

Processing Gradable Adjectives in Context: A Visual World Study*

Helena Aparicio
University of Chicago

Ming Xiang
University of Chicago

Christopher Kennedy
University of Chicago

Abstract Both relative adjectives (RAs) like ‘*big*’ and absolute adjectives (AAs) like ‘*empty*’ are sensitive to the context: in the former case, the context determines how much size is required to count as big; in the latter, the context determines how much deviation from total emptiness is allowed to count as empty. Whereas it is generally agreed that the role of context with RAs is to fix the value of a threshold variable, the status of absolute adjective thresholds, and therefore the role of context in their interpretation, remains an object of debate. Some researchers have argued that all gradable adjectives have context-sensitive threshold variables that are assigned values by the same mechanisms (Lassiter & Goodman 2013). Others have claimed that AAs have fixed, endpoint-oriented meanings and that sensitivity to context arises from pragmatic reasoning about imprecision (Kennedy 2007; Syrett, Kennedy & Lidz 2009; van Rooij 2011; Burnett 2014; Qing & Franke 2014). In an eye-tracking Visual World experiment, we investigate RAs and AAs used as restrictive modifiers. We find that target identification is significantly faster for both types of adjectives when the visual context supports a restrictive interpretation of the predicate, although this effect is considerably delayed in the case of AAs. We conclude that for RAs, the target facilitation effect is driven by the lexical semantics of the predicate itself. However, it is argued that the extra processing cost observed with AAs results from pragmatic reasoning about imprecision.

Keywords: gradable adjectives, thresholds, context-sensitivity, Visual World, eye-tracking

* We would like to thank Aaron Hill, Jackson Lee and Gabriel Aparicio for invaluable technical support, and Michelle Namkoong and Sonia Juan Rubio for help in the creation of part of the visual stimuli. Thanks to Christina Kim, Gallagher Flinn, Julian Grove, Katie Franich and Pedro Brugarolas for punctual help on different aspects of the experiment. Finally, we would also like to thank the five anonymous SALT 25 reviewers, as well as the audiences of the GLiF seminar at the UPF, the 28th CUNY conference on Human Sentence Processing, SALT 25, and the workshop on *Gradability, Scale Structure, and Vagueness: Experimental Perspectives* celebrated in Madrid in May 2015. This project was supported by a NSF grant (BCS 1227144) to C. Kennedy and M. Xiang.

1 Introduction: Context-sensitivity and the Relative vs. Absolute Distinction

Gradable relative (1a) and absolute (1b) adjectives in their positive form have context-sensitive meanings (Rotstein & Winter 2004; Kennedy & McNally 2005; Kennedy 2007; McNally 2011, Sassoon & Toledo 2011 among many others).

- (1) a. John is *tall*.
 b. The movie theater is *empty*.

Relative adjectives are context-sensitive because their criteria of application change across contexts: different contexts pick out different sets of individuals of which the adjective can be truthfully predicated. This can be observed in example (1), where the cut-off point that distinguishes those individuals that count as tall from those that do not varies as a function of the features of the context. For instance, the cut-off point determining what individuals fall in the extension of the predicate '*tall*' will be set to a higher value in a context in which the height of basketball players is discussed. In contrast, in a context where the question under discussion is the height of male gymnasts (who tend to be much shorter than male basketball players), the cut-off point will be significantly lower. Thus, if John is 5' 10ft tall, (1a) is true in case John is a gymnast, but false in case he is a basketball player.

The criteria of applicability of absolute adjectives can also vary from context to context. Clearly, sentence (1b) can be truthfully uttered in a context in which there are no people in the movie theater. Intuitively, it also seems possible to felicitously utter (1b) in a context where there are only two people sitting in the back row of the theater during the premier of a popular movie. In such a context, the fact that the movie theater is not *completely* empty can be ignored for the purpose of uttering (1b). Other contexts, however, impose further restrictions on the amount of available information that can be ignored without rendering an utterance of (1b) infelicitous. Example (1b) stops being acceptable if uttered to describe the movie theater during a fire-emergency that requires the building to be evacuated. In this context, it seems clear that the same two people sitting in the back of the room can no longer be ignored. Unlike the previous context, in such a context '*empty*' must mean *completely* empty.

However, relative and absolute adjectives differ in their degrees of context-sensitivity. While the cut-off points, or thresholds, of relative adjectives display a great deal of contextual variability (e.g. it is possible to refer to a toddler as tall, even though he is possibly much shorter than a short adult), the thresholds of absolute adjectives are more rigid and tend to receive endpoint-oriented interpretations (e.g. in order to count as '*empty*', the movie theater in example (1b) must be very close to the maximal degree of emptiness). In other words, when it comes to absolute adjectives, contextual sensitivity seems to be reduced to how much distance from

the endpoint-oriented interpretation of the threshold is tolerated in a given context.

Since relative adjectives show unrestricted contextual variability, existing accounts of the relative/absolute distinction assume that the threshold of relative adjectives is always set by accessing contextual information. However, theories differ with respect to the status they assign to absolute thresholds. Broadly speaking, proposals can be split based on whether they take relative *and* absolute thresholds to be resolved via context, or whether they take the value of absolute thresholds to be semantically provided. Among the theories adhering to the former view is [Lassiter & Goodman 2013](#), who propose a Bayesian pragmatic account ([Frank & Goodman 2012](#)) of the relative/absolute distinction. [Lassiter & Goodman \(2013\)](#) assume that the meaning of the positive form of any gradable adjective contains a free variable over thresholds. The value of the threshold is derived via a coordination process where both speaker and listener recursively reason about each other's knowledge state and conversational goals in order to arrive at the correct utterance interpretation. The relative/absolute distinction arises from prior assumptions about the statistical properties of the comparison class used for the interpretation of the adjective (e.g. in the two contexts discussed in relation to example (1a) above, the priors would consist of a probability distribution over the heights of adult male gymnasts, and over the heights of adult male basketball players, respectively). Relative-like interpretations arise when the prior distribution is approximately normal. Absolute-like interpretations, on the other hand, result from skewed priors. Restricted context-sensitivity of absolute adjectives is expected, as the posterior probability distribution for possible threshold values presents a much smaller standard deviation when the adjective prior is skewed, compared to when it is roughly normal. Since likely values for the threshold are fairly restricted, it is correctly predicted that absolute adjectives should give rise to little uncertainty about the threshold value, thus leaving less room for context-driven variability.

The second type of theory assumes that absolute thresholds are not subject to contextual variability: truth-conditionally, absolute adjectives always have endpoint-oriented meanings. Context-sensitive uses of absolute adjectives are claimed to emerge from pragmatic reasoning about *imprecision* ([Kennedy 2007](#); [Syrett et al. 2009](#); [van Rooij 2011](#); [Burnett 2014](#); [Qing & Franke 2014](#)). Here, we focus on one possible implementation of this view, that of [Kennedy \(2007\)](#). Building on previous work on the topology of adjectival scales ([Kennedy & McNally 2005](#)), [Kennedy \(2007\)](#) traces the relative/absolute distinction back to the structural properties of the scale denoted by the adjective. [Kennedy](#) proposes that context-sensitivity with relative adjectives derives from the fact that their scales are open. Since the scale is unbounded in at least one of its end-points, the value of the threshold cannot be provided by the scale itself and in most cases must be computed with respect to a contextually provided comparison class. Absolute adjectives, on the other

hand, live in closed scales. In general, absolute standards either have maximal or minimal values that can be conventionally provided by one of the scale's endpoints. Context-sensitive uses of absolute adjectives do not result from standard-shifts of the adjectival threshold but rather from pragmatic reasoning about imprecision. Therefore, in this framework, context-sensitivity displayed by relative and absolute adjectives is fundamentally different: relative adjectives are semantically context-sensitive, whereas absolute adjectives are pragmatically context-sensitive.

Kennedy (2007) does not provide an explicit account to model tolerance for imprecision. However, there exist compatible approaches to this phenomenon that can derive non-minimal/maximal standard interpretations without having to assume shiftable absolute thresholds in the semantics. Lasersohn (1999) proposes that any expression in the language has a *pragmatic halos* formed by expressions of the same semantic type that differ only in ways that are pragmatically ignorable. For instance, in example (1b) repeated in (2), the adjective '*empty*' would have a pragmatic halo formed by adjective meanings in which the threshold is close enough to the maximal degree of emptiness. Thus, when (2) is uttered in a context where there are two people sitting in the back row of the theater, the speaker is uttering something that is strictly speaking false but that is still considered acceptable and informative in that particular context.

(2) The movie theater is *empty*.

A second type of approach is represented by the granularity account of imprecision, developed initially by Krifka (2007) to model imprecision in the domain of numeral expressions. Granularity-based accounts assume that any scale can be conceptualized with different levels of granularity. The level of granularity is determined, among other things, by the amount of distinctions each context makes among the degrees that form the relevant scale. In the adjectival domain the relevant scale would correspond to the scale provided by the adjective. For instance, in the imprecise interpretation of sentence (2) a coarser adjectival scale would be at play such that two people in the room would be mapped to the maximal degree of emptiness. A more precise context would distinguish between zero and two people in the room such that only the former would be mapped into the maximal degree of emptiness.

All the theories described above can account for the different patterns of context-sensitivity associated with relative and absolute adjectives. However, adjudicating among the existing proposals on purely theoretical grounds is difficult. In the present experiment, we report results from a Visual World eye-tracking experiment that investigates how context influences reference resolution during online processing of gradable adjectives.

2 Visual World studies of adjectival modification and reference resolution

The Visual World eye-tracking paradigm has proven useful in investigating questions about the influence of context during the resolution of temporarily ambiguous referents. In this methodology, participants' eye-movements are recorded while looking at visual displays as they hear utterances. In general, experimental manipulations vary features of the visual context, as well as the properties of the linguistic input. Based on the eye-movement patterns that emerge as a function of these manipulations, researchers make inferences regarding the ways in which contextual information is integrated during linguistic online processing. For instance, in seminal work on this paradigm, Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy (1995) showed that contextual information, introduced by the manipulation of the visual display, was immediately adopted by the listeners to guide their online parsing decisions.

In this section, we focus on experimental results that make use of the Visual World paradigm to investigate the influence of context in reference resolution of definite descriptions involving prenominal adjectival modification. In a Visual World study, Sedivy et al. (1999) investigate how the presence of a contrast set (or comparison class) in the visual display influences the processing of sentences containing relative adjectives like *'tall'*. In the experiment, participants heard instructions such as *'Pick up the tall glass'* while looking at displays of four objects. Two conditions were tested: the Contrast condition supported a contrasting interpretation of the adjective by including, alongside the target object (e.g. a tall glass), a contrast object that could be described by the noun but not the adjective in the instruction (e.g. a short glass). In the second condition, the No-Contrast condition, the contrasting object was substituted with a distractor, i.e. an object that could not be described either by the head noun or the modifier in the instruction. Finally, Both conditions contained a competitor that presented a higher degree of the property in the instruction when compared to the target, but was judged to be better described by an unmodified noun (e.g. a pitcher that was taller than the glass, but was itself not tall for a pitcher).

The main finding of the experiment was that participant's fixations converged on

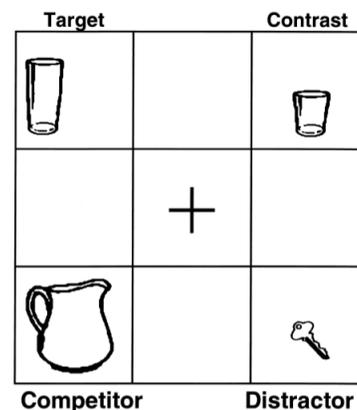


Figure 1: Trial example of the visual stimuli used in the Contrast Condition in Sedivy et al. (1999) study.

the target faster in the Contrast condition than they did in the No-Contrast condition. Crucially, in the Contrast condition participants zoomed into the target object during the adjective window at a point in which the head noun had not yet been processed. Thus, participants were able to identify the target at a point in which the linguistic instruction was still compatible with either the target or the competitor, suggesting that the information about the contrasting object was used very quickly.

Sedivy and colleagues point out that there are at least two interpretations compatible with these findings. The first one is that the effect of referential Contrast is driven by the use of the restrictive prenominal modifier, a use that triggers pragmatic reasoning about a set of referents that contrast with respect to the degree to which they possess the property denoted by the modifier (Altmann & Steedman 1988; Tanenhaus et al. 1995). If hearers assume that the speaker is observing the Gricean Maxim of Quantity, uttering the prenominal modifier would be maximally informative only in those cases in which the modifier is used to disambiguate between two possible referents. This view predicts that effects of referential contrast should arise with any prenominal modifier, as long as the context supports a restrictive interpretation.

A second option is that the effect of contrast is driven by aspects of the lexical semantics of relative adjectives. As discussed in §1, the threshold of relative adjectives must be resolved via a contextually supplied comparison class. Therefore, participants could be integrating the visual contextual clues as part of the lexical processing of the adjective itself. Under this view, the slower visual identification of the target object observed in the No-Contrast condition reflects the difficulty of processing a relative adjective in the absence of appropriate contextual support. This account predicts that adjectives that do not have context-sensitive thresholds, should not be sensitive to the presence or absence of a contrasting object in the visual display. However, this prediction does not seem to be fully borne out. In a subsequent eye-tracking study that replicated the same method and design, Sedivy (2005) tested color and material adjectives such as ‘*wooden*’ or ‘*plastic*’. Since neither of these two adjective types have context-sensitive thresholds, no effect of referential contrast is predicted to arise. As expected, color adjectives did not reveal any significant difference between the Contrast and the No-Contrast condition. However, material adjectives did give rise to a Contrast effect such that trials containing a contrasting object showed facilitated target identification. This latter finding is somewhat surprising if the effect of Contrast is lexically driven. For this reason, Sedivy (2005) rejected a purely lexical account of referential effects of contrast, and argued that listeners’ pragmatic reasoning about utterance informativity must be responsible for these effects.

Informativity-based accounts of referential contrast might also be able to explain the absence of contrast effects for color adjectives reported in Sedivy (2005). In a

follow-up eye-tracking study investigating color adjectives, Sedivy (2003) finds that the presence of a contrastive object only facilitated the identification of the target in those cases in which the color was predictable from the head noun, e.g. when the target was a yellow banana and the contrastive object was a brown banana. Sedivy argues that overt uses of predictable color adjectives must be taken to be highly informative by participants, as otherwise the speaker would be overspecifying by using a modifier that is redundant given the noun. This account is also compatible with the absence of contrast effects for color adjectives found in Sedivy (2005), where all color adjectives were unpredictable from the head-noun.¹

Based on the findings discussed in this section, it is clear that accounts of referential effects of contrast that uniquely rely on the grammatical properties of the different types of adjectives cannot explain the variety of effects reported in the literature. However, this does not necessarily entail that grammatical factors, such as lexical meaning, do not play a role in determining patterns of integration of contextual information during online processing. Another logical possibility is that certain types of adjectives, like color adjectives, only show informativity-based effects of contrast while other types of adjectives, like relative adjectives, show a combination of both. It might turn out that upon comparison of the timing and magnitude of the effects observed with different types of adjectives, a different picture emerges where lexical-semantic and informativity-based effects of referential contrast can in fact be teased apart. Defining the signature of these two different types of effects of referential contrast would allow us to provide empirical evidence to settle the issue of whether context-sensitivity displayed by absolute adjectives is driven by lexical-semantic factors. To date, this is a question that remains to be explored and will constitute one of the main aims of the present experiment.

3 The Current Study

The current study builds upon Sedivy et al.'s (1999) findings. In particular, we focus on the question of whether listeners' rapid sensitivity to context is driven by informativity-based pragmatic reasoning, lexical-semantic processing of a comparison class, or a mixture of both. A better understanding of the contributions of these different effects will also allow us to more precisely locate the source of the context-sensitivity of relative and absolute adjectives. With these goals in mind, the present study replicates Sedivy et al.'s (1999) design and extends it to absolute adjectives. However, we depart from their design in that we do not use real world

¹ As will be discussed below, our results do not replicate Sedivy's (2005) finding for orthogonal color adjectives. Our data show a clear effect of referential contrast for color adjectives, even though our stimuli consisted of geometric shapes for which the head-noun was fully unpredictable from the color adjective.

objects, but rather geometric shapes. We do so in order to control for head-noun effects that might result from world knowledge. Since geometric shapes are idealized mathematical objects, we assume that they should give rise to less variability based on world knowledge.

Three types of adjectives were tested in one experiment: relative adjectives ($n=9$), absolute adjectives ($n=4$), and color adjectives ($n=4$). Following Sedivy et al. (1999), two critical kinds of visual displays were tested (see Figure 2). In the Contrast condition, the visual display contains: 1) a TARGET object (e.g. a tall cylinder) that participants are requested to click on; 2) a COMPETITOR that shares the target property but presents a different shape (e.g. a tall spiral);² 3) a CONTRAST object that belongs to the same comparison class as the target, but could not be described by the adjective in the instruction (e.g. a short cylinder); and 4) a DISTRACTOR object that could not be described by the adjective in the instruction, nor does it belong to the same comparison class (e.g. a wavy line). The No-contrast condition was created by substituting the contrasting object with a second distractor. With the exception of color-adjective trials, none of the shapes in the visual array shared color. Finally, all absolute-adjective target and competitor pictures were consistent with precise interpretations, as they supported an interpretation of the predicate that was compatible with an end-point oriented standard.

Fifty experimental items (20 containing relative adjectives, 10 containing absolute adjectives and 20 containing color adjectives) were constructed. The adjective-noun pairs tested in the experiment are presented in Table 3.³ Conditions were

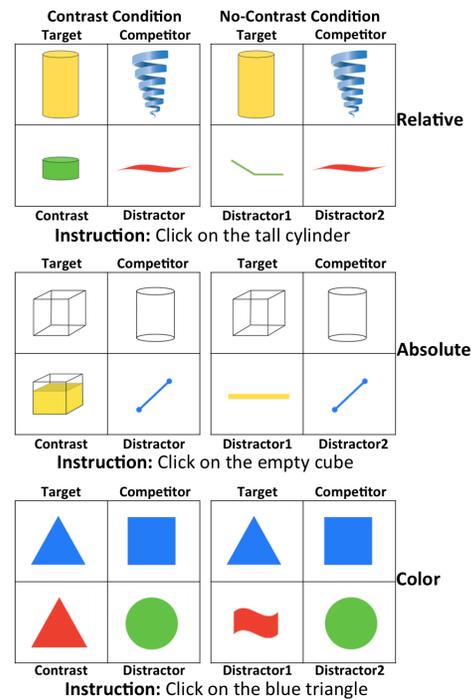


Figure 2: Item Examples

² Our competitor objects differ from Sedivy et al.'s (1999) design in that they all presented the same degree of the adjectival property as the target object, and they could all be felicitously described by the adjective in the auditory instruction. See §2 for details about the properties of the competitors used in Sedivy et al. (1999).

³ The bigram frequency of each of the adjective-noun pairs involving a gradable adjective in Table 3 was calculated using the Brown corpus. Most of the bigrams were unattested. Only three combinations

distributed in two lists using a Latin Square design. Both the order of the trials within each list and the position of the four pictures within each trial were randomized.

Relative	Noun	Absolute	Noun
long	line	closed	circle
short	spiral/cylinder/line	flat	circle/triangle
tall	spiral/cylinder	full	cube/cylinder
big	square/triangle	empty	cube/cylinder
small	square/triangle	Color	Noun
thick	line	blue	circle/triangle/square/line
thin	line	red	circle/triangle/square/line/oval
wide	rectangle/oval	green	circle/triangle/square/rectangle
narrow	rectangle/oval	yellow	circle/triangle/square/line/rectangle

Figure 3: Adjectives-Noun pairs tested in the experiment

Each list was complemented with 60 filler trials. All adjectives used in filler trials were color adjectives (*red*, *green*, *yellow* and *blue*), and pictures always consisted of 2D shapes with plain colors.

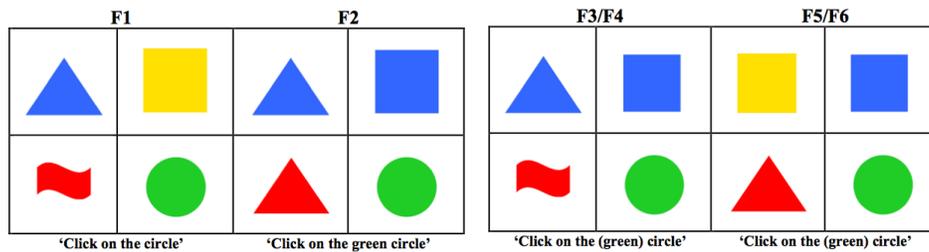


Figure 4: Fillers

Six different types of fillers (10 trials per type) were constructed (see Figure 4). In the first type (F1), none of the figures shares shape or color and the instruction does not contain a modifier. In the second type of filler (F2), the visual display is equivalent to the Contrast condition in the color-adjective trials. However, these filler trials differ from the Contrast condition in that the auditory instruction targets the

were attested: *straight line* occurred 12 times, *thin line* occurred 2 times. Finally, *long line* was attested once.

distractor. In the third type of filler (F3), none of the objects share shape, although two of the pictures share color. The instruction contains a modifier but it does not target any of the two shapes that share color. The fourth type of filler (F4) only differs from F3 in that the instruction does not include a modifier. In the fifth type of filler (F5), none of the figures in the visual array shares color. However, two of the shapes belong to the same comparison class. The instruction contains a modifier and targets one of the two pictures that does not share shape with any of the other pictures in the visual array. Finally, the sixth type of filler (F6) is like F5, except that the instruction does not make use of a color adjective.

We now turn to the predictions made by our experimental design. First, we expect that all adjectives will at least show an informativity-based effect of referential contrast. We use color adjectives as a baseline for the informativity-based effect of contrast linked to the use of restrictive modification.⁴ With respect to relative adjectives, we expect that the presence of the contrasting object in the visual display should be integrated early during lexical processing of the adjective, since the computation of relative thresholds should be facilitated by a contextually supplied comparison class. Any difference we observe between these two types of adjectives would provide empirical evidence for two different sources of the contrast effect. Regarding absolute adjectives, if their thresholds are set via the same mechanism as relative adjectives (Lassiter & Goodman 2013; Lassiter & Goodman 2015), it is expected that the patterns of the effect of referential contrast observed with absolute adjectives should be similar to those found with relative adjectives. Divergence between absolute and relative adjectives would suggest that context-sensitivity of these two kinds of adjectives is driven by different mechanisms.

3.1 Materials

3.1.1 Visual Stimuli

Pictures used in experimental trials as targets, contrasts and competitors (a total of 108) were normed in a series of three description-picture matching studies on MTurk. The purpose of the norming studies was to standardize the interpretation preferences of the visual stimuli within and across adjective types. More specifically, the norming studies ensured that all target and competitor objects were recognized to satisfy the relevant adjectival property, whereas contrast objects (used in the Contrast condition) were recognized to NOT instantiate the relevant adjectival property. Due to space constraints, we do not report further details about the results of the three norming

⁴ Although color adjectives manifest various types of context-dependence (see e.g. Kennedy & McNally 2010), they have not been argued to show the same kind of comparison class-based context dependence that we are interested in here.

studies here. In addition, 81 more images were used as distractors. Whenever possible, distractors were drawn from the pool of objects that had been used as target, competitor or contrast in other trials.

3.2 Auditory Stimuli

Auditory stimuli were recorded in a sound booth by a female native speaker of English. For each recording, the onsets and offset of the adjective were measured in order to determine the mean duration of the three groups of adjectives tested. For relative adjectives, the mean duration was 469 ms (SD = 74.17). For absolute adjectives the adjective mean duration was 442 ms (SD = 53.12). Finally, for color adjectives the mean duration was 466 ms (SD = 46.09). The general mean for all adjective types is 459 ms. A linear regression using adjective type as predictor did not reveal any differences among the three groups ($F(2, 97) = 1.45, p > .2$), ensuring that none of the three types was on average significantly longer than the other two.

3.3 Apparatus

Eye movements were recorded with a Tobii T60 Eye-tracker sampling at 60 Hz. Viewing was binocular and both eyes were tracked, although analyses were performed on data belonging to the right eye exclusively.

3.4 Procedure

Participants saw a visual display with four pictures. Their eye movements were tracked while listening to instructions such as ‘*Click on the tall cylinder*’. Participants were instructed to click on the picture that they thought fitted the description in the auditory instruction best. Only clicks that took place after the offset of the auditory instruction triggered the next trial. There was a 2-second long preview window between the onset of the visual display and the onset of the auditory instruction. Before each trial, a fixation cross appeared in the middle of the screen. A red box framing the cross appeared when participants fixated on it. Participants were instructed to click on the cross when the red box appeared in order to proceed to the next trial. This was done so that eye movements to the four objects could be measured from a default position that was equidistant to the four pictures in the display. At the beginning of the experiment, participants had four practice trials to help them become familiar with the task.

3.5 Participants

Participants were fifty-one undergraduate and graduate students at the University of Chicago (34 females, $M = 20.7$, range 18-34). All participants were native speakers of American English. Undergraduate students did the experiment to fulfill a research awareness requirement for a linguistics course. Graduate students were paid \$10. All participants had normal or corrected to normal vision. Subjects were excluded from data analysis if they met at least one of the following two criteria: 1) track loss for a given subject was higher than 40%; and 2) before the head noun became available, a subject did minimal scanning of any part of the display (i.e., when the aggregated proportion of fixations to the four pictures in the display was $<10\%$ of the total recorded fixations, probably because the subject was only fixating on the fixation cross in the center of the screen). The latter criterion intends to exclude participants who were passively waiting for the head noun information before processing the instruction. The application of these two criteria resulted in the exclusion of 11 subjects. The results reported in the following section correspond to data from 40 participants between the ages of 18-34 (26 females, $M = 20.57$).

3.6 Results

Analyses were performed on the adjective window. The adjective window was calculated using the mean duration for all three types of adjectives (459 ms). For each trial, the adjective window was offset by 200 ms from the onset of the adjective to adjust for the time required to plan and implement an eye-movement. Based on the visual inspection of the results (see figure 5), three time windows of 150 ms each, starting at 350 ms from the adjective onset (including the 200ms offset), were created for data analysis: window 1 (W1) spans from 350-500 ms, window 2 (W2) spans from 500-650. Finally window 3 (W3) was set to 650-800 ms. Given the adjective mean duration, the first two windows (W1, W2) contain fixations reflecting processing of the adjective, whereas window 3 contains fixations reflecting processing of the head noun.

Analyses were run on the aggregated proportion of fixations in each of the three time windows (see figure 5). The first analysis (Analysis I) had the objective of establishing a baseline for processing of the adjectival predicate in the absence of contextual support. Thus, this analysis focused on the No-Contrast condition only. Specifically, for each adjective type we examined how quickly participants started to fixate significantly more on those objects that could be described by the adjectival property (i.e. the target and the competitor), as opposed to those objects that could not (i.e. the two distractors).

We therefore collapsed fixations to target and competitor on the one hand, and

Processing Gradable Adjectives in Context

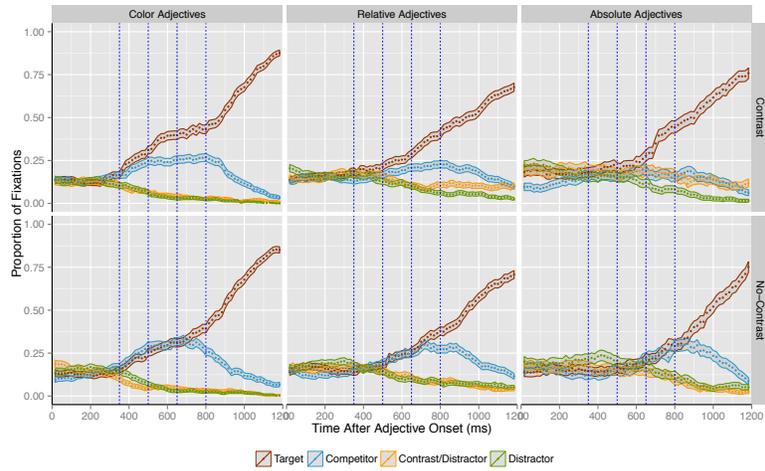


Figure 5: Proportions of fixations to each of the four objects in the display over time starting at the adjective onset for each adjective type. The vertical dashed blue lines mark the boundaries of the three windows defined for data analysis.

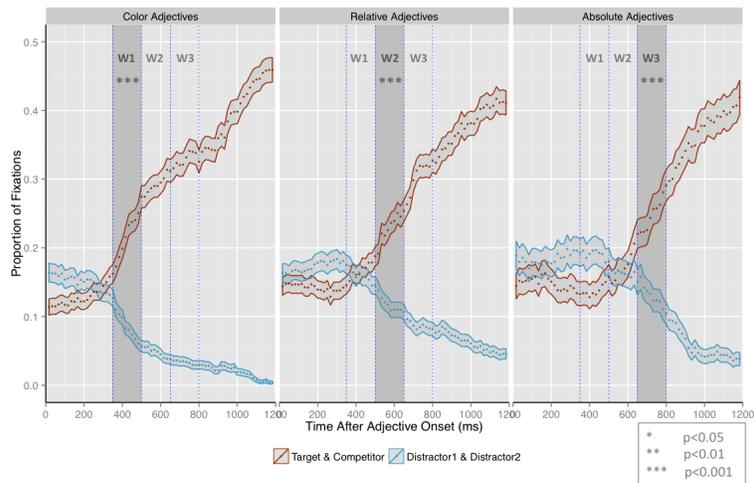


Figure 6: Proportions of fixations to target/competitor vs. distractors (No-Contrast condition) over time starting at the adjective onset. The grayed time windows correspond to the first time window in which a significant difference was found.

fixations to the two distractors on the other. Figure 6 plots the grand average of the proportions of fixations to target/competitor vs. two distractors. Visual identification of the adjective property (looks at target and competitor) is faster for color adjectives than for gradable adjectives (relative and absolute adjectives). ANOVAs run on each time-window and each adjective type showed that in the case of color adjectives, participants' looks to the target and competitor start to significantly diverge from looks to the two Distractors in W1 ($F(1, 39) = 55.57, p < .001$). The first significant effect for relative adjectives takes place in W2 ($F(1, 39) = 43.50, p < .001$), whereas for absolute adjectives, the first significant effect does not take place until W3 ($F(1, 39) = 17.92, p < .001$). The effect remains significant in all subsequent windows for both color (W2 and W3) and relative adjectives (W3).

The second analysis (Analysis II) consisted of a comparison of the Contrast and the No-Contrast conditions with respect to target vs. competitor disambiguation. The goal of this comparison was to determine whether the presence of the contrasting object facilitated target identification in the Contrast condition when compared to the No-Contrast condition. Figure 7 contains the plotted proportions of fixations to these two objects in each condition for each adjective type.

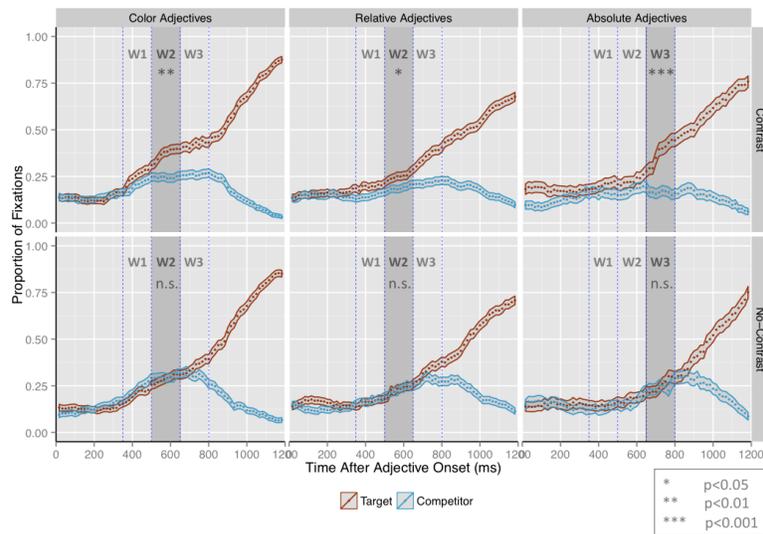


Figure 7: Proportions of fixations to target vs. competitor over time starting at the adjective onset. The grayed time windows correspond to the first time window in which a significant difference was found.

A one-way ANOVA using object type (target vs. competitor) as a factor was run for each time window and adjective type. We found that the presence of the contrasting object facilitated target identification in the contrast condition for all three

adjective types. However, the effect of contrast arises in different time windows. For **color adjectives**, the first significant effect in the Contrast condition takes place in W2 ($F(1, 39) = 8.91, p < .05$), such that participants looked at the target object significantly more than they did to the Competitor. The effect does not reach significance in the same time window of the No-Contrast condition ($F(1, 39) = 0.07, p > .7$), suggesting that at this point participants had not yet discriminated between target and competitor. **Relative adjectives** present a similar situation, with the first significant effect in the Contrast condition arising at W2 ($F(1, 39) = 5.87, p < .05$). Again, the effect is driven by an increase in looks to the target object in this window. The analysis of W2 in the No-Contrast condition did not reach significance ($F(1, 39) = 5.89e-05, p > .9$), revealing that at this point participants had not yet committed to the target object. The first significant effect for **absolute adjectives** in the Contrast condition does not appear until W3 ($F(1, 39) = 13.82, p < .001$), whereas the effect in W3 of the No-Contrast condition is, again, not significant ($F(1, 39) = 0.03, p > .8$).

3.7 Discussion

Our results show that target identification is facilitated in the Contrast condition for all three types of adjectives. However, the patterns of the contrast effects observed with each adjective type differ in significant ways. Our main interest lies in the timing of the contrast effect (Analysis II), relative to the timing at which participants could identify the adjectival property when no contrast was present (Analysis I). We discuss these results below.

The No-Contrast condition alone provided us with a baseline of how quickly participants can identify different types of adjectival properties (in a visual display). Our results show that the color properties are identified faster (W1) than properties introduced by relative (W2) and absolute adjectives (W3). This could either be due to the visual saliency of color, or the fact that the interpretations of both relative and absolute adjectives are more context-dependent and therefore more complex (incurring in higher processing costs), or a combination of both.

When Contrast and No-Contrast conditions are compared, the presence of the contrasting object facilitated the processing of all three types of adjectives, but with different time course. For color and relative adjectives this contrast benefit takes place in W2, whereas for absolute adjectives the effect occurs in W3. By integrating the results of the two analyses performed on the data, it can be observed that, for color adjectives, disambiguation of target and competitor in the Contrast condition (W2) takes place *after* the adjective property is identifiable in the No-Contrast condition (W1). However, for relative adjectives, disambiguation of target and competitor in the contrast condition takes place in the *same* window (W2) in which the adjective

property is identified in the absence of contrast (W2). For absolute adjectives, all the relevant effects take place in a late time window (W3) in which the head noun information is already available in the speech input.

We take the relative timing of the two analyses reported above to be informative regarding the question of whether an observed effect of contrast is driven by the computation of the lexical semantics of an adjective (based on the presence of a threshold variable) or by informativity-based pragmatic reasoning (based on the use of a modifier). **Color adjectives**, which we assume do not evoke a contextually salient comparison class in their interpretation, set up a baseline for our evaluation of the overall results. Most importantly, even though the presence of a contrasting object facilitated the disambiguation between the target and the competitor (during the time window W2), this effect took place after the time window in which the color property itself could be independently identified in the No-Contrast condition (during W1). We take this result to suggest that the contrast effect observed for the color adjectives is driven by informativity-based pragmatic reasoning, the computation of which slightly lags behind the lexical semantic processing that identifies the adjectival property itself. We also note that our findings are not consistent with the results reported in Sedivy (2003), where the effect of contrast for color adjectives only arises when the color property is predictable from the object. Given that we used geometric shapes, in our visual stimuli the relationship between the object and the color property was always unpredictable. We do not have an explanation for this difference between our results and previously reported findings and so we must leave this issue for further research.

In light of the findings from color adjectives, we further suggest that the contrast effect found for **relative adjectives** was mostly driven by the fact that the presence of the contrasting object provides a contextual comparison class that facilitates the computation of the adjectival threshold. The critical difference between color and relative adjectives is that, for relative adjectives, the time window in which the contrasting object facilitated the disambiguation of the target and the competitor (i.e. W2) overlaps with the time window in which the adjectival property could be identified in the No-contrast condition (i.e. discriminating the target and the competitor from the two distractors). We take this to mean that the contrast effect for relative adjectives has (at least partially) a different source compared to color adjectives—the contrasting object provided a contextually salient comparison class and facilitated the lexical semantic processing for relative adjectives. Together, our findings for color and relative adjectives replicate previous results that listeners are very sensitive to the presence of contrast in the visual display, but we provide a more nuanced interpretation of the effects of contrast.

Absolute adjectives present a more complex pattern of results. The presence of a contrasting object also facilitated the identification of the target. However,

the effect of contrast did not appear until the time window W3, by which point information about the head-noun was already available to the hearer. Similarly, the identification of the adjectival property in the No-Contrast condition did not take place until W3. Therefore, the processing of absolute adjectives is overall delayed, compared to the other two types of adjectives. With the data we have, it is difficult to tease apart the effects driven by the head noun information and the effects driven by the processing of the absolute adjectives. But we would like to speculate that imprecision might be playing a role in shaping the observed results on absolute adjectives.

One possibility, is that the delay in lexical processing arises from setting the contextual threshold of imprecision. The idea that absolute adjectives always require setting the value of a contextual threshold of imprecision has been proposed by van Rooij (2011), who adopts a delineation approach to model the relative/absolute distinction. Like most delineation approaches, van Rooij assumes that the variability in interpretation of relative adjectives is driven by comparison class variance (Klein 1980; Sassoon & Toledo 2011; Burnett 2014). Absolute adjectives' insensitivity to context can be straightforwardly explained if these predicates are systematically evaluated with respect to the maximal domain of individuals in the model. Since comparison class variability is eliminated, the criteria of application of these predicates remains constant across contexts and thus no standard-shift effects are predicted to arise with absolute adjectives. Thus, in this account the threshold of absolute adjectives is always endpoint-oriented. Context-sensitive interpretations of absolute adjectives involving non-endpoint-oriented thresholds result from changes in the *contextual standard of precision*. van Rooij formalizes this pragmatic type of context-sensitivity by assuming a set of models \mathbb{M} that share both the domain of contexts C and individuals I , but differ in the standard of precision each of them imposes on the use of the predicate. In other words, each model in \mathbb{M} is associated with a different valuation function V , such that for any two models M and M' in \mathbb{M} , it is the case that $V_M \neq V_{M'}$.

Going back to our results, it is possible that, from a processing perspective, committing to a precise (or close to precise) interpretation of the absolute predicate is costlier than committing to an imprecise interpretation, as a variety of contexts can support an imprecise interpretation, but very few situations can actually license a precise one. If setting the contextual threshold of imprecision is a necessary condition for determining which of the objects in the visual display instantiate the property denoted by the absolute adjective, it is likely that this calculation might be incurring extra processing costs (especially in those cases in which the context imposes high precision requirements, as is always the case for the visual stimuli used in our experiment). If this account is on the right track, it would explain the overall delay in processing found for absolute adjectives in our results. Regarding the effect

of contrast, it is possible that the contrasting object facilitated the computation of the contextual threshold of imprecision, leading to faster identification of the target. This explanation would be compatible with assuming that absolute thresholds are context-independent (Kennedy 2007; Syrett et al. 2009; van Rooij 2011; Burnett 2014; Qing & Franke 2014), and would attribute contextual sensitivity of absolute adjectives to imprecision calculation.

4 Conclusion

The goal of the experiment presented in this paper has been to investigate whether contextual information is integrated in comparable ways during the processing of relative and absolute adjectives. Our results show that relative and absolute adjectives interact with context in different ways. For relative adjectives, contextual information is integrated early as part of lexical processing. Our results constitute further supporting evidence for the claim that relative adjectives have context-dependent thresholds. The picture that has emerged for absolute adjectives is more complex. Absolute adjectives pattern both similar and differently from relative adjectives. Like relative adjectives, absolute adjectives show sensitivity to context. However, absolute adjectives are unique in that they clearly involve higher processing costs, as all significant effects did not take place until the noun information was available. We suggest that imprecision calculation might play an important role both in accounting for the interpretation and the processing cost of absolute adjectives.

References

- Altmann, Gerry & Mark Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30(3). 191–238.
- Burnett, Heather. 2014. A delineation solution to the puzzles of absolute adjectives. *Linguistics and philosophy* 37(1). 1–39.
- Frank, Michael C. & Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998–998.
- Kennedy, Christopher. 2007. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30(1). 1–45.
- Kennedy, Christopher & Louise McNally. 2005. Scale structure and the semantic typology of gradable predicates. *Language* 81(2). 345–381.
- Kennedy, Christopher & Louise McNally. 2010. Color, context, and compositionality. *Synthese* 174(1). 79–98.
- Klein, Ewan. 1980. A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4. 1–45.
- Krifka, Manfred. 2007. Approximate interpretations of number words: a case

- of strategic communication. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 111–126. Humboldt-Universität zu Berlin, Philosophische Fakultät II.
- Lasersohn, Peter. 1999. Pragmatic halos. *Language* 75(3). 522–551.
- Lassiter, Daniel & Noah D. Goodman. 2013. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Proceedings of SALT 23*, 587–610.
- Lassiter, Daniel & Noah D. Goodman. 2015. Adjectival vagueness in a Bayesian model of interpretation. *Synthese*.
- McNally, Louise. 2011. The relative role of property type and scale structure in explaining the behavior of gradable adjectives.
- Qing, Ciyang & Michael Franke. 2014. Gradable adjectives, vagueness, and optimal language use: a speaker-oriented model. In *Proceedings of SALT 24*, 23–41.
- van Rooij, Robert. 2011. Vagueness and linguistics. In G. Ronzitti (ed.), *Vagueness: a guide*, chap. 6, 123–179. Dordrecht: Springer.
- Rotstein, Carmen & Yoad Winter. 2004. Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics* 12(3). 259–288.
- Sassoon, Galit & Assaf Toledo. 2011. Absolute and relative adjectives and their comparison classes.
- Sedivy, Julie C. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32(1). 3–23.
- Sedivy, Julie C. 2005. Evaluating explanations for referential context effects: Evidence for gricean mechanisms in online language interpretation. In John C. Trueswell & Michael K. Tanenhaus (eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, 345–364. Cambridge, MA: MIT Press.
- Sedivy, Julie C., Michael K. Tanenhaus, Craig G. Chambers & Gregory N. Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71(2). 109–147.
- Syrett, Kristen, Christopher Kennedy & Jeffrey Lidz. 2009. Meaning and context in children’s understanding of gradable adjectives. *Journal of Semantics* 27(1). 1–35.
- Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information during spoken language comprehension. *Science* 268(5217). 1632–1634.

Helena Aparicio
Department of Linguistics
University of Chicago
1101 E. 58th St.
Chicago, IL 60637
haparicio@uchicago.edu

Christopher Kennedy
Department of Linguistics
University of Chicago
1101 E. 58th St.
Chicago, IL 60637
ck@uchicago.edu

Ming Xiang
Department of Linguistics
University of Chicago
1101 E. 58th St.
Chicago, IL 60637
mxiang@uchicago.edu