# Children's and adults' interpretation of gradable adjectives: the role of the comparison class

*Merle Weicker and Petra Schulz*

Goethe University Frankfurt

## 1 Introduction

The question of which cues listeners consider when interpreting relative gradable adjectives (GA) such as *tall* or *small* has been the topic of much recent research. Theoretical approaches agree that a threshold needs to be calculated relative to a comparison class, but no consensus has been reached regarding **how** thresholds are determined. Experimental studies with children and adults mostly investigated the interpretation of GAs by looking at non-linguistic cues such as the visual context, but left open how linguistic cues, in concert with non-linguistic cues, are used by listeners. The present study investigates how children and adults interpret GAs when provided with linguistic and with non-linguistic cues regarding the comparison class to shed light on the role of the comparison class for determining the threshold.

In order to interpret sentences with a GA, the listener needs to calculate a threshold. This threshold identifies the "cutoff point" for the entities that have the property described by the adjective, e.g., what counts as tall (e.g., Kennedy 2007). Semantic accounts typically maintain that the threshold is determined by the context in some way or another; hence, it can vary. The threshold for GAs is assumed to be calculated relative to a comparison class (e.g., von Stechow 1984). The comparison class can be made explicit by a *for*-phrase (*Fido is tall for a Chihuahua/for a dog*), but often information about the comparison class is implicit and has to be inferred from other cues, which can be linguistic or non-linguistic in nature. Linguistic cues include the modified head noun (*Fido is a tall Chihuahua/dog*); non-linguistic cues comprise the visual context (e.g., other dogs), the discourse context (e.g., the topic of the conversation), and world knowledge (e.g., about the typical size of dogs). Informally speaking, the listener must decide on the appropriate comparison class in order to calculate the threshold. Even if the comparison class is established, however, listeners can be uncertain about the threshold. This characteristic of GAs, referred to as "vagueness", is reflected by fuzzy boundaries, the existence of borderline cases, and by their susceptibility to the Sorites Paradox (see Morzycki 2015, Solt 2015).

The current study asks whether listeners use cues provided by the modified noun or by the visual context to calculate the threshold for GAs. By comparing children's and adults' interpretation patterns, we are able to detect which interpretive properties emerge early, indicating a default interpretation. Adults' and children's response patterns are discussed as to whether they indicate that thresholds are determined solely by the comparison class or that uncertainty exists regarding the threshold. We argue that these experimental findings directly relate to different theoretical semantic approaches of GAs.

## 2 Semantics of gradable adjectives and implications for acquisition

Two main semantic approaches have been suggested for the analysis of GAs, referred to as (i) inherent vagueness and (ii) degree-based (see Morzycki 2015). (i) Inherent vagueness approaches analyze GAs as denoting partial functions (type $<e,t>$) that yield a positive

extension, a negative extension, and an extension gap (e.g., Klein 1980). Individuals that satisfy a GA are in the positive extension, individuals that do not satisfy the GA are in the negative extension. Individuals that are neither in the positive nor in the negative extension fall into the extension gap. These individuals constitute the borderline cases, which are characteristic for vague adjectives. The classification of what belongs to the positive and negative extension and to the extension gap varies depending on the comparison class, which is determined by the context. The partitioning of the respective set of individuals can be more or less precise; this may result in a listener being uncertain regarding the threshold for what counts as definitely tall, for example. Common to inherent vagueness approaches is the assumption that vagueness is an inherent property of particular predicates, here GAs.

(ii) Degree-based approaches generally analyze GAs as relations between individuals and degrees on a scale (type $<d,<e,t>>$) (e.g., Cresswell 1976). Degrees are understood as representations of measurement. A scale is a set of degrees that is totally ordered regarding a dimension such as height. To evaluate whether the property denoted by a GA is true of an individual, the individual must be related to a degree that exceeds a certain threshold or standard of comparison to a sufficiently salient extent (e.g., Kennedy 2007). For the positive unmarked form, a null degree morpheme ('pos') is assumed to serve this function. The degree determined by 'pos', that is the threshold it introduces, depends on the context. According to Kennedy (2007), vagueness results from epistemic uncertainty about whether an individual exceeds the threshold to an extent sufficiently large to be salient in context and from "metalinguistic resistance to treating highly similar objects differently relative to the property expressed by the positive form" (Kennedy, 2007: 42). Degree-based approaches relate vagueness to the compositional derivation of GAs, i.e., the existence of a 'pos' morpheme, and, unlike inherent vagueness approaches, do not assume that vagueness is an inherent property of predicates.

Importantly, although degree-based and inherent vagueness approaches analyze GAs differently, they predict the same interpretive outcome of the positive form: context-sensitive and vague. Both approaches leave open how exactly the context provides the threshold, i.e. which cues are actually relevant to calculate the threshold. Experimental data from adults and children can contribute to answering this question. Adults are competent speakers with rich linguistic knowledge and world knowledge who are expected to know which cues they can use to determine the comparison class and to calculate the threshold accordingly. To date, it is unclear which comparison class they choose, if multiple options are available.

Children are in the process of detecting which cues exist and how to employ them for the interpretation of GAs. Accordingly, acquisition data can reveal whether any and if so which of the cues are already used early on as a default. Acquisition data may also contribute to the debate of which semantic analysis of GAs is on the right track. This is because the inherent vagueness and the degree-based approaches result in different acquisition steps: if vagueness is an inherent property of GAs, vagueness should be reflected in children's interpretation from an early age. If vagueness follows from the compositional derivation of GAs, however, children's initial interpretation of GAs may show no indications of vagueness.

## 3 Previous experimental research

Experimental studies on the interpretation of GAs have focused on how listeners determine the comparison class and calculate the respective threshold. While adults could be asked directly about which comparison class they have in mind in a specific context, children's interpretation were inferred from their responses to questions or their object choices in a specific experimental setup. Tessler et al. (2020) found that adults' choices of the comparison

class were affected by the visual context and by the head noun modified by the adjective. Because in this study participants were asked about the comparison class but not about their actual threshold, it is open whether adults' thresholds differ for different comparison classes.

Acquisition studies have revealed that at age 2 children could use non-linguistic cues such as their world knowledge for the interpretation of GAs. If another non-linguistic cue was available, i.e., other objects presented in the visual setup, children used this visual cue, and not their world knowledge, to calculate the threshold (Ebeling & Gelman 1988). To our knowledge, only one child study employed visual cues, i.e., unknown objects from the same or a different category, together with linguistic cues, i.e., novel names for the objects presented (Barner & Snedeker 2008). The findings suggest that the 4-year-old children prefer linguistic cues for determining the comparison class and that they calculate the threshold relative to that comparison class. While this experimental research has substantially advanced our knowledge regarding children's and adults' interpretation of GAs, many issues remain open, most notably the role of linguistic cues compared to non-linguistic cues addressed in our study.

# 4 Our study

Our study investigates the influence of **both** linguistic and non-linguistic cues on participants' interpretation of GAs to uncover whether the comparison class is determined based on linguistic and/or non-linguistic cues and whether changes in the comparison class result in participants' adjustment of their thresholds. More specifically, we ask whether the type of head noun (here: basic-level vs. superordinate-level existing nouns) and/or the visual context (here: one vs. two object categories) affect the threshold for GAs. Accordingly, a novel picture-choice task was developed containing linguistic and non-linguistic cues, which could be considered for the calculation of the threshold. This experimental setup differs from previous studies in that it allows to tease apart whether threshold adjustment is due to changes in the modified noun or in the visual context or in both. The method of picture choice is well-established in experimental research on adjective interpretation in children (Barner & Snedeker 2008, Syrett et al. 2010) and in adults (Solt & Gotzner 2010). Accordingly, we used the same experimental set up for adults and children, which enabled us to directly compare children's and adults' interpretation patterns and to detect whether adults show interpretation patterns that are not yet available to children. The participants (43 children aged 3-5 years and 26 adults, all monolingual German) saw pictures of objects, which varied regarding their size, and had to fulfil requests (e.g., *Please hand me the big toys*) by selecting those objects that matched their interpretation of the request. Then, based on participants' choices, the individual thresholds were determined.

## 4.1 Materials

We tested the GAs *big* and *small*. For each adjective, participants saw 3 types of visual contexts we refer to as baseline, upper expansion, and lower expansion (Figure 1). In the baseline, participants saw 8 toys from the same basic-level category (water balloons or space hoppers) varying in size. In the upper expansion, participants saw 12 toys: the 8 water balloons from the baseline and 4 bigger toys from a second basic-level category (soccer balls). In the lower expansion, participants saw also 12 objects, but the additional toys (buckets) were smaller compared to the 8 space hoppers from the baseline. Each object was depicted on a picture card. All picture cards were of the same size independent of the size of the object depicted.

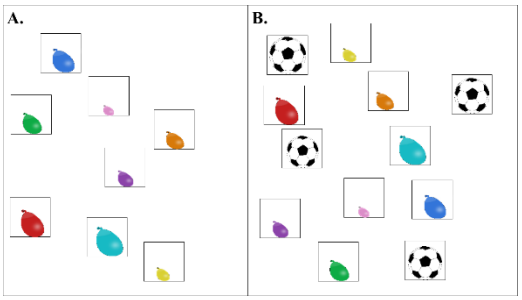


**Figure 1. A = Example baseline context; B = Example upper expansion context.**

The test prompts contained the adjective *small* or *big*, which modified a basic-level noun (*water balloons* or *space hoppers*) or a superordinate-level noun (*toys*). The prompts were spoken with neutral, non-contrastive intonation. Combination of type of visual context and type of modified noun resulted in 6 test trials per adjective (Table 1). In addition, 8 filler trials were included per session, increasing the distance between test trials with the same prompt or with the same visual context in order to minimize potential influences from the prior presentation.

**Table 1. Test trials in order of presentation.**

| Trial | Visual context | Modified noun (*Please hand me the big/small…*) |
|---|---|---|
| 1 | Baseline | Basic-level (*water balloons*) |
| 2 | Upper expansion | Superordinate-level (*toys*) |
| 3 | Upper expansion | Basic-level (*water balloons*) |
| 4 | Baseline | Basic-level (*space hoppers*) |
| 5 | Lower expansion | Superordinate-level (*toys*) |
| 6 | Lower expansion | Basic-level (*space hoppers*) |

## 4.2 Procedure

The experiment was administered in 2 sessions. In session 1 participants received the *big*-trials; in session 2 the *small*-trials. For the test trials, the experimenter distributed all objects cards in random order on a table (Figure 1). The experimenter uttered the test prompt and the participants' task was to select the matching object cards. In order to respond to the request, participants had to establish their own ordering to calculate the threshold because the objects were not presented in a pre-ordered fashion, which would have suggested a specific relation a between the objects at the outset. Presenting objects in unordered fashion is ecologically more valid because it mirrors the often non-transparent relations between objects in reality.

## 4.3 Results

Participants' individual thresholds were determined by calculating their respective cutoff points: for *big*, this was the smallest object selected; for *small*, it was the biggest object selected. To find out whether the threshold was affected by changes in the modified noun and/or the visual context, we compared the cutoff points across trials 1, 2, 3, and across trials 4, 5, 6, respectively (Table 1).

Children's cutoff points for *big* differed significantly across trials 1, 2, 3 (Friedman test; $p = .001$) and across trials 4, 5, 6 ($p = .005$). Post-hoc pairwise comparisons revealed a difference between trial 1 and 2 ($p = .046$), between trial 2 and 3 ($p = .004$), and between trial 4 and 5 ($p = .035$). For *small*, the cutoff points did not differ significantly across trials 1, 2, 3 ($p = .236$), but across trials 4, 5, 6 ($p < .001$). Post-hoc pairwise comparisons revealed a difference between trial 4 and 5 ($p = .033$) and between trial 5 and 6 ($p = .002$), respectively. As for the adults, the same effects of trial were found for *big* (trials 1, 2, 3: $p = .001$; trials 4, 5, 6: $p = .003$) and *small* (trials 1, 2, 3: $p = .144$; trials 4, 5, 6: $p < .001$).

Post-hoc pairwise comparisons revealed a difference between trial 1 and 2 ($p = .003$) and between trial 4 and 5 ($p = .031$) for *big* as well as between trial 1 and 2 for *small* ($p = .002$).

## 5 Discussion

Focusing on GAs, we asked whether changes in the comparison class result in listeners' adjustment of their thresholds and whether the comparison class is determined based on linguistic and/or non-linguistic cues. Our data reveal, first, that children's and adults' thresholds for *big* and *small* are indeed context-sensitive: their thresholds across trials differed significantly, with the only exception of *small* in one visual context (upper expansion) (see Barner & Snedeker 2008, for a similar divergent pattern). It may be that this pattern is related to the fact that *small* and *big* are associated with the same dimension (size), with *small* as the negative member of the antonym pair. When evaluating smallness participants may "ignore" bigger objects in the upper expansion resulting in no adjustment of the threshold.

Our second finding concerns the visual cues to the comparison class. When the visual contexts differed and the noun was the same (trial 1 vs. 3 and trial 4 vs. 6), children's and adults' thresholds for *big* and *small* did not change, but when different visual contexts were accompanied by different head nouns (trial 1 vs. 2 and 4 vs. 5) their thresholds changed. These results suggest that the noun plays a more crucial role than the visual context for calculating the threshold, given that linguistic and non-linguistic cues are available but are not congruent. The child response patterns provide additional evidence for the prominent role of the noun: children's threshold for *big* (trial 2 vs. 3) and *small* (trial 5 vs. 6) changed when the noun changed from basic- to superordinate-level, although the visual context was the same. This suggests that children use linguistic cues, here the modified noun, to determine the comparison class and to calculate the threshold relative to this class, in line with previous findings using novel objects and novel nouns (Barner & Snedeker 2008). In a nutshell, children have different thresholds for different nouns and the same threshold for the same nouns.

Why is this pattern not observed in the adults? We argue that adults exhibit more flexibility than children and that this 'flexibility' is related to the vagueness of relative GA. Given that in our experimental setup the size differences between the single objects were quite small, the objects in the visual context were very similar with respect to the property encoded by the adjective. It is possible that adults are indifferent to small differences regarding the relevant property and accept that there is no precise boundary between 'big' and 'not big' and 'small' and 'not small' objects. Consequently, this boundary can shift from trial to trial but it does not need to. Which objects fall into the borderline area may change as well: a borderline case may be selected in one trial but not in another or it may or may not be selected across trials. This finding is in line with previous studies that observed manifestations of vagueness, using a different experimental setup (Alxatib & Pelletier 2011, Égré & Zehr 2018).

In summary, adults' interpretation of GAs reflect the properties described by semantic theory: adult listeners calculate a threshold that is context-sensitive and vague. Children's interpretation mirrors these properties only partly. Like adults, child listeners calculate a threshold that is context-sensitive: the threshold is dependent on a comparison class, which is determined based on linguistic cues. But, crucially, children's threshold is not vague. Therefore, we suggest that our data from acquisition speak in favour of approaches that derive vagueness from the compositional semantics of GAs, involving the 'pos'-morpheme, rather than in favour of approaches that regard vagueness as an inherent property of predicates. This proposal raises an important question for future research on vagueness: given that vagueness is a pervasive phenomenon in language and exists across different lexical categories (see Morzycki 2015), we can ask what these categories, including adjectives, have in common.

## References


Alxatib, Sam & Pelletier, Francis J. (2011). The psychology of vagueness: borderline cases and contradictions. *Mind & Language* 26. 287–326.

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

Cresswell, Max J. (1976). The semantics of degree. In Barbara Partee (ed.), *Montague Grammar,* 261-292. New York. Academic Press.

Ebeling, Karen S. & Gelman, Susan A. (1988). Coordination of size standards by young children. *Child Development* 59. 888-869.

Égré, Paul & Zehr, Jeremy (2018). Are gaps preferred to gluts? A closer look at borderline contradictions. In Elena Castroviejo, Louise McNally, Galit W. Sassoon (eds.), *The Semantics of Gradability, Vagueness, and Scale Structure,* 25-58. Berlin: Springer.

Kennedy, Christopher (2007). Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30. 1-45.

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

Morzycki, Marcin (2015). *Modification*. Cambridge. Cambridge University Press.

Solt, Stephanie (2015). Vagueness and imprecision: Empirical foundations. *Annual Review of Linguistics* 1. 107-127.

Solt, Stephanie & Gotzner, Nicole (2010). *Expensive, not expensive or cheap? An experimental investigation of vague predicates.* Paper presented at the 11th Szklarska Poreba Workshop on Experimental Pragmasemantics, Szklarska Poreba.

Syrett, Kristen Kennedy, Christopher & Jeffrey Lidz (2010). Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics* 27. 1-35.

Tessler, Michael H., Tsvilodub, Polina, Snedeker, Jesse, & Roger P. Levy (2020). Informational goals, sentence structure, and comparison class inference. In Stephanie Denison., Michael Mack, Yang Xu, Blair C. Armstrong (eds.*), Proceedings of the 42nd Annual Meeting of the Cognitive Science Society,* 537-543. Austin, TX: Cognitive Science Society.

von Stechow, Arnim (1984). Comparing semantic theories of comparison. *Journal of Semantics* 3. 1-77.