradual completion verbs'. In P. M. Bertinetto, V. Bianchi, J. Higginbotham, & M. Squartini (Eds.), *Temporal reference, aspect and actionality: Semantic and syntactic perspectives* (Vol. 1) (pp. 11-26). Torino, Italy: Rosenberg and Sellier.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 277-360). New York, NY: Wiley.

Bhat, D. N. S. (1994). *The adjectival category: Criteria for differentiation and identification*. Amsterdam: John Benjamins Publishing Company.

Bierwisch, M. (1967). Some semantic universals of German adjectivals. *Foundations of Language*, *3*, 1-36.

Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation* (pp. 71-262). (1987). *Grammatische und konzeptuelle Aspekte von Dimensionsadjectiven*. *Springer series in language and communication* (Vol. 26). Berlin: Akademie-Verlag, Springer.

Blackwell, A. (1998). *Adjectives in acquisition: On the semantic and syntactic development of English adjectives*. Doctoral dissertation, Boston University, Boston, MA.

Blackwell, A. (2005). Acquiring the English adjective lexicon: Relationships with input properties and adjectival semantic typology. *Journal of Child Language*, *32*, 535-562.

Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax*. The Hague, Mouton.

Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when and why. *Cognitive Psychology*, *6*, 380-420.

Bloom, L., Lightbown, P., & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development, 40*, (Serial No. 160).

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language*, *24*, 511-533.

Boas, H. C. (2000). *Resultative constructions in English and German*. Doctoral dissertation, University of North Carolina, Chapel Hill.

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 defence of the diversity of comparatives and positives. *Linguistiche Berichte*, *36*, 1-9.

Bolinger, D. (1967a). Adjectives in English: Attribution and predication. *Lingua*, *18*, 1-34.

Bolinger, D. (1967b). Adjective comparison: A semantic scale. *Journal of English Linguistics*, *1*, 2-10.

Bolinger, D. (1972). *Degree words*. The Hague: Mouton.

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.

Bornstein, M. (1985). On the development of color naming in young children: Data and theory. *Brain and Language*, *26*, 72-93.

Bresnan, J. W. (1973). Syntax of the comparative clause construction in English. *Linguistic Inquiry*, *4*, 275-473.

Brewer, W., & Stone, J. B. (1975). Acquisition of spatial antonym pairs. *Journal of Experimental Child Psychology*, *19*, 299-307.

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, Evanston, IL.

Buresh, J., & Woodward, A. (2007). Infants track action goals within and across agents. *Cognition*, *104*, 287-314.

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

Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, *8*, F13-F20.

Carston, R. (1998). Informativeness, relevance, and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179-236). Amsterdam: John Benjamins.

Casasola, M., Bhaqwat, J., & Ferguson, K. T. (2006). Precursors to verb learning: Infants' understanding of motion events. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action* *meets word: How children learn verbs* (pp. 160-190). New York, NY: Oxford University Press.

Cassidy, B. A. (1980). *Age and the comprehension of spatial adjectives*. Masters Thesis, Portland State University.

Chierchia, G. (2001). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. Ms., University of Milan.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 157-168). Somerville, MA: Cascadilla Press.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition, 41*, 83-121.

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language*, *22*, 497-529.

Cinque, G. (1994). On the evidence for partial N movement in the Romance DP. In G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi, & R. Zanuttini (Eds.), *Paths toward Universal Grammar* (pp. 85-110). Georgetown: Georgetown University Press.

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. New York, NY: Oxford University Press

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

Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition*, *64*, 1-37.

Clark, E. V. (2004). How language acquisition builds on cognitive development. *TRENDS in Cognitive Science*, *Language and Conceptual Development Series*, *8*, 472-478.

Clark, E. V. (2006). Color, reference, and expertise in language acquisition. *Journal of Experimental Child Psychology*, 94, 339-343.

Clark, E. V., & Carpenter, K. L. (1989). On children's uses of *from*, *by*, and *with* in oblique noun phrases. *Journal of Child Language, 16*, 349-364.

Clark, E. V., & Carpenter, K. L. (1994). The notion of source in language acquisition. In P. Bloom (Ed.), *Language acquisition: Core readings* (pp. 251-284). Cambridge: MIT Press.

Clark, E. V., Carpenter, K. L., & Deutsch, W. (1995). Reference states and reversals: Undoing actions with verbs. *Journal of Child Language, 22*, 633-662.

Clark, H. (1970). The primitive nature of children's relational concepts. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 269-278). New York, NY: Wiley.

Clark, H. (1973). Space, time, semantics, and the child. In T. Moore (Ed.), *Cognitive development and the acquisition of* language (pp. 27-63). New York, NY: Academic Press.

Clark, R. (1970). Concerning the logic of predicate modifiers. *Noûs*, *4*, 311-355.

Coley, J. D., & Gelman, S. A. (1989). The effects of object orientation and object type on children's interpretation of the word *big*. *Child Development*, *60*, 372-380.

Coots, J. H. (1976). *Children's knowledge and acquisition of polar spatial adjectives*. Doctoral dissertation, University of Utah, Salt Lake City, UT.

Corver, N. (1990). The syntax of left branch constructions. Doctoral dissertation, Tilburg University, The Netherlands.

Crawford, L. E., Regier, T., & Huttenlocher, J. (2000). Linguistic and non-linguistic spatial categorization. *Cognition*, *75*, 209-235.

Cresswell, M. J. (1976). The semantics of degree. In B. H. Partee (Ed.), *Montague grammar* (pp. 261-292). New York, NY: Academic Press

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.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.

Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*, 60-78.

Csibra, G., Biro, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*, 111-133.

Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of 'pure reason' in infancy. *Cognition*, *72*, 237-267.

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

Dixon, R. M. W. (1977). Where have all the adjectives gone? *Studies in Language*, 1, 19-80.

Donaldson, M., & Wales, R. J. (1970). On the acquisition of some relational terms. In J. R. Hayes & R. Brown (Eds.), *Cognition and the development of language* (pp. 235-268). New York, NY: John Wiley and Sons, Inc.

Dowty, D. R. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.

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

Dromi, E. (1987). *Early lexical development*. New York, NY: Cambridge University Press.

Dummett, M. (1975). Wang's paradox. *Synthèse*, *30*, 301-324.

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.

Edgington, D. (2001). The philosophical problem of vagueness. *Legal Theory*, *7*, 371-378.

Ehri, L. (1976). Comprehension and production of adjectives and seriation. *Journal of Child Language*, *3*, 369-384.

Eilers, R. E., Kimbrough Oller, D., & Ellington, J. (1974). The acquisition of word-meaning for dimensional adjectives: The long and short of it. *Journal of Child Language*, *1*, 195-204.

Engel, R. (1989). On degrees. *The Journal of Philosophy*, *86*, 23-37.

Eschenbach, C. (2000). Geometric structures of frames of reference and natural language semantics. *Spatial Cognition and Computation*, *1*, 329-348.

Eschenbach, C., Tschander, L., Habel, C., & Kulik, L. (2000). Lexical specifications of paths. In W. B. C. Freksa, C. Habel, & K. Wender (Eds.), S*patial cognition II: Integrating abstract theories, empirical studies, formal methods, and practical applications* (pp. 127-144). Berlin: Springer.

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

Fauconnier, G. (1975). Pragmatic scales and logical structure. *Linguistic Inquiry*, *6*, 353-375.

Feider, H. (1973). Comparatives in early child language. *Glossa*, *7*, 3-20.

Filip, H. (2007). Events and maximalization: The case of telicity and perfectivity. Ms., University of Florida, Gainesville, FL.

Fillmore, C. (1968a). The case for case. In E. Bach & R. Harms (Eds.), *Universals of linguistic theory* (pp. 1-91). New York, NY: Holt, Rinehart & Winston.

Fillmore, C. (1968b). Lexical entries for verbs. *Foundations of Language*, *4*, 373-393.

Fillmore, C. J. (1971). Entailment rules in a semantic theory. In J. F. Rosenberg & C. Travis (Eds.), *Readings in the philosophy of language* (pp. 533-547). Englewood Cliffs, NJ: Prentice-Hall.

Finch-Williams, A. (1981). *Biggerest or biggester: A study in children s acquisition of linguistic features of comparative adjectives and nonlinguistic knowledge of seriation*. Doctoral dissertation, University of Kansas, Lawrence, KS.

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, *30*, 265-300.

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, *92*, 333-375.

Fodor, J. (1981). The present status of the innateness controversy. In *Representations* (pp. 257-315). Cambridge, MA: MIT Press.

Fox, D., & Hackl, M. (2006). The universal density of measurement. *Linguistics and Philosophy*, *29*, 537-586.

Frazier, L., Clifton, Jr., C., & Stolterfoht, B. (to appear). Scale structure: Processing minimum standard and maximum standard scalar adjectives. *Cognition*.

Freeman, N. H., Sinha, C. G., & Stedmon, J. A. (1980). The allative bias in three-year-olds is almost proof against task naturalness. *Journal of Child Language*, *8*, 283-296.

Gasser, M., & Smith, L. (1998). Learning noun and adjective meanings: A connectionist account. *Language and Cognitive Processes*, *13*, 269-306.

Gathercole, V. C. (1979). *Birdies like birdseed the bester than buns: A study of relational comparatives and their acquisition*. Doctoral dissertation, University of Kansas, Lawrence, KS.

Gathercole, V. C. (1982). Decrements in children's responses to *big* and *tall*: A reconsideration of the potential cognitive and semantic causes. *Journal of Experimental Child Psychology*, *34*, 156-173.

Gawron, J. M. (2006a). Generalized paths. In E. Georgala & J. Howell (Eds.), *Proccedings of Semantics and Linguistic Theory XV*. Ithaca, NY: CLC Publications.

Gawron, J. M. (2006b). Paths and scalar change. Paper presented at the Chicago Workshop on Scalar Meaning, May 19-20, Chicago, IL.

Gazdar, G. (1979). *Pragmatics*. New York, NY: Academic Press.

Geach, P. (1956). Good and evil. *Analysis*, *17*, 33-42.

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., & Taylor, M. (1984). How two-year-old children interpret proper and common names for unfamiliar objects. *Child Development*, *55*, 1535-1540.

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.

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, *49*, 988-998.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kučzaj II (Ed.), *Language development* (Vol. 2)*: Language, thought and culture* (pp. 301-334). Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215-256). Cambridge, UK: Cambridge University Press.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *TRENDS in Cognitive Science*, *7*, 287-292.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165-193.

Geurts, B. (1998). Scalars. In P. Ludewig & B. Geurts (Eds.), *Lexicalische Semantik aus kognitiver Sicht* (pp. 95-117). Tübingen: Narr Verlag.

Gillette, J., Gleitman, H., Gleitman, L. R., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*, 135-176.

Gitterman, D., & Johnston, J. R. (1983).Talking about comparisons: A study of young children's comparative adjective usage. *Journal of Child Language*, *10*, 605-621.

Givón, T. (1970). Notes on the semantic structure of English adjectives. *Language*, *46*, 816-837.

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

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*, 23-64.

Goede, K. (1989). Language acquisition and development of children's *bigger* and *more* judgements. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation* (pp. 419-432). (1987). *Grammatische und konzeptuelle Aspekte von Dimensionsadjectiven*. *Springer series in language and communication* (Vol. 26). Berlin: Akademie-Verlag, Springer.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Goldberg, A. E., & Jackendoff, R. (2004). The English resultative as a family of constructions. *Language*, *80*, 532-568.

Goldberg, A. E., Casenhiser, D., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, *15*, 289-316.

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

Goldin-Meadow, S., Seligman, M., & Gelman, R. (1976). Language in the two-year-old. *Cognition*, *4*, 189-202.

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.

Graham, S., Cameron, C., & Welder, A. (2005). Preschoolers' extension of familiar adjectives. *Journal of Experimental Child Psychology*, *91*, 205-226.

Graham, S., Welder, A., & McCrimmon, A. (2003). Hot dogs and zavy cats: Preschoolers' and adults' expectations about familiar and novel adjectives. *Brain and Language*, *84*, 16-37.

Graziano-King, J. (1999). *Acquisition of comparative forms in English*. Doctoral dissertation, City University of New York.

Graziano-King, J., & Cairns, H. S. (2005). Acquisition of English comparative adjectives. *Journal of Child Language*, *2*, 345-373.

Green, G. (1972). Some observations on the syntax and semantics of instrumental verbs. In P. M. Peranteau, J. N. Levi, & G. C. Phares (Eds.), *Papers from the 8th Regional Meeting of the Chicago Linguistic Society* (pp. 83-97). Chicago, IL: Chicago Linguistic Society.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics 3* (pp. 41-58). New York, NY: Academic Press.

Grimshaw, J. (1982). Subcategorization and grammatical relations. In A. Zaenen (Ed.), *Subjects and other subjects: Proceedings of the Harvard conference on the representation of grammatical relations* (pp. 35-55). Bloomington, IN: Indiana University Linguistics Club.

Grimshaw, J. (1991). Extended projections. Ms., Brandeis University, Waltham, MA.

Gruber, J. S. (1965). *Studies in lexical relations*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA. Reprinted (1970) by Indiana University Linguistics Club, Bloomington, IN.

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001). At the semantics/pragmatics interface in child language. In *Proceedings of Semantics and Linguistic Theory XI* (pp. 231-247). Ithaca, NY: CLC Publications.

Hacquard, V. (2006). Aspects of *too* and *enough* constructions. In E. Georgala & J. Howell (Eds.), *Proceedings of Semantics and Linguistic Theory XV*. Ithaca, NY: CLC Publications.

Hale, K., Ihọnụ, Ụ. P., & Manfredi, V. (1995). Igbo bipositional verbs in a syntactic theory of argument structure. In A. Akinlabi (Ed.), *Theoretical approaches to African linguistics* (pp. 83-107). Trenton, NJ: Africa World Press.

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.

Hallett, S. (1976). *Over-extension phenomena in children's acquisition of spatial adjectives*. Doctoral Dissertation, Indiana University, Bloomington, IN.

Hamann, C. (1991). Adjectives. In A. von Stechow & D. Wunderlich (Eds.),*Semantik: Ein internationals Handbuch der zeitgenössischen Forschung* (pp. 657-673). Berlin: Walter de Gruyter.

Hare, R. M. (1952). *The language of morals*. Oxford: Oxford University Press.

Harris, P. L., & Folch, L. (1985). Decrement in the understanding of *big* among English- and Spanish-speaking children. *Journal of Child Language*, *12*, 685-690.

Harris, P. L., Morris, J. E., & Meerum Terwogt, M. (1986). The early acquisition of spatial adjectives: A cross-linguistic study. *Journal of Child Language*, *13*, 335-352.

Hay, J., Kennedy, C., & Levin, B. (1999). Scalar structure underlies telicity in ‘degree achievements’. In T. Matthews & D. Strolovitch (Eds.), *Proceedings of Semantics and Linguistic Theory IX* (pp. 127-144). Ithaca, NY: CLC Publications.

Heim, I. (1985). Notes on comparatives and related matters. Ms. University of Texas, Austin.

Heim, I. (1990). E-type pronouns and donkey anaphora. *Linguistics and Philosophy*, *13*, 137-177.

Heim, I. (2000). Degree operators and scope. In B. Jackson & T. Matthews (Eds.), *Proceedings of Semantics and Linguistic Theory X* (pp. 40-64). Ithaca, NY: CLC Publications.

Hellan, L. (1981). *Towards an integrated analysis of comparatives*. Tübingen: Günter Narr Verlag.

Higginbotham, J. (1985). On semantics. *Linguistic Inquiry*, *16*, 547-593.

Hirschberg, J. (1985). A theory of scalar implicature. Doctoral dissertation, University of Pennsylvania, Philadelphia, PA.

Hitzeman, J. (1992). The selectional properties and entailments of *almost*. In C. Canakis, G. Chan, & J. Denton (Eds.), *Papers from the 28th Regional Meeting of the Chicago Linguistic Society* (pp. 225-238). Chicago, IL: Chicago Linguistic Society.

Hoeksema, J. (1983). Negative polarity and the comparative. *Natural Language & Linguistic Theory*, *1*, 403-434.

Hollich, G. (2003). SuperCoder [Computer program]. Retrieved from http://hincapie.psych.purdue.edu/Splitscreen/SuperCoder.dmg.

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.

Horn, L. (1972). On the semantic properties of the logical operators in English. Doctoral dissertation, University of California, Los Angeles. [Distributed by Indiana University Linguistics Club]

Huang, Y., Snedeker, J., & Spelke, R. (submitted). What exactly do numbers mean? Ms., Harvard University, Cambridge, MA, paper presented at the 26th Annual Meeting of the Cognitive Science Society, August 5-7, Chicago, IL.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*, 597-626.

Hurewitz, F., Papafragou, A., Gleitman, L. R., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development*, *2*, 77-96.

Huttenlocher, J., & Higgins, E. (1971). Adjectives, comparatives, and syllogisms. *Psychological Review*, *78*, 487-504.

Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review*, *90*, 72-93.

Izvorski, R. (1995). A solution to the subcomparative paradox. In J. Camacho, L. Choueiri, & M. Watanabe (Eds.), *Proceedings of the 14th West Coast Conference on Formal Linguistics* (pp. 203-219). Stanford, CA: CSLI Publications.

Jackendoff, R. (1978). Grammar as evidence for conceptual structure. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 201-228). Cambridge, MA: MIT Press.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1991). Parts and boundaries. *Cognition*, *41*, 9-45.

Jackendoff, R. (1992). *Languages of the mind: Essays on mental representation*. Cambridge, MA: MIT Press.

Jackendoff, R. (1996). The proper treatment of measuring out, telicity, and perhaps even quantification in English. *Natural Language and Linguistic Theory*, *14*, 305-354.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reasoning*. Chicago, IL: University of Chicago Press.

Kadmon, N. (1990). Uniqueness. *Linguistics and Philosophy*, *13*, 273-323.

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

Kamp, H., & Roßdeutscher, A. (1992). Remarks on lexical structure, DRS-construction and lexically driven inferences. *Arbcitspapiere des Sonderforschungsbereichs*, *340*. IMS, Universität Stuttgart.

Kamp, J. A. W. (1975). Two theories of adjectives. In E. Keenan (Ed.), *Formal semantics of natural language* (pp. 123-155). Cambridge: Cambridge University Press.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. *Cambridge studies in linguistics* (Vol. 24). New York, NY: Cambridge University Press.

Katz, J. (1964). Semantic theory and the meaning of *good*. *The Journal of Philosophy*, *61*, 739-766.

Katz, J. (1967). Recent issues in semantic theory. *Foundations of Language*, *3*, 124-194.

Katz, J. (1972). *Semantic theory*. New York, NY: Harper and Row.

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.

Kearns, K. (2007). Telic senses of deadjectival verbs. *Lingua*, *117*, 26-66.

Keil, F., & Carroll, J. (1980). The child's acquisition of *tall*: Implications for an alternative view of semantic development. *Papers and Reports on Child Language Development*, *19*, 21-28.

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. (2001). Polar opposition and the ontology of 'degrees'. *Linguistics and Philosophy*, *24*, 33-70.

Kennedy, C. (2006). Vagueness, imprecision and scale structure. Paper presented at the Chicago Workshop on Scalar Meaning, May 19-20, Chicago, IL.

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

Kennedy, C., & Levin, B. (to appear). Measure of change: The adjectival core of degree achievements. In L. McNally & C. Kennedy (Eds.), *Adjectives and adverbs: Syntax, semantics and discourse*. Oxford: Oxford University Press.

Kennedy, C., & McNally, L. (1999). From event structure to scale structure: Degree modification in deverbal adjectives. In T. Matthews & D. Strolovitch (Eds.), *Proceedings of Semantics and Linguistic Theory IX* (pp. 163-180). Ithaca, NY: CLC Publications.

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

Kim, M. (2002). Does Korean have adjectives? In T. Ionin, H. Ko, & A. Nevins (Eds.), *Proceedings of HUMIT 2001*, *MIT Working Papers in Linguistics* (Vol. 43) (pp. 71-89). Cambridge, MA: MIT.

Klatzky, R. L., Clark, E. V., & Macken, M. (1973). Asymmetries in the acquisition of polar adjectives: Linguistic or conceptual? *Journal of Experimental Child Psychology*, *16*, 32-46.

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

Klein, E. (1982). The interpretation of adjectival comparatives. *Journal of Linguistics*, *18*, 113-136.

Klein, E. (1991). Comparatives. In A. von Stechow & D. Wunderlich (Eds.), *Semantik: Ein internationals Handbuch der zeitgenössischen Forschung* (pp. 673-691). Berlin: Walter de Gruyter.

Klibanoff, R. S. (2000). *Conceptual and perceptual factors contributing to the role of basic level categories in children's acquisition of adjectives*. Doctoral dissertation, Northwestern University, Evanston, IL.

Klibanoff, R. S., & Waxman S. R. (1998). Preschoolers' acquisition of novel adjectives and the role of basic-level kind. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development* (pp. 442-453). Somerville, MA: Cascadilla Press.

Klibanoff, R. S., & Waxman S. R. (2000). Basic level object categories support the acquisition of novel adjectives: Evidence from preschool-aged children. *Child Development*, *71*, 649-659.

Kowalski, K., & Zimiles, H. (2006). The relation between children's conceptual functioning with color and color term acquisition. *Journal of Experimental Child Psychology*, 94, 301-321.

Krämer, I. (2005). When does *many* mean a lot? Discourse pragmatics of the weak-strong distinction. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 353-364). Somerville, MA: Cascadilla Press.

Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. In I. Sag & A. Szabolcsi (Eds.), *Lexical matters* (pp. 29-54). Stanford, CA: CSLI Publications.

Krifka, M. (1998). The origins of telicity. In S. Rothstein (Ed.), *Events and grammar* (pp. 197-235). Dordrecht: Kluwer.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Lakusta, L. (2005). *Source and goal asymmetry in non-linguistic motion event representations*. Doctoral dissertation, The Johns Hopkins University, Baltimore, MD.

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, *96*, 1-33.

Lakusta, L., Wagner, L., O'Hearn, K., & Landau, B. (2007). Conceptual foundations of spatial language: Evidence for a goal bias in infants. *Language, Learning and Development*, *3*, 179-197.

Landau, B., & Jackendoff, R. (1993). *What* and *where* in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*, 217-265.

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

Landman, F. (2000). *Events and plurality: The Jerusalem lectures. Studies in linguistics and philosophy* (Vol. 76). Dordrecht: Kluwer.

Langacker, R. (1987). *Foundations of cognitive grammar* (Vol. 1). Stanford, CA: Stanford University Press.

Langford, C. H. (1968). The notion of analysis in Moore's philosophy. In P.A. Schilpp (Ed.), *The philosophy of G.E. Moore* (3rd ed., first published, 1942) (pp. 319-342). Evanston, IL: Northwestern University Press.

Larson, R. (1995). Olga is a beautiful dancer. Paper presented at the LSA Annual Meeting, January 5-8, New Orleans, LA.

Lasersohn, P. (1999). Pragmatic halos. *Language*, *75*, 522-551.

Layton, T. L., & Stick, S. L. (1978). Comprehension and production of comparatives and superlatives. *Journal of Child Language*, *6*, 511-527.

Lee, J., & Naigles, L. (2005). The input to verb learning in Mandarin Chinese: A role for syntactic bootstrapping. *Developmental Psychology*, *41*, 529-540.

Levin, B., & Rappaport Hovav, M. (1991). Wiping the slate clean: A lexical semantic exploration. *Cognition*, *41*, 123-155.

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use and the interpretation of language* (pp. 303-323). Berlin: Walter de Gruyter.

Link, G. (1987). Algebraic semantics of event structures. In J. Groenendijk, M. Stokhof, & F. Veltman (Eds.), *Proceedings of the Sixth Amsterdam Colloquium* (pp. 243-262). Amsterdam: ITLI.

Liu, J., Golinkoff, R. M., & Sak, K. (2001). One cow does not an animal make: Young children can extend novel words at the superordinate level. *Child Development*, *72*, 1674-1694.

Ludlow, P. (1989). Implicit comparison classes. *Linguistics and Philosophy*, *12*, 519-533.

Lumsden, E. A., & Poteat, B. W. S. (1968). The salience of the vertical dimension in the concept of *bigger* in five- and six-year-olds. *Journal of Verbal Learning and Verbal Behavior* 7: 404-408.

Macnamara, J. (1982). *Names for things: A study of human learning*. Cambridge, MA: MIT Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mandler, J. (1992). How to build a baby II: Conceptual primitives. *Psychological Review*, *99*, 587-604.

Mandler, J. (2004). Thought before language. *TRENDS in Cognitive Science*, *Language and Conceptual Development Series*, *8*, 508-513.

Manning, C., & Schütze, H. (1999). *Foundations of statistics natural language processing*. Cambridge, MA: MIT Press.

Maratsos, M. (1973). Decrease in the understanding of the word *big* in preschool children. *Child Development*, *44*, 747-752.

Maratsos, M. (1974). When is a high thing the big one? *Developmental Psychology*, *10*, 367-375.

Maratsos, M. (1976). *The use of definite and indefinite reference in young children: An experimental study of semantic acquisition*. New York, NY: Cambridge University Press.

Maratsos, M. (1988). Crosslinguistic analysis, universals, and language acquisition. In F. Kessel (Ed.), *The development of language and language researchers: Essays in honor of Roger Brown* (pp. 121-152). Hillsdale, NJ: Lawrence Erlbaum.

Markman, E., & Wachtel, G. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology*, *20*, 121-157.

Markowitz, J. R. (1975). *The acquisition of spatial adjectives in their nominal, comparative and superlative forms among moderately retarded children*. Doctoral dissertation, Indiana University, Bloomington, IN.

Marschark, M. (1977). Lexical marking and the acquisition of relational size concepts. *Child Development*, *48*, 1049-1051.

McCawley, J. D. (1968). The role of semantics in grammar. In In E. Bach & R. Harms (Eds.), *Universals of linguistic theory* (pp. 125-170). New York, NY: Holt, Rinehart & Winston.

McConnell-Ginet, S. (1973). *Comparative constructions in English: A syntactic and semantic analysis*. Doctoral dissertation, University of Rochester, NY.

McConnell-Ginet, S. (1979). On the deep (and surface) adjective *good*. In L. Waugh & F. van Coetsem (Eds.), *Contributions to grammatical studies: Semantics and syntax* (pp. 132-150). Cornell Linguistics Contributions II. Leiden: E. J. Brill.

Meier, C. (2003). The meaning of *too, enough* and *so… that. Natural Language Semantics*, *11*, 69-107.

Meltzoff, A. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*, 838-850.

Meroni, L. (2006). *Putting children in context: Experimental studies on children's interpretation of definite noun phrases*. Doctoral dissertation, University of Maryland, College Park.

Miller, K., Schmitt, C., Chang, H., & Munn, A. (2005). Young children understand *some* implicatures. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 389-400). Somerville, MA: Cascadilla Press.

Mintz, T. (2005). Linguistic and conceptual influences on adjective acquisition in 24- and 36-month olds. *Developmental Psychology*, *41*, 17-29.

Mintz, T., & Gleitman, L. R. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition*, *84*, 267-293.

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.

Moore, D. (1999). Comparatives and superlatives: Lexical before functional. In A. Greenhill, H. Littlefield, & C. Tano (Eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development* (pp. 474-481). Somerville, MA: Cascadilla Press.

Morzycki, M. (2006). Atelicity and measure phrases: Licensing measure phrase modification across AP, PP, and VP. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics* (pp. 279-287). Somerville, MA: Cascadilla Proceedings Project.

Morzycki, M. (2006). Size adjectives and adnominal degree modification. In E. Georgala & J. Howell (Eds.), *Proceedings of Semantics and Linguistic Theory XV*. Ithaca, NY: CLC Publications.

Munn, A., Miller, K., & Schmitt, C. (2006). Maximalilty and plurality in children's interpretations of definites. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 377-387). Somerville, MA: Cascadilla Press.

Murphy, G. (1982). Cue validity and levels of categorization. *Psychological Bulletin*, *91*, 174-177.

Musolino, J. (2004). The semantics and acquisition of number words: Integrating linguistic and developmental perspectives. *Cognition*, *93*, 1-41.

Nadig, A., Sedivy, J., Joshi, A., & Bortfeld, H. (2003). The development of discourse constraints on the interpretation of adjectives. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 568-79). Somerville, MA: Cascadilla Press.

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

Naigles, L. (2002). Form is easy, meaning is hard: Resolving a paradox in early child language. *Cognition*, *86*, 157-199.

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.

Neale, S. (1990). *Descriptions*. Cambridge, MA: MIT Press.

Nelson, K. (1976). Some attributes of adjectives used by young children. *Cognition*, *4*, 1-31.

Nelson, K., & Benedict, H. (1974). The comprehension of relative, absolute, and contrastive adjectives by young children. *Journal of Psycholinguistic Research*, *3*, 333-342.

Ninio, A. (1999a). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language*, *26*, 619-653.

Ninio, A. (1999b). Learning from examples in syntactic development. Keynote address at the 11th Australasian Human Development Conference, Sydney University, July 9, Sydney, Australia.

Noveck, I. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165-188.

O'Hanlon, C. G., & Roberson, D. (2006). Learning in context: Linguistic and attentional constraints on children's color term learning. *Journal of Experimental Child Psychology*, *94*, 274-300.

Olawsky, K. (2004). What is a noun? What is an adjective? Problems of classification in Dagbani. *Journal of African Languages and Linguistics*, *25*, 127-148.

Onishi, K. (2005). Infants' understanding of goal directed behavior. Doctoral dissertation, University of Illinois, Urbana-Champaign.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *86*, 253-282.

Papafragou, A., & Schwarz, N. (2005-2006). Most wanted. *Language Acquisition: A Journal of Developmental Linguistics*, *13*, 207-251.

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition: A Journal of Developmental Linguistics*, *12*, 71-82.

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: Reidel.

Piccin, T., & Waxman, S. R. (in press). Why nouns trump verbs in word learning: New evidence from children and adults in the Human Simulation Paradigm. *Language Learning and Development*.

Platts, M. (1979). *Ways of meaning: An introduction to a philosophy of language*. London: Routledge & Kegal Paul.

Prasada, S. (1992). *Acquisition of adjective meanings: A lexical semantic approach*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Prasada, S., & Cummins, M. (2001). Structural constraints on the interpretation of novel count nouns. In A. H-J. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 623-632). Somerville, MA: Cascadilla Press.

Premack, D., & Premack, A. (1997). Infants attribute value ± to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience*, *9*, 848-856.

Pulverman, R. (2005). *The relationship between language development and event processing: Lexical acquisition and attention to manner and path*. Doctoral dissertation, University of Delaware, Newark, DE.

Pulverman, R., & Golinkoff, R. M. (2004). Seven-month-olds' attention to potential verb referents in nonlinguistic events. In A. Brugos, L. Micciulla, & C.E. Smith (Eds.), *Proceedings of the 28th Annual Boston University Conference on Language Development* (pp. 473-480). Somerville, MA: Cascadilla Press.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Raffman, D. (1994). Vagueness without paradox. *The Philosophical Review*, *103*, 41-74.

Ramchand, G. (1997). *Aspect and predication*. Oxford: Clarendon Press.

Rappaport Hovav, M., & Levin, B. (2001). An event structure account of English resultatives. *Language*, *77*, 766-797.

Ravn, K. E., & Gelman, S. A. (1984). Rule usage in children's understanding of *big* and *little*. *Child Development*, *55*, 2141-2150.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, MA: MIT Press.

Rett, J. (2007). *Degree modification in natural language*. Doctoral dissertation, Rutgers, The State University of New Jersey.

Rips, L., & Turnbull, W. (1980). How big is big? Relative and absolute properties in memory. *Cognition*, *8*, 145-174.

Roberts, C. (2003). Uniqueness in definite noun phrases. *Linguistics and Philosophy*, *26*, 287-350.

Rolf, B. (1984). Sorites. *Synthese*, *58*, 219-250.

Ross, W. D. (1930). *The right and the good*. Oxford: Clarendon 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.

Russell, B. (1905). On denoting. *Mind*, *14*, 479-493.

Ryalls, B. O. (2000). Dimensional adjectives: Factors affecting children's ability to compare objects using novel words. *Journal of Experimental Child Psychology 76*, 26-49.

Ryalls, B. O., Winslow, E., & Smith, L. B. (1998). A semantic congruity effect in children's acquisition of high and low. *Journal of Memory and Language*, *39*, 543-557.

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. In K. Nelson (Ed.), *Children's language* (Vol. 4). Hillsdale, NJ: Lawrence Erlbaum Associates.

Samuelson, L., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *73*, 1-33.

Sandhofer, C. M. (2002). *Language input and the process of learning words: Evidence from dimensional adjectives*. Doctoral dissertation, Indiana University, Bloomington, IN.

Sandhofer, C. M., & Smith, L .B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology*, *35*, 668-679.

Sandhofer, C. M., & Smith, L .B. (2001). Why children learn color and size words so differently: Evidence from adults' learning of artificial terms. *Journal of Experimental Psychology*, 130, 600-620.

Sandhofer, C. M., & Thom, E. (2006). Taking the task seriously: Reflections on measures of color acquisition. *Journal of Experimental Child Psychology*, *94*, 344-348.

Sapir, E. (1944). Grading: A study in semantics. *Philosophical Science*, *11***,** 93-116. Reprinted in D. G. Mandelbaum (Ed.), *Selected writings of Edward Sapir in language, culture, and personality* (pp. 122-149). (1949). Berkeley, CA: University of California.

Schwartz, S. P. (1977). Introduction. In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds* (pp. 13-41). Ithaca, NY: Cornell University Press.

Schwarzschild, R. (2002). The grammar of measurement. In B. Jackson (Ed.), *Proceedings of Semantics and Linguistic Theory XII* (pp. 225-245). Ithaca, NY: CLC Publications.

Schwarzschild, R. (2005). Measure phrases as modifiers of adjective. *Recherches Linguistiques de Vincennes*, *35*, 207-228.

Schwarzschild, R., & Wilkinson, K. (2002). Quantifiers in comparatives: A semantics of degree based on intervals. *Natural Language Semantics*, *10*, 1-41.

Sena, R., & Smith, L. B. (1990). New evidence on the development of the word *big*. *Child Development*, *61*, 1034-1052.

Sera, M., & Smith, L. B. (1987). *Big* and *little*: ‘Nominal’ and relative uses. *Cognitive Development*, *2*, 89-111.

Seuren, P. (1978). The structure and selection of positive and negative gradable adjectives. In D. Farkas, W., M. Jacobsen, & K. W. Todrys (Eds.), *Papers from the Parasession on the Lexicon, 14th Regional Meeting of the Chicago Linguistic Society* (pp. 336-46). Chicago, IL: Chicago Linguistic Society.

Sharpe, D., Fonte, I., & Christe, E. (1998). Big mice, big animals, big problems: the acquisition of adjective interpretation rules. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development* (pp. 675-683). Somerville, MA: Cascadilla Press.

Siegel, M. E. A. (1976). *Capturing the adjective*. Doctoral dissertation, University of Massachusetts, Amherst.

Siegel, M. E. A. (1979). Measure adjectives in Montague grammar. In S. Davis & M. Mithun (Eds.), *Linguistics, philosophy, and Montague grammar*. Conference on Montague grammar, philosophy, and linguistics, State University of New York at Albany. (pp. 223-262). Austin, TX: University of Texas Press.

Smiley, P., & Huttenlocher, J. (1995). Conceptual development and the child's early words for events, objects, and persons. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 21-62). Hillsdale, NJ: Lawrence Erlbaum.

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, *24*, 99-142.

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.

Smith, L. B., Rattermann, M. J., & Sera, M. (1988). *Higher* and *lower*: Comparative and categorical interpretations by children. *Cognitive Development*, *3*, 341-357.

Snedeker, J. Brent, M., & Gleitman, L. R. (submitted). The changing character of the mental lexicon: An information-based account of early word learning.

Sodian, B. (2004). Do infants apply the principle of rational action to human agents? Infant *Behavior & Development*, *27*, 31-41.

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

Spelke, E. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, *15*, 626-636.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.

Sproat, R., & Shih, C. (1991). The cross-linguistic distribution of adjective ordering restrictions. In C. Georgopoulos & R. Ishihara (Eds.), *Interdisciplinary approaches to language: Essays in Honor of S.-Y. Kuroda* (pp. 565-593). Dordrecht: Kluwer.

Staab, S., & Hahn, U. (1998). Grading on the fly. In M.A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Meeting of the Cognitive Science Society* (pp. 1019-1024). Mahwah, NJ: Lawrence Erlbaum Associates.

Strawson, P. F. (1950). On referring. *Mind*, *59*, 320-344.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*, 103-114.

Syrett, K., Bradley, E., Lidz, J., & Kennedy, C. (2005). Shifting standards: Children's understanding of gradable adjectives. Paper presented at the Annual Meeting of the Linguistic Society of America, January 6-9, Oakland, CA.

Syrett, K., Bradley, E., Lidz, J., & Kennedy, C. (2006). Shifting standards: Children's understanding of gradable adjectives. In K. Ud Deen, J. Nomura, B. Schulz, & B. D. Schwartz (Eds.), *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition - North* America (pp. 353-364). Honolulu, HI. Cambridge MA: UConn Occasional Papers in Linguistics 4.

Syrett, K., Kennedy, C., & Lidz, J. (submitted). Meaning and context in children's understanding of gradable adjectives.

Syrett, K., Lidz, J., & Kennedy, C. (2006a). The semantic typology of gradable adjectives: Experimental evidence from adult and child language. Poster presented at the *CUNY Human Sentence Processing Conference*, March 23-25, New York, NY.

Syrett, K., Lidz, J., & Kennedy, C. (2006b). Scalar structure and the semantic representations of gradable adjectives in child language. Paper presented at the Main Session of the 42nd Regional Meeting of the Chicago Linguistic Society, April 6-8, Chicago, IL.

Talmy, L. (1983). How language structures space. In H. L. Pick, Jr., & L. P. Acredolo (Eds.), *Spatial orientation*: *Theory, research, and application* (pp. 225-282). New York, NY: Plenum Press. Based on the proceedings of the Conference on Spatial Orientation and Perception, July 14-16, 1980, University of Minnesota, Minneapolis, MN.

Talmy, L. (1985a). Grammatical categories and the lexicon. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 3) (pp. 57-149). Cambridge: Cambridge University Press.

Talmy, L. (1985b). Force dynamics in language and thought. In W. H. Eilfort, P. D. Kroeber, & K. L. Peterson (Eds.), *Papers from the Parasession on Causatives and Agentivity, 21st Regional Meeting of the Chicago Linguistic Society* (pp. 293-337). Chicago, IL: Chicago Linguistic Society.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*, 49-100.

Tanz, C. (1977). Polar exploration: *Hot* and *cold*, *cool* and *cold*. *Journal of Child Language*, *4*, 477-478.

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24*, 535-566.

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

Tenny, C. (1987). Grammaticalizing aspect and affectedness. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Tenny, C. (1992). The aspectual interface hypothesis. In I. Sag & A. Szabolsci (Eds.), *Lexical matters* (pp. 1-28). Stanford, CA: Center for the Study of Language and Information.

Tenny, C. (1994). *Aspectual roles and the syntax-semantics interface*. Dordrecht: Kluwer.

Teodorescu, A. (2006). Adjective ordering revisited. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics* (pp. 399-407). Somerville, MA: Cascadilla Proceedings Project.

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.

Thorpe, K., Baumgartner, H., & Fernald, A. (2006). Children's developing ability to interpret adjective-noun combinations. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 631-642). Somerville, MA: Cascadilla Press.

Tortora, C. (1998). Verbs of inherently directed motion are compatible with resultative phrases. *Linguistic Inquiry*, *29*, 338-345.

Townsend, D. (1976). Do children interpret 'marked' comparative adjectives as their opposites? *Journal of Child Language*, *3*, 385-396.

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

Unger, P. K. (1975). *Ignorance: A case for skepticism*. Oxford: Clarendon Press.

Vanden Wyngaerd, G. (2001). Measuring events. *Language*, *77*, 61-90.

Vendler, Z. (1957). Verbs and times. *Philosophical Review*, *66*, 143-160.

Vendler, Z. (1968). *Adjectives and nominalizations*. *Papers on formal linguistics* (Vol. 5). The Hague: Mouton.

Vendler, Z. (1972). *Res cogitans*. Ithaca, NY: Cornell University Press.

Verkuyl, H. (1972). *On the compositional nature of the aspects*. Dordrecht: Reidel.

Verkuyl, H. (1993). *A theory of aspectuality: The interaction between temporal and atemporal structure*. Cambridge: Cambridge University 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.

Wagner, L., & Carey, S. (2005). 12-month-old infants represent probable endings of motion events. *Infancy*, *7*, 73-83.

Wallace, J. (1972). Positive, comparative, superlative. *The Journal of Philosophy*, *69*, 773-782.

Wang, H. (1995). Dong jie shi shu bu jie gou pei jia yan jiu [Studies on the valence of resultative complement constructions]. In Y. Shen & D. Zheng (Eds.), *Xian dai Han yu pei jia yu fa yan jiu* [*Studies on valent grammar in modern Chinese*] (pp. 144-167). Beijing: Beijing da xue chu ban she.

Warren, B. (1988). Ambiguity and vagueness in adjectives. *Studia Linguistica*, *42*, 122-171.

Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition*, *70*, B35-B50.

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

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

Wechsler, S. (2001). An analysis of English resultatives under the event-argument homomorphism model of telicity. In *Proceedings of the 3rd Workshop on Text Structure*, University of Texas, Austin, October 13-15, 2000. Austin, TX: University of Texas Linguistics Department.

Wechsler, S. (2005). Weighing in on scales: A reply to Goldberg and Jackendoff. *Language*, *81*, 465-473.

Wetzer, H. (1992). 'Nouny' and 'verby' adjectivals: A typology of predicative adjectival constructions. In M. Kefer & J. Van Der Auwera (Eds.), *Meaning and grammar: Cross-linguistic perspectives* (pp 223-262). Berlin: Mouton de Gruyter.

Wheeler, S. C. (1972). Attributives and their modifiers. *Noûs*, *6*, 310-334.

Wierzbicka, A. (1972). *Semantic Primitives*. Frankfurt: Athenäum.

Wierzbicka, A. (1989). Semantic primitives: The expanding set. *Quaderni di Semantica* X, *2*, 309-32.

Williams, A. (2005). *Complex causatives and verbal valence*. Doctoral dissertation, University of Pennsylvania, Philadelphia, PA.

Williams, R. J., Jr. (1977). *The acquisition of spatial adjectives by black and white children*. Doctoral dissertation, The Florida State University, Tallahassee, FL.

Winter, Y. (to appear). Closure and telicity across categories. In *Proceedings of Semantics and Linguistic Theory XVI*. Ithaca, NY: CLC Publications.

Wojdak, R. (2001). An Argument for Category Neutrality? In K. Mewgerdoomian & L.A. Bar-el (Eds.), *Proceedings of the 20th West Coast Conference on Formal Linguistics* (pp. 621-634). Somerville, MA: Cascadilla Press.

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1-34.

Woodward, A. L., & Markman, E. M. (1998). Early word learning. In W. Damon, D. Kuhn, & R. Siegler (Eds.), *Handbook of child psychology* (Vol. 2)*: Cognition, perception and language* (pp. 371-420). New York, NY: John Wiley and Sons.

Woodward, A., (2005). The infant origins of intentional understanding. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 33) (pp. 229-262). Oxford: Elsevier.

Wright, C. (1975). On the coherence of vague predicates. *Synthese*, *30*, 325-365.

Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245-272.

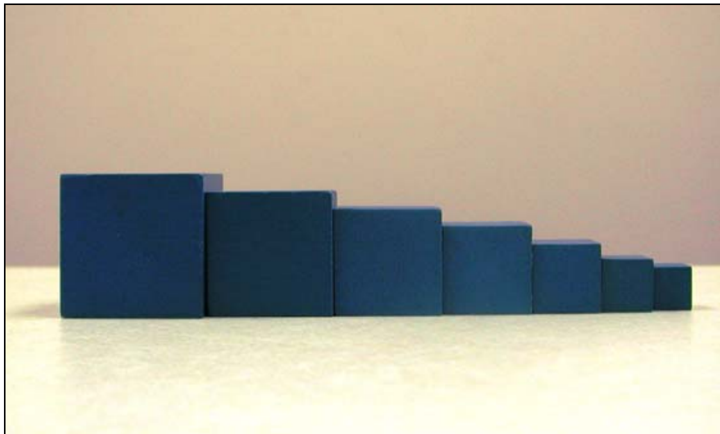
Yoon, Y. (1996). Total and partial predicates and the weak and strong interpretations. *Natural Language Semantics*, *4*, 217-236.

Ziff, P. (1960). *Semantic analysis*. Ithaca, NY: Cornell University Press.

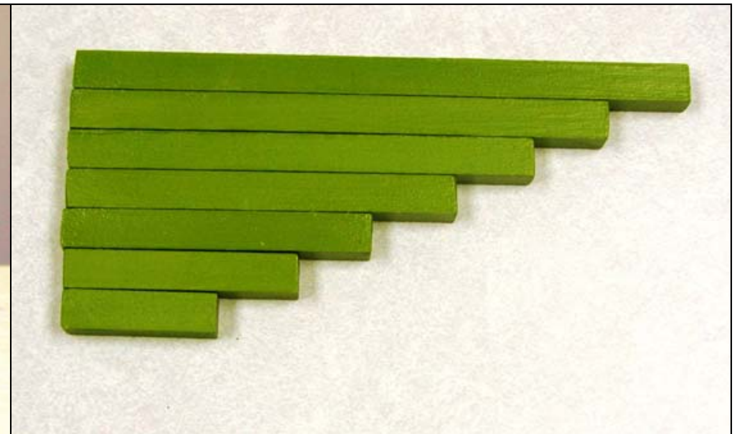
Zwarts, J. (2005). Prepositional aspect and the algebra of paths. *Linguistics and Philosophy*, *28*, 739-779.

Appendices

Appendix A: Stimuli used in the Scalar Judgment Task



big



long

spotted

full

Appendix B: Training stimuli for Experiment 1 (Presupposition Assessment Task)

adjective	stimuli	pragmatic status of request
happy	pictures of a happy face and an angry face	felicitous
round	pictures of a green triangle and a blue square	infelicitous
red	pictures of a red circle and a red square	infelicitous
blue	pictures of a yellow bird and a blue bird	felicitous

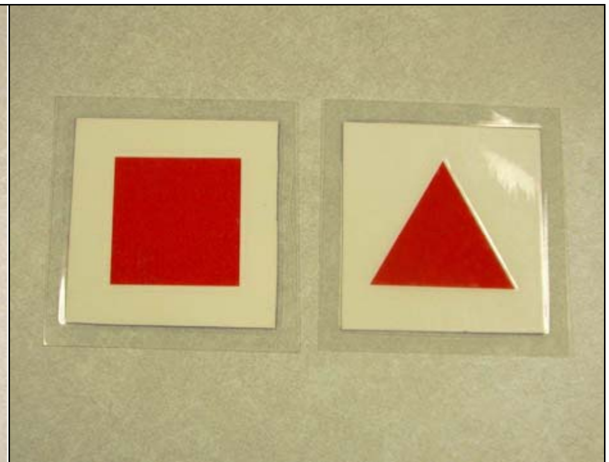
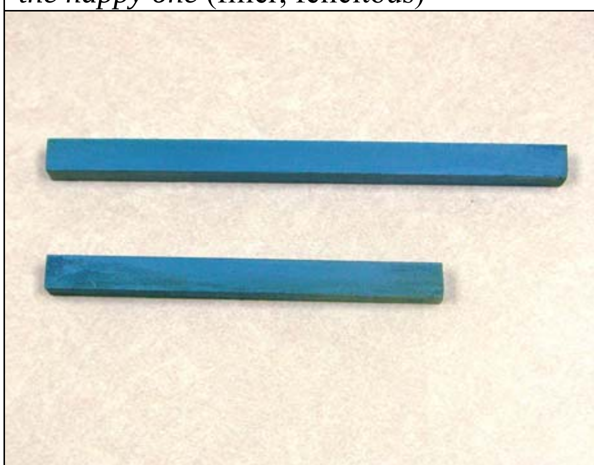
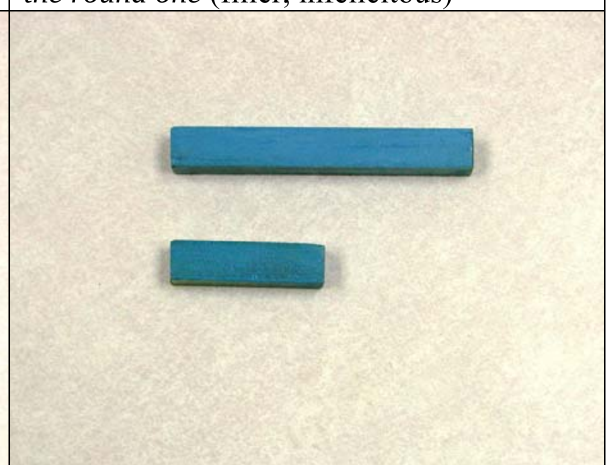
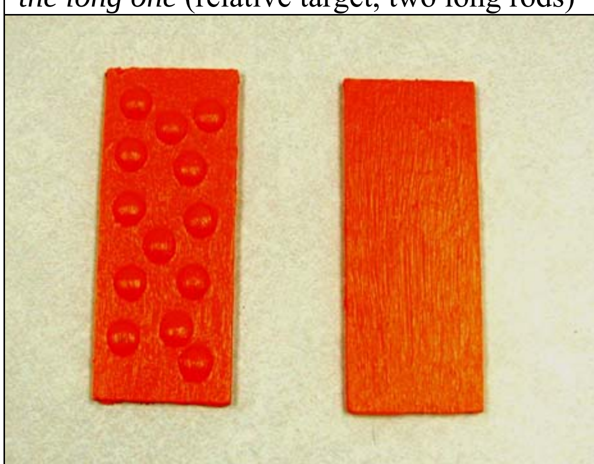
Appendix C: Target stimuli for Experiment 1 (Presupposition Assessment Task)

adjective	stimuli	pragmatic status of request
relative		
big	two blocks, one bigger than the other (blocks #1 and 3 from SJT)	felicitous
big	two blocks, one bigger than the other, both smaller than the first pair (blocks #5 and 7 from SJT)	felicitous
long	two rods, one longer than the other (rods #1 and 3 from SJT)	felicitous
long	two rods, one longer than the other, both shorter than the first pair (rods #5 and 7 from SJT)	felicitous
absolute		
spotted	two disks, one with a few spots, one without any (disks #5 and 7 from SJT)	felicitous
spotted	two disks, one more spotted than the other (disks #1 and 4 from SJT)	infelicitous
full	full container and container filled more than halfway (containers #1 and 3 from SJT)	felicitous
full	two containers filled somewhat, one more than the other, but neither full (containers #4 and 6 from SJT)	infelicitous

Appendix D: Control stimuli for Experiment 1 (Presupposition Assessment Task)

adjective	stimuli	pragmatic status of request
color		
yellow	pictures of a yellow bird and a black bird	infelicitous
green	purple yo-yo and yellow yo-yo	infelicitous
red	pictures of a red square and a red circle	infelicitous
red	red poker chip and a white poker chip	felicitous
shape		
round	pictures of a red triangle and a red square	infelicitous
square	pictures of a blue square and a yellow circle	felicitous
mood		
happy	pictures of a sad face and an angry face	infelicitous
happy	pictures of an angry face and a sad face	infelicitous
sad	pictures of a happy face and a sad face	felicitous

Appendix E: Examples of stimuli used in Experiments 1 and 2 (Presupposition Assessment Task)

	
the happy one (filler, felicitous)	*the round one* (filler, infelicitous)
	
the long one (relative target, two long rods)	*the long one* (relative target, two short rods)
	
the bumpy one (abs. min. target, felicitous)	*the full one* (abs. max. target, infelicitous)

Appendix F: Information about stimuli, requests, and responses for Experiment 1 (Presupposition Assessment Task)

request (*Give me the ____ one.*)	adjective status	pragmatic status of request	stimuli	correct response
big	relative	felicitous	two blocks, one bigger than the other	give the object that exceeds the cutoff for the context (e.g., the bigger one)
long	relative	felicitous	two rods, one longer than the other	give the object that exceeds the cutoff for the context (e.g., the longer one)
full	absolute, maximum standard	felicitous	two containers, only one full, the other filled to some degree	give the object that manifests the property (e.g., the full one)
full	absolute, maximum standard	infelicitous	two containers, neither full, but one fuller than the other	give neither object
spotted	absolute, minimum standard	felicitous	two disks, only one spotted	give the object that manifests the property (e.g., the spotted one)
spotted	absolute, minimum standard	infelicitous	two disks, both spotted, but one more than the other	give neither object

Appendix G: Co-occurrence of 20 adverbs and adjectives in the spoken BNC

	maximum standard absolute GAs					relative GAs				
restricted adverbs	*clean*	*dry*	*empty*	*full*	*straight*	*big*	*high*	*long*	*tall*	*wide*
almost	0	1	0	1	1	0	0	0	0	0
completely	0	1	4	4	0	0	0	0	0	0
entirely	0	0	0	0	0	0	0	0	0	0
half	0	0	6	6	0	0	0	0	0	0
totally	0	0	0	2	0	0	0	0	0	0
non-restricted adverbs	0	0	1	1	0	1	6	3	0	0
extremely	10	2	0	3	3	42	9	17	2	8
really	0	0	1	0	0	0	0	1	0	0
relatively	0	5	0	6	0	143	77	216	5	9
too	27	9	1	16	10	182	214	252	11	28
very	0	0	1	1	0	1	6	3	0	0

Appendix H: Ten most frequently modified adjectives for each of 10 different adverbs

adverbs	10 most frequently co-occurring adjectives[117]
restricted adverbs	
almost	*impossible* (20), *certain* (16), *inevitable* (6), *ready* (5), *identical* (3), *double* (3), *unthinkable* (3), *sure* (3), *universal* (2), *right* (2), *flat* (2), *complete* (2), *continuous* (2)
completely	*different* (96), *wrong* (17), *new* (16), *anonymous* (11), *separate* (8), *irrelevant* (8), *confidential* (7), *black* (5), *clear* (5), *free* (5)
entirely	*different* (25), *sure* (6), *new* (6), *wrong* (6), *clear* (5), *consistent* (4), *sensible* (4), *separate* (4), *happy* (4), *dependent* (3), *true* (3), *possible* (3), *appropriate* (3)
half	*dead* (7), *grand* (7), *full* (6), *asleep* (6), *past* (5), *empty* (5), *yearly* (4), *okay* (3)
totally	*different* (101), *wrong* (23), *opposed* (9), *unnecessary* (8), *unacceptable* (7), *separate* (6), *honest* (6), *new* (6), *dependent* (5), *independent* (4), *convinced* (4), *true* (4)
non-restricted adverbs	
extremely	*difficult* (47), *important* (31), *good* (3), *valuable* (12), *helpful* (8), *high* (6), *interesting* (6), *large* (6), *expensive* (6), *nice* (5), *frustrating* (5), *cold* (5), *useful* (5)
really	*good* (432), *nice* (291), *bad* (104), *funny* (99), *big* (42), *lovely* (39), *important* (39), *weird* (36), *hard* (31), *sure* (30)
relatively	*small* (43), *new* (15), *short* (12), *straightforward* (9), *low* (9), *cheap* (8), *minor* (8), *high* (7), *little* (6), *recent* (6), *easy* (6)
too	*bad* (354), *big* (143), *late* (137), *long* (71), *high* (68), *hot* (67), *small* (61), *good* (57), *busy* (53), *sure* (50)
very	*good* (1972), *nice* (782), *difficult* (569), *important* (562), *small* (252), *interesting* (241), *high* (201), *happy* (191), *big* (182), *strong* (165)

[117] Cells have less than 10 items if the rest of the adjectives appearing with that adverb shared a frequency of one or two. Cells had more than 10 items if more than one adjective in the list shared the same frequency.

Appendix I: Ten most frequent adverbial modifiers for each of 10 different adjectives

adjectives	10 most frequently co-occurring adverbs[118]
maximum standard absolute GAs	
clean	*very* (27), *really* (10), *quite* (6), *just* (6), *always* (4), *as* (3), *so* (3), *ever so* (3), *like* (3), *sort of* (2), *spotlessly* (2), *perfectly* (2)
dry	*very* (9), *so* (7), *too* (5), *as* (5), *a bit* (5), *mostly* (5), *fairly* (4), *quite* (4), *virtually* (3), *mainly* (3)
empty	*half* (6), *completely* (4), *then* (2), *virtually* (2)
full	*in* (27), *very* (16), *so* (13), *always* (11), *as* (10), *quite* (10), *too* (6), *half* (6), *just* (6), *a bit* (4), *completely* (4), *still* (4)
straight	*just* (17), *very* (10), *so* (7), *that* (4), *less* (4), *quite* (4), *absolutely* (3), *then* (3), *there* (3), *dead* (3), *fairly* (3), *here* (3), *now* (3), *pretty* (3), *really* (3)
relative GAs	
big	*very* (182), *too* (143), *as* (72), *quite* (63), *really* (42), *so* (38), *that* (38), *a bit* (21), *real* (14), *fairly* (13)
high	*very* (214), *too* (77), *as* (47), *quite* (44), *so* (35), *fairly* (20), *pretty* (12), *really* (9), *relatively* (9), *that* (8)
long	*very* (252), *too* (216), *so* (209), *as* (125), *that* (98), *there* (34), *for* (33), *quite* (21), *awful* (18), *a bit* (17), *really* (17)
tall	*quite* (14), *as* (12), *very* (11), *too* (5), *fairly* (4), *pretty* (3), *a bit* (2), *really* (2)
wide	*very* (28), *as* (15), *too* (9), *quite* (9), *really* (8), *so* (7), *fairly* (4), *just* (4), *more* (4), *narrowly* (2)

[118] The number of items in each cell deviates from 10 if the conditions mentioned in the previous footnote are met.

Appendix J: Example of response sheet for adult background information

The following information will serve as an informative linguistic tool for the interpretation of the data we obtain from this experiment. We do not evaluate individual participants; rather, we look for trends among groups of participants. Your completion of this form is voluntary. You are free to omit any information that you do not wish to share.

(1) Gender (circle one): male female

(2) Age: ____ years

(3) Native Language(s): ______________________________

(4) Other languages you have studied or with which you have experience: ______________

____

(5) Have you ever had any speech or hearing impairment? (circle one) yes no
If yes, please explain. ______________________________

____

(6) Is there anything else you would like to share? ______________________________

____

Appendix K: Script with instructions for adult participants

In this experiment, you will be watching a short video (less than 5 minutes long) in which you will see a sequence of *5 experimental trials*.

Each trial will *begin* with a picture of a baby and the sound of a baby giggling. You will then see a blank screen and hear a voice say something. Now, because this experiment was designed for two-and-a-half-year-olds, the voice you hear will sound like child-directed speech. During each trial, you will see some pictures of objects, accompanied by a description of those objects.

At the *end* of each trial, you will see two new objects, one on each side of the screen. The voice will say, "Look! They're different!" You will then be asked to answer a question about these objects. To answer the question, you will need to choose either the object on the left side of the screen or the object on the right side of the screen. Note the corresponding labels next to the video screen. You should mark your answer for each trial on the appropriate response sheet by choosing either LEFT or RIGHT, then flip the page to get ready for the next trial. Again, you'll know when the next trial is beginning, because you will see the picture of the baby and hear giggling.

Please note that you MUST choose an answer for each trial. You should not look back at your answers for previous trials. (This will not help you anyway, because of the experiment design.)

Appendix L: Example of response sheet for adult response packet

Trial 1
Answer (Mark one.)
LEFT RIGHT