Proceedings of SALT 26: 836–854, 2016

Imprecision is pragmatic: Evidence from referential processing*

Timothy Leffel
University of Chicago

Ming Xiang
University of Chicago

Christopher Kennedy
University of Chicago

Abstract Gradable adjectives (GAs) provide an ideal domain for evaluating theories of the interface betwen semantic interpretation and context: relative and absolute GAs are both context dependent, but absolute adjectives can have precise meanings in a way that relative adjectives cannot. We provide processing evidence for the hypothesis that imprecision in absolute adjectives is a pragmatic phenomenon: absolute thresholds are semantically scalar endpoints, and imprecise uses are derived by a global *pragmatic threshold*, which specifies how much deviation from literal meaning is pragmatically ignorable in a context (see e.g. Lasersohn 1999). The evidence comes from a set of visual world eye-tracking experiments adapted from the design of Sedivy, Tanenhaus, Chambers & Carlson (1999) and Aparicio, Xiang & Kennedy (2015). We measured the time-course of reference resolution in phrases of the form *the Adj Noun* while varying whether *Adj* is relative or absolute, as well as the degree to which the intended referent (target) and a competing object exemplified *Adj*. We found a contrast-based processing advantage for absolute GAs when the target exemplified *Adj* imprecisely but the competitor exemplified *Adj* precisely. However, this advantage was found to be sensitive to the relative/absolute distinction, as well as to the degree to which objects in the display exemplified the relevant scalar property. Collectively, our results provide evidence for a difference in the mental processes underlying relative versus absolute threshold fixing, which we argue is best modeled by a pragmatic theory of imprecision.

Keywords: semantic processing, gradability, vagueness, precision, visual world paradigm

1 Introduction: Context dependence and gradable adjectives

The underlined expressions in (1) and (2) all have context-dependent interpretations: whether a room containing some fixed quantity of people counts as *packed* or *full*,

* Thanks to Dan Lassiter, Christina Kim, Jackson Lee, and audiences at SALT 2016 and U Chicago Language Processing Lab. Thanks *especially* to Helena Aparicio, whose experimental materials were invaluable at every stage of this study. This research was funded in part by NSF Award BCS 1227144.

©2016 Leffel, Xiang, and Kennedy

for example, depends on how big the room is, how many people tend to occupy it, the relevant fire code, etc. Similarly *a lot of people* does not indicate an exact quantity, and frequently neither does *a hundred people* (e.g. in an estimate).

(1) a. The room is packed.
b. There's a lot of people here.
c. Chris is tall.
d. The antenna is long.

(2) a. The room is full.
b. There's a hundred ppl here.
c. Chris is six feet tall.
d. The antenna is straight.

There is also a difference between the expressions in (1) and (2): it is possible to interpret the expressions in (2) precisely, whereas the expressions in (1) are inherently vague. Even though *a lot* and *a hundred* can both name indeterminate amounts in some contexts, *a lot* cannot pick out specific quantities in any context, while *a hundred* clearly names a precise quantity as a matter of literal meaning. We refer to contextually variable use of (2)-type expressions as **imprecision**.

Vagueness is generally thought of as a semantic phenomenon in the sense that the criterion of applicability for a vague expression (the **threshold** or standard, θ) is fixed during the process of semantic interpretation (by a phrase or a variable assignment). Less of a consensus exists about the status of imprecision, however. In this paper we will argue that imprecision is a pragmatic phenomenon involving interpretive mechanisms distinct from those that regulate vagueness.

To investigate the distinction between vagueness and imprecision, we use the empirical domain of gradable adjectives (GAs), which are characterized by their ability to hold of objects to varying degrees, and by their context-sensitivity. GAs form an ideal domain for evaluating theories of context dependence because they fall into two subcategories that differ minimally from one another in the availability of precise interpretations. **Relative standard** gradable adjectives like *tall* and *long* in (1) are inherently vague and context-dependent; **absolute standard** gradable adjectives like *straight* or *full* in (2) can (but need not) be interpreted precisely.[1]

Relative and absolute adjectives are highly similar to one another in many respects: both are demonstrably gradable and participate in degree constructions like comparatives, superlatives, and equatives (*taller, fuller*, etc.). Such similarities provide motivation for degree semantics, according to which gradable expressions are associated with **scales**, or abstract representations of measurement (e.g. *tall*'s scale is height), and truth-conditions are stated in terms of relations between degrees making up those scales (see e.g. Kennedy 1999 and refs therein). For example *John is taller than Mary* receives an interpretation like "John's height is greater than

1 Throughout we focus on maximum standard absolute adjectives (*full*, *straight*, etc.), because they are better suited than minimum standard adjectives (*bent*, *late*) to be compared directly with relative GAs. We sometimes use "absolute adjectives" to abbreviate "maximum standard absolute GAs."

Mary's height."

In degree semantics, vagueness in positive-form (unmarked) relative adjectives is usually modeled in terms of free variables: a sentence like *John is tall* is true iff John's measure of height (given by $\mu_{\mathsf{tall}}(\mathsf{john})$) exceeds a contextually determined threshold for tallness, denoted by a free variable θ_{tall} ranging over degrees of height. Uncertainty about the exact value of θ_{tall} is a central determinant of the vagueness of *tall*.

(3) $\quad [\![\text{John is tall}_{pos}]\!] \rightsquigarrow \mu_{\mathsf{tall}}(\mathsf{john}) > \theta_{\mathsf{tall}}$

θ_A is valued by a variable assignment (or sequence of parameters), similar to pronouns. Call the assignment or parameter that values such variables c, and call whatever free degree variables are valued by c **semantic thresholds**.

Using the distribution of degree modifiers (e.g. *completely* [*full*, **tall*]), Kennedy & McNally (2005) showed that absolute adjectives are associated with scales that have endpoints (closed scales), while relative adjectives are associated with scales lacking endpoints (open scales). A major question in the theory of gradability has to do with whether the free-variable semantics in (3) should be generalized to all GAs. In some sense the null hypothesis should be that there is no semantic distinction between relative and absolute adjectives outside of what knowledge of scale structure provides. According to this position, every gradable adjective A contributes a threshold variable θ_A, and scale structure simply constrains its potential values.[2] Since absolute adjectives denote on closed scales, scalar endpoints are possible values of θ_A, and precise interpretations arise simply when $[\![\theta_A]\!]^c = \max_{>}(D_A)$ (where D_A is the scale for A). But since relative adjectives denote on open scales, endpoint thresholds are not possible, and hence – given uncertainty – neither are precise meanings. This forms the basis of Lassiter & Goodman's (2014) theory of the relative/absolute distinction: the grammar provides no distinction between the two adjective types, and thresholds are uniformly free variables. According to this view, an endpoint threshold has no privileged semantic status – it is just one of many possible choices for θ_A. However, it is one that can be favored by extra linguistic knowledge of distributional differences between objects in the domains of absolute vs relative adjectives. Lassiter and Goodman develop an elegant Bayesian model of pragmatic decision making about variable values that uses this information to promote interpretations of absolute adjectives in which the threshold is at or close to the scalar endpoint.

Other approaches to the relative/absolute distinction hold that there is a fundamental difference between relative and absolute GAs, such that the former have

2 Other factors can of course also constrain the value of θ_A. The role of comparison classes in specifying θ_A becomes important in §2 and §4, when we discuss factors that affect the processing of GAs.

context-dependent semantic denotations of the sort described above, while the latter have denotations in which the threshold is effectively fixed to a scalar endpoint (see e.g. Kennedy & McNally 2005; Kennedy 2007; Toledo & Sassoon 2011; Burnett 2014). As a concrete example, Kennedy & McNally (2005:358) propose that while GAs have a common semantic core, the *pos* morpheme specifies different thresholds depending on adjective type. This distinction can be modeled in different ways; for the purpose of this paper, we will simply assume the existence of a grammatical principle that assigns the positive form of a relative adjective a context-dependent meaning of the sort shown in (4a), and assigns the positive form of a maximum standard absolute adjective the context-invariant denotation in (4b).

(4) a. $[\![pos_{\mathbf{rel}}(A)]\!] = \lambda x[\mu_A(x) \geq \theta_A]$ for relative std. A
b. $[\![pos_{\mathbf{max}}(A)]\!] = \lambda x[\mu_A(x) \geq \mathbf{max}_A]$ for max. std. absolute A

On this view, although both relative and absolute GAs are semantically scalar, the positive form of the former has a denotation that is sensitive to a semantic threshold but the positive form of the latter does not. This approach to the relative/absolute distinction explains why only absolute adjectives have precise interpretations, but must be supplemented by some other, non-semantic mechanism to account for the possibility of imprecise readings. If maximum standards are specified semantically, then interpreting an adjective A imprecisely involves some process of weakening whereby objects close (enough) to $\max_{>}(D_A)$ are treated as though they satisfy the property denoted by the adjective. Drawing on ideas from Kennedy 2007; Potts 2008; Lasersohn 1999, a.o., we introduce the notion of a **pragmatic threshold**, which represents the degree to which an expression's meanings is allowed or expected to diverge from its literal, lexically specified meaning in a context. Viewed another way, a pragmatic threshold (in this case) specifies how large of a final segment of the scale is "close enough" to the upper endpoint. We might model this notion for absolute GAs as in (5) and (6): imprecise interpretations are derived by substituting the literal meaning with an alternative form (a pragmatic alternative) whose standard is "shifted" by θ_A^{imp}.

(5) The pragmatic threshold θ_A^{imp} for a maximum standard GA A is the largest deviation from $\max_{>}(D_A)$ that is pragmatically ignorable in a context.

(6) The set of pragmatic alternatives to A_{pos} is the set[3]
$Alt(A_{pos}) = \{\lambda x.\mu_A(x) \geq (\mathbf{max}_A - \theta_A^{imp}) \mid \theta_A^{imp}$ is a pragmatic threshold$\}$

The intuition is that imprecision arises when part of an utterance is "replaced" by a weaker but otherwise identical form (how much weaker depends on context). This

3 NB the definition of subtraction on degrees will depend on the exact ontology of degrees assumed.

process is distinct from contextual variability introduced by semantic thresholds, not least because pragmatic thresholds need not always be specified, but the end result is – in the case of imprecision – a kind of uncertainty about what is communicated by an absolute adjective that is isomorphic to the uncertainty that arises in an analysis that relies on a semantic threshold.

Assuming unmodified relative adjectives in the positive form always have semantic thresholds, our original question – whether imprecision in absolute adjectives is a semantic or pragmatic phenomenon – can be reformulated as an opposition between the following "semantic" **(HS)** and "pragmatic" **(HP)** hypotheses about imprecision (where c is the assignment or parameter that values free degree variables).

(7) Imprecise interpretations of absolute adjectives are derived via...
(HS) c assigning a non-endpoint value to a semantic threshold variable.
(HP) a global pragmatic threshold relaxing the literal meaning.

Notice that **(HS)** and **(HP)** can deliver identical utterance meanings, at least in the kinds of simple positive-form predications discussed here: any meaning delivered by assigning a non-maximal threshold under **(HS)** can be delivered by pragmatically weakening a maximal threshold under **(HP)**. This suggests that judgments about truth or satisfaction are not the kind of data that will adjudicate between **(HS)** and **(HP)**. Instead, we turn to language processing to compare these two views, based on the following observations. **(HS)** states that the resolution strategy for context dependence should be the same for both relative and absolute GAs; namely, whatever strategy is relevant for fixing the value of a semantic threshold variable (possibly in a probabilistic way, as in Lassiter & Goodman 2014). In contrast, **(HP)** says that the resolution strategy depends on adjective type. This opens the door for methodologies that quantify aspects of semantic versus pragmatic processing (not just contribution to truth-conditions), which are known to have different time-courses and psychological properties. In §3 we present two on-line comprehension experiments the results of which suggest that the process of resolving imprecision in absolute GAs is distinct from the process of fixing the denotation of relative GAs, in a way hat is in line with the hypothesis that the former have pragmatic rather than semantic thresholds. This is expected on **(HP)** but not on **(HS)**, as we discuss in §4.

2 Gradability and referential contrast in the visual world paradigm

The visual world paradigm (VWP) is a format for eye-tracking research that is useful for investigating interactions between visual context and real-time language processing, especially those involving referring expressions (see Huettig, Rommers & Meyer 2011 for a recent review). In a typical VWP trial, a participant views a

collection of objects while being verbally directed to interact with one of them (the referential **target**). During the auditory stimulus, eye-movements to objects in the scene are recorded, and can then be time-locked to regions of interest in the stimulus (e.g. the onset of a word). Fixation patterns across experimental conditions can then be analyzed to yield information about (for instance) how context influences the time-course of syntactic or semantic processing of particular phrases or words.

In an influential VWP study, Sedivy et al. (1999) investigated the incremental processing of simple referring expressions containing attributive adjectives. Expt. 2 of Sedivy et al. (1999) compared the time-course of target-identification for instructions like "touch the tall glass" across two kinds of visual displays: contrast displays (Fig. 1, left), and no-contrast displays (Fig. 1, right).[4] In addition to a target (here the tall glass), all displays contained a **competitor** object (the tall pitcher) that satisfied the adjective but not the noun, thus rendering the instruction ambiguous until noun information is available. Contrast displays also contained a referential **contrast** object (here the short glass), an item satisfying the noun but not the adjective. This allowed Sedivy et al. to evaluate the hypothesis that a contrast object should facilitate referential processing in modified noun phrases (for reasons discussed shortly), an effect which can be detected by analyzing the relative speed of reference resolution in trials with a contrast object versus those without one.

Sedivy et al.'s hypothesis was indeed supported: target identification (operationalized as a high proportion of trials with a look to the target at some time) was established significantly earlier in contrast displays than in no-contrast displays. Critically, the advantage was detectable even before head noun information could have been accessed – suggesting an extremely rapid time-course of the underlying semantic/pragmatic processes. Borrowing terminology from Sedivy (2003) a.o., we refer to this kind of processing advantage as a **referential contrast effect** (RCE).

The RCE as described above is traditionally understood as a combination of Manner- and Quantity-based Gricean reasoning: given a prompt of the form *the Adj Noun*, the comprehender considers the briefer alternative phrase *the Noun*, concluding it must not have been sufficient to identify the referent (or else the speaker would have used it, given that it is briefer and just as informative). But if *the Noun* is not sufficient when *the Adj Noun* is, then two objects must satisfy *Noun* while only one satisfies *Adj Noun*; this conclusion creates a bias in favor of the target in a display like Fig. 1 (left). The modifier *Adj* is thus interpreted in such a way that the existence of a multitude of objects satisfying *Noun* can be inferred *even before noun information is available*. This reasoning results in what is known as the "contrastive inference" associated with restrictive modification, and has been observed repeatedly in referential processing across a variety of experimental

4 They also manipulated the "typicality" of the target, or how well it exemplified the adjective. Typicality as it relates to the present experiments is discussed in §4.

Figure 1 Contrast and no-contrast displays from Sedivy et al.'s (1999) Expt. 2.

paradigms (Sedivy 2002,2003; Grodner, Gibson & Watson 2005; Leffel, Lauter, Westerlund & Pylkkänen 2014, a.o.).

However, Gricean reasoning is not the only potential source of processing facilitation in this kind of context: a contrast object could also facilitate the *lexical* processing of certain kinds of adjectives. For example whether something is *tall* or not depends upon the value of θ_{tall}, which can be "fixed by context." As noted above, relative adjectives are usually assumed to be sensitive to comparison classes, and it is with respect to such a comparison class that θ_{tall} is valued. One way of specifying a comparison class is via head noun information, e.g. the inferred value of θ_{tall} will tend to be higher in the phrase *the tall building* than in *the tall chair* (Rips & Turnbull 1980). Without information about comparison classes, how does one know whether an arbitrary object satisfies a relative adjective? Put another way, to interpret (say) *the tall glass*, one must first interpret *tall*; but interpreting *tall* requires assigning a value to a semantic threshold θ_{tall}; and a salient comparison class constrains possible values of θ_{tall}, thus facilitating the processing of *tall* and hence the entire phrase. This is another source of facilitation in contexts of the variety discussed here: a contrast object can form a comparison class with the target, upon which grounds the target can be classified as *tall*. But in no-contrast displays, it is exclusively properties of the target that are available to determine whether it satisfies *tall*. In other words, the presence of a contrast object could facilitate the lexical processing of GAs, an effect that would percolate to the phrases containing them.

These two potential sources of RCE-type facilitation can be characterized as:

(8) a. **Manner contrast:** processing facilitation guided by Gricean reasoning about the contrastive function of restrictive modification
b. **Threshold contrast:** processing facilitation guided by contextual cues that constrain the value of θ_A (e.g. a comparison class)

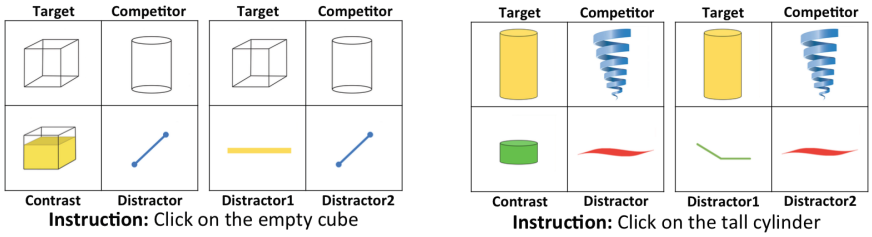

Figure 2 Sample items from Aparicio et al. 2015, for *empty* (left) and *tall* (right). Within each panel, contrast is left; no-contrast is right.

These concepts are, importantly, not mutually exclusive: the processing of *the tall glass* might be aided by (8a) and (8b) – since *tall* is both restrictive and requires a comparison class – whereas RCE-type facilitation in e.g. *the wooden horse* is not attributable (8a) (and hence likely *is* attributable to (8b)), under the reasonable assumption that interpreting *wooden* does not require a comparison class.

A VWP study conducted by Aparicio et al. (2015) suggests that the factors in (8) do indeed interact with lexical semantics to reveal different processing profiles for relative versus absolute gradable adjectives. Using a design similar to Sedivy et al.'s (1999), Aparicio et al. compared reference resolution latencies for definite descriptions containing relative, color, and absolute adjectives as modifiers, in contrast and no-contrast displays. To neutralize effects of typicality (see fn. 4), they used images of artificially constructed geometric shapes instead of real-world artifacts (sample items shown in Fig. 2).

Aparicio et al. found a RCE for both relative and precise uses of absolute adjectives; however, the advantage became significant temporally earlier for relative GAs than it did for absolute GAs. This timing difference provides preliminary evidence for different processing profiles for relative and absolute GAs: relative adjectives induced an earlier RCE because in a visual display like Fig. 2, the contrast object would provide both a manner contrast boost and a threshold contrast boost to the target (right), whereas only a manner-based RCE would apply to absolute targets (left), because in this study they were always precise exemplars.

Summarizing, we have made three key distinctions: (*i*) semantic versus pragmatic thresholds; (*ii*) the semantic and pragmatic hypotheses about imprecision (**(HS)** and **(HP)**, resp.); and (*iii*) Manner-driven versus threshold-driven effects of referential contrast. From these opposing notions, we can extract predictions from **(HS)** and **(HP)** about how manipulation of precision in absolute adjectives should affect processing time-course: if **(HS)** is true, then absolute adjectives have semantic thresholds, and therefore a comparison class should uniformly facilitate processing in the same way that it does for relative GAs. But if **(HP)** is true, then non-

endpoint standards for (lexically precise) absolute adjectives are derived by pragmatic thresholds. A natural expectation on this latter view is thus that threshold contrast effects should appear for absolute GAs only when they are used imprecisely: loosely speaking, absolute GAs become "relative-like" when used imprecisely, so sensitivity to comparison classes (here, threshold contrast) should surface for absolute GAs only when they are interpreted relative to a pragmatic threshold (i.e. imprecisely).

Since the Aparicio et al.'s (2015) study did not test absolute targets that are compatible with non-endpoint standards, it did not provide direct direct information about threshold fixing in absolute adjectives. Next we report two experiments designed to do exactly this: by giving participants the opportunity to select non-endpoint standards, the process of relative vs absolute threshold fixing can be studied more directly.

3 Experiments

We extended the methodology of Sedivy et al. (1999) and Aparicio et al. (2015) in two experiments comparing the time-course of reference resolution in definite descriptions containing gradable adjectives as attributive modifiers. We varyied whether descriptions contained a relative or absolute adjective as a modifier, as well as whether displays contained a contrast object. We also systematically manipulated the relative degree to which targets and competitors exemplified scalar properties. In the first experiment ("**T<C**") targets were exemplars that possess a lesser degree of the relevant property than competitors. This means that in absolute trials, participants were forced to interpret the adjective imprecisely and select a non-endpoint target. In the second experiment ("**T>C**"), targets were exemplars that possesed a greater degree of the adjectival property than competitors, so that an endpoint standard (precise interpretation) would satisfy the target but not the competitor in absolute items. Because of the similarity in format between **T<C** and **T>C**, we describe the methods and materials jointly, and present only results and discussion separately.

3.1 Methods and materials

Participants. A total of fifty-five native English speakers participated the study, and were compensated either $10 or course credit (all $\geq$ 18 y.o.; 36 female). Participants whose rate of tracking loss exceeded 40% were removed from the dataset for analysis, resulting in a sample of 21 participants for **T<C** and 22 for **T>C**.

Stimuli. Auditory stimuli were instructions of the form "click on the Adj Noun," recorded in a sound booth by a female native speaker of English. There were 10 critical auditory stimuli for relative adjective trials, and 10 critical stimuli for absolute maximum standard adjective trials. The adjectives included *long*/*short*, *big*/*small*,

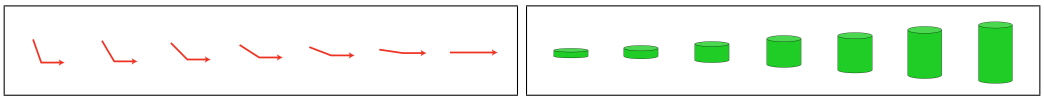

Figure 3 Example image sets for absolute trials containing *straight* (left), and for relative trials containing *tall*/*short* (right)

wide/*narrow*, *tall*/*short*, and *thick*/*thin* for relative standard, and *full*, *empty*, *straight*, *flat*, and *closed* for maximum standard. We also included 10 minimum standard items, but for reasons of space and relevance these are not discussed in this paper. We measured onset and offset times for the adjective in each experimental prompt, resulting in mean adjective durations of 469ms ($sd = 74$) and 442ms ($sd = 53$) for relative and absolute, respectively. There was no significant difference in adjective duration between relative and absolute items (see Aparicio et al. 2015:§3.2). Prompts were also recorded for 80 filler items, some of which did not contain an attributive adjective, and others of which contained color adjectives.

Visual stimuli were images of geometric shapes, arranged in the four corners of 2×2 grids like those in Fig. 1 and 2 (gridlines were not visible to participants). Images were normed to ensure that those used as targets and competitors matched the descriptions they were paired with to a suitable degree. Each non-distractor image was chosen from a set of seven related images, differing only in the degree to which it exemplified some scalar property. All images were thus associated with a scale position ranging from 1 (worst exemplar) to 7 (best exemplar). For example the leftmost image in Fig. 3 (right) is a 1 on the tallness scale, and the rightmost a 7; similarly for Fig. 3 (left) and the straightness scale. Since prompts sometimes utilized both elements of antonym pairs, the leftmost cylinder in Fig. 3 was also used as a 7 on the shortness scale, with the rightmost cylinder as a position 1, for instance.

All visual displays contained images depicting a target object, a competitor object, and a distractor object. The fourth image was either a contrast object (in contrast items) or an additional distractor (in no-contrast items). After data collection was complete, two items were removed from the **T<C** dataset for analysis because they were found to contain errors.

Each of the 20 critical auditory stimuli (relative and maximum absolute adjectives) was paired with two critical visual stimuli (in contrast and no-contrast conditions, distributed across two lists), resulting in 40 critical visual stimuli. The 80 filler items had a variety of forms; ten contained "imprecise" colors (e.g. a maroon target object referred to as *red*), included to reinforce the presence of descriptions not perfectly matching any displayed object.

Factorial design. Both experiments employed the same 2×2 design, with manipulations of adjective type (levels: relative and absolute) and contrast (levels:

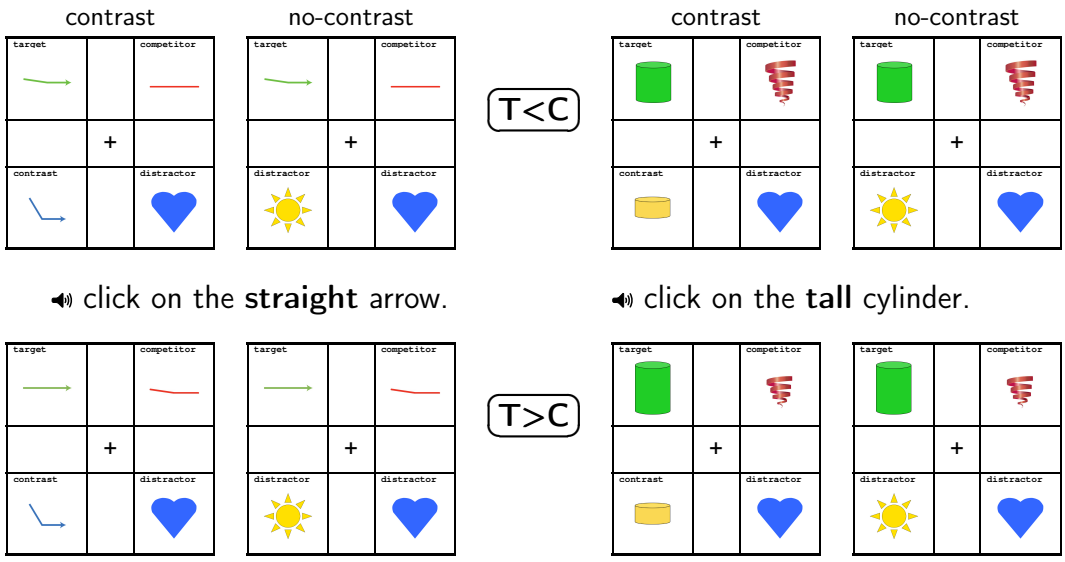

Figure 4 Sample items for **T<C** (top) and **T>C** (bottom), absolute (left) and relative (right). Within adjective type, contrast is left; no-contrast, right.

contrast and no-contrast). All four conditions are illustrated for the **T<C** version and the **T>C** version in Fig. 4 (top) and Fig. 4 (bottom), respectively.

The difference between **T<C** and **T>C** concerned the nature of target and competitor objects, all of which were chosen from one of the sets of 7-point scale images (see Fig. 4). In **T<C**, targets were images with scale position 6 for absolute trials and 5 for relative trials, while competitors had position 7 for absolute and 6 for relative. Target and competitor images were selected from image sets with the same adjective scale but different noun category (e.g. an empty cylinder target with an empty cube competitor).[5] In other words, target objects were uniformly worse exemplars of the adjectives in prompts than were competitors. Importantly, the difference between position-7 and non-position 7 objects for absolute trials is the difference between an object that satisfies an endpoint standard (a precise meaning) and one that doesn't.

In the **T>C** version, targets were objects of position 7 for absolute and 6 for relative, while competitors were position 6 for absolute and 5 for relative. Here, the relationship is the reverse of **T<C**: in **T>C** absolute targets satisfied the precise interpretation of the target adjective, while competitors did not; and relative targets

5 Different positions were chosen for relative and absolute competitors to ensure that the difference was clearly visible for relative trials, but that the lower-positioned items were still reasonably good exemplars of the adjective in absolute trials. The same reasoning applies to targets in **T>C**.

were better exemplars of the associated adjective than were competitors.

Equipment. Auditory prompts were presented over an external speaker set at a fixed volume across participants. Eye-movements were recorded by a Tobii T60 eye-tracker, whose monitor component presented the visual displays. Gaze location samples were drawn at 60 Hz (i.e. a data point was recorded every $\approx$ 16.67ms). Data from both eyes were recorded, but all analyses were performed on right-eye data.

Procedure and trial structure. Before the experiment began, we instructed participants to sit in a comfortable and stable position in front of the monitor. We then calibrated the eye-tracking apparatus on the current participant, and provided them with instructions for completing their task. We told participants that in each trial, they would hear an instruction to click on one of four images appearing on the screen. Before each auditory prompt, there was a two-second delay that we encouraged participants to use to familiarize themselves with the objects on the screen. To ensure eye-movement trajectories originated in a uniform location across participants which was also equidistant from all displayed objects, participants were required to look at a centrally located fixation cross before each trial. Once they fixated the cross, a red box would appear around it and the next trial would begin only after the participant clicked inside the red box. We further informed participants that there may not always be a perfect match for instructions (to prepare them for imprecise descriptions), and that in such a case they should select the object that they thought was the closest match. Participants then completed four practice trials, none of which contained adjectives or critical objects appearing in experimental items. After a final opportunity to ask questions, participants completed all trials while their eye-movements were recorded. The entire procedure lasted 10 to 15 minutes.

3.2 Experiment 1: T<C results and discussion

T<C results. For each participant and trial, a series of raw gaze-locations was converted into a sequence of binary data points indicating which of the four regions containing an object (if any) the participant was looking at during each time point. The temporal region of interest for the present experiments begins at the onset of the adjective in the prompt, because it is only at this point that looks directed towards a particular object can be attributed to information introduced by the adjective.

For each time point, we averaged looks to individual objects over items, yielding a dataset that specifies mean proportion of looks to each object in each time point and condition. These data (further collapsed over subjects) are plotted for targets versus competitors in Fig. 5, as a function of time (ms) after adjective onset. Mean noun onset time within each adjective type is represented by a blue vertical line, pushed forward 200ms to compensate for eye-movement planning.

We are interested primarily in comparing – across conditions – when participants

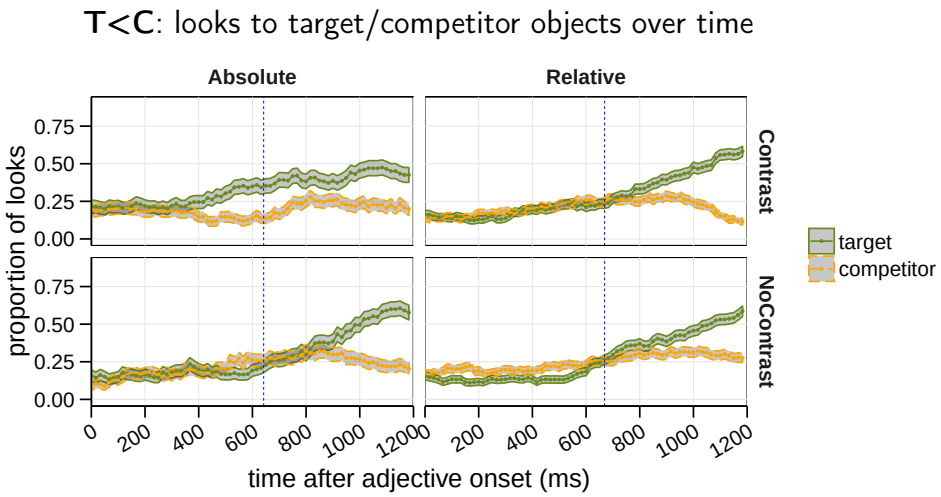

Figure 5 Mean proportion of looks to target (green/solid) versus competitor (orange/dashed), averaged over subjects and items within each condition. Absolute (left), relative (right); contrast (top), no-contrast (bottom). Ribbon width represents standard error.

are significantly more likely to fixate the target than they are the competitor: when looks to the target become reliably more frequent than looks to the competitor, this is an indication that the referent has been correctly identified, and hence that the phrase has been successfully processed (or at the very least that there is a meaningful bias toward the correct referent). Thus the goal is to assess how the temporal relationship between looks to the target and looks to the competitor differs across contrast and no-contrast trials (separately for relative/absolute). The RCE occurs when looks to the target diverge from looks to the competitor earlier in contrast displays than in the corresponding no-contrast displays.

To analyze the data statistically, we first defined four consecutive time-windows of 150ms in duration, beginning at 350ms post-adjective onset. Any activity during the first two windows is attributable to adjective effects, since noun information is unavailable until around 650ms. For each adjective type and time-window, we then performed repeated-measures ANOVAs on look-proportions, with object (target versus competitor) and condition (contrast versus no-contrast) as factors.

Absolute adjective type. A repeated-measures ANOVA revealed a highly significant interaction between object (target/competitor) and condition (contrast/no-contrast) in the second (500-650ms) time-window, such that looks to the target were significantly higher in contrast items than in no-contrast items ($F(1,20) = 16.40$,

$p < .001$; generalized eta-squared $\eta_G^2 = .16$).[6] This indicates the presence of a referential contrast effect – a target-identification advantage in the presence of a contrast object. The interaction remained marginally significant in the third window ($F(1,20) = 3.31$, $p < .09$), albeit with a smaller effect size of $\eta_G^2 = .05$.

Relative adjective type. The same analysis failed to identify significant interactions between target/competitor and contrast/no-contrast in all time-windows (smallest $p > .22$). This indicates the absence of a referential contrast effect for relative adjectives in this particular experiment: the presence of a contrast object resulted in no detectable advantage in target-identification speed.

T<C summary. Our results demonstrate differential effects of referential contrast on the processing of relative and imprecise absolute GAs. For descriptions containing absolute GAs as modifiers, participants identified targets more quickly in the presence of a contrast object than without one. Even before noun information is available (in the second window), there is a clear RCE that is manifest as an earlier preference for targets in contrast items than in no-contrast items.

Interestingly, we found no significant effects of referential contrast in the processing of relative adjectives: looks to targets rose at roughly similar rates across contrast and no-contrast displays for all windows analyzed. This differs from the absolute results and from Aparicio et al.'s (2015) finding of a strong and early contrast-based facilitation for relative GAs. We return to these differences in §4.

3.3 Experiment 2: T>C results and discussion

T>C results. The analytical logic and statistical procedures detailed in §3.2 were applied to the results of **T>C**: using repeated-measures ANOVAs, we compared mean looks to target versus competitor in four consecutive time-windows, in order to quantify the effect of a contrast object on reference resolution latency.

Absolute adjective type. The analysis revealed only a brief, weak, and marginally significant interaction between object and condition in the second time-window, such that looks to the target were slightly higher in the contrast condition than in the no-contrast condition ($F(1,25) = 3.51$, $p < .08$; $\eta_G^2 = .03$). This is the only window during which there was even a marginal target advantage in contrast over no-contrast, indicating a weak or non-existent RCE for absolute adjectives when the target was precise and the competitor was not.

Relative adjective type. The analysis revealed a significant interaction between object and condition for relative trials, but it appeared only in later time-windows. In the third window, there was a marginally significant and small advantage for

6 Generalized eta-squared (η_G^2) is an effect-size measure for certain repeated measures designs (Olejnik & Algina 2003); a value of .16 represents a medium-sized effect on Bakeman's (2005:383) suggested classification.

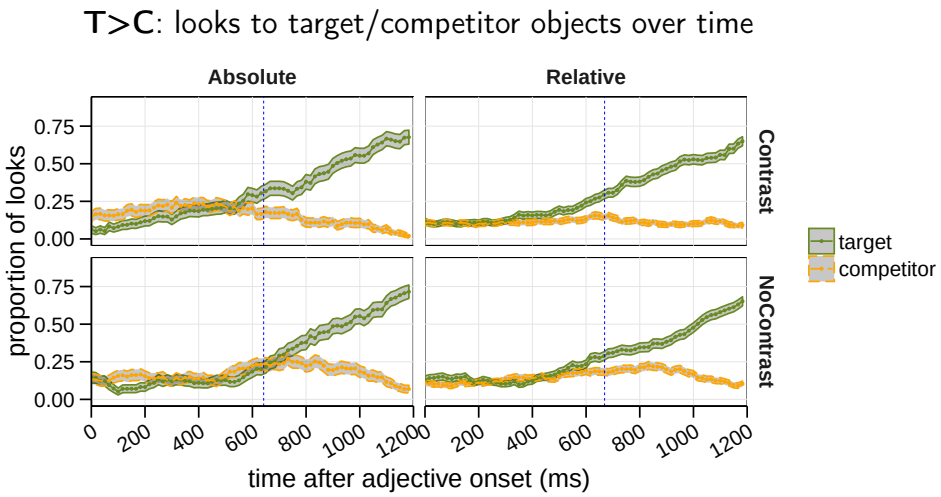

Figure 6 Mean proportion of looks to target (green/solid) versus competitor objects (orange/dashed); ribbon width represents standard error.

targets in contrast over no-contrast ($F(1,25) = 3.16$, $p < .09$; $\eta^2_G = .02$). But the interaction persisted and became stronger in both magnitude and reliability (in the same direction) during the fourth window, with $F(1,25) = 13.08$, $p < .005$; $\eta^2_G = .07$. This pattern indicates a late-occurring RCE in relative items when the target was a better exemplar than the competitor.

T>C summary. For absolute adjectives we found only a marginal referential contrast effect with precise targets and imprecise competitors. Relative adjectives, on the other hand, are associated with a late-occurring contrast effect when the target is a better exemplar than the competitor. Both findings deviate from the results in the first experiment **T<C**, in which a RCE was observed for absolute but not relative.

4 General discussion

In two experiments, we found that an object that introduces referential contrast affects the processing of relative and absolute adjectives in different ways. We saw that when imprecise interpretations are made salient, the presence of a contrast object facilitated the processing of absolute adjectives only when the target was imprecise (**T<C**). On the other hand, the processing of relative adjectives was facilitated when the target was a better exemplar than the competitor (**T>C**), but not when the relationship was reversed (**T<C**). Recall our earlier suggestion that RCEs observed in the literature (e.g. Sedivy et al. 1999) arise from two different sources: the presence of a contrast object can trigger both a manner-based facilitation effect guided by Gricean reasoning, and a threshold-based facilitation effect that speeds

the lexical processing of gradable adjectives by making salient a comparison class. Previous accounts of referential contrast have primarily focused on the Manner-based effect (see e.g. Sedivy 2002; Grodner et al. 2005; Grodner & Sedivy 2011), but the current results support the view that at least part of the RCE is based on threshold contrast. This conclusion is based on the reasoning that a manner-based RCE should be constant across attributive adjectives, and would not explain the observed differences between absolute and relative adjectives. However, if the contrast object influences how thresholds are set for GAs, and in the meantime if relative and absolute adjectives resort to different mechanisms to compute their thresholds, we have the basis for an explanation of these differences.

Let's start with relative standard GAs, which (we assume) have interpretations that are dependent on semantic thresholds. As discussed in §1, the semantic threshold of a relative adjective is constrained by many contextual factors, one of which is a comparison class: a set of objects that are similar in some key respect, but which vary along the dimension measured by the adjective. We hypothesize that in our experiments, the target, contrast (when present), *and* competitor form a comparison class: all are similar types of abstract shapes, and they manifest a continuum of values along the relevant scales (size, fullness, straightness, etc.). We further hypothesize that, given a comparison class, threshold selection is sensitive to a general "strongest meaning" principle that favors a higher semantic threshold, which makes the adjective true of fewer objects in the comparison class, over a lower threshold, which makes it true of more objects (cf. van Deemter 2006). If this is correct, then the absence of a referential contrast effect for the target in the **T<C** experiment can be explained in terms of a conflict between the high threshold preference and the Manner contrast effect: the former favors selection of a standard that makes the adjective true of the competitor but false of the target, while the latter favors selection of a standard that makes the adjective true of the target (and of the competitor). In contrast, in the **T>C** experiment, in which the target had higher a scale position than the competitor, the high threshold preference and the Manner effect both favor selection of a standard that makes the adjective true of the target, and we observe a corresponding facilitation in reference resolution.

Turning to maximum standard absolute GAs, the results of the **T<C** experiment already indicate that **(HS)** – the hypothesis that imprecise interpretations of absolute adjectives involve selection of a non-maximal semantic threshold – is incorrect. If the reasoning outlined above is correct, there should be a preference in this scenario for a high threshold; in this case one that would make the adjective true of the competitor, which has a maximal degree of the measured property, and false of the target, which satisfies the property only imprecisely. Such a threshold would be in conflict with the effects of Manner contrast, as with relative adjectives, and would eliminate the referential contrast effect. But this is not what we saw; indeed, we saw

exactly the opposite: there was a strong referential contrast effect in **T<C**, but only a marginal and weak one in **T>C**.

We claim that this pattern is best explained by **(HP)**, the view that imprecision involves calculation of a pragmatic threshold of imprecision, i.e. a determination of how much deviation from the literal, maximum-standard denotation of an absolute adjective is tolerated in context. The result of such a computation is a *weakening* of the semantic content of the adjective; it is therefore no surprise that we do not see a high threshold preference (favoring the competitor) and corresponding conflict with the Manner contrast effect (favoring the target) in the contrast condition of the **T<C** experiment, as was the case for relative adjectives. Instead, we claim, the presence of the contrast item facilitates reference resolution to the target because Manner contrast and threshold contrast are aligned: the former favors resolution to an object of the same category as the contrast, and the latter facilitates identification of an appropriate pragmatic threshold for imprecision, resulting in rapid fixation on the target object. In the **T>C** experiment, however, the target is a precise exemplar, obviating the need for pragmatic threshold calculation and eliminating any processing benefit that results when Manner and threshold contrast are aligned. The result is only a marginal facilitation of reference resolution to the target, arising (presumably) from Manner contrast alone. While it is somewhat puzzling that the effect is not stronger or more reliable than it is, the results we obtained trend in the same direction as those reported by Aparicio et al. (2015), who found a significant, but small and delayed RCE for precise exemplars of absolute adjectives in the presence of a contrast item.

5 Conclusions

In two eye-tracking experiments, we provided evidence that imprecise interpretations of absolute adjectives involve a different kind of context dependence than is involved in the interpretation of vague relative adjectives. We suggest that the difference be modeled in terms of different kinds of thresholds: the "semantic" thresholds associated with relative threshold are valued by a free degree variable present in the lexical semantics; but the "pragmatic thresholds" associated with absolute adjectives are shifted by global pragmatic parameters that specify a degree of tolerance in deviation from literal meaning in a context.

We have focused on the particular case of relative and absolute GAs here because these two expressions form a semantic natural class, but we expect that our results will generalize to a broader class of expressions for which precise meanings are available but not obligatory, including measure phrases, numerals, etc. If correct, the relative/absolute distinction is just one instance of a more widespread kind of interaction between the symbolic system that builds sentence meanings ("semantics"), and more domain-general reasoning capacities that lead us to draw inferences about

speaker meanings ("pragmatics") on the basis of blending world knowledge with the output of the symbol-manipulation system.

References

Aparicio, Helena, Ming Xiang & Chris Kennedy. 2015. Processing gradable adjectives in context: A visual world study. In Sarah D'Antonio, Mary Moroney & Carol Rose Little (eds.), *Semantics and Linguistic Theory (SALT)* 25, 413–432.

Bakeman, Roger. 2005. Recommended effect size statistics for repeated measures designs. *Behavior Research Methods* 37(3). 379–384.

Burnett, Heather. 2014. A delineation solution to the puzzles of absolute adjectives. *Linguistics and Philosophy* 37(1). 1–39.

van Deemter, Kees. 2006. Generating referring expressions that involve gradable properties. *Computational Linguistics* 32(2). 195–222.

Grodner, Daniel, Edward Gibson & Duane Watson. 2005. The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition* 95(3). 275–296.

Grodner, Daniel & Julie Sedivy. 2011. The effect of speaker-specific information on pragmatic inferences. In Edward Gibson & Neal Pearlmutter (eds.), *The Processing and Acquisition of Reference*, 239–272. MIT Press.

Huettig, Falk, Joost Rommers & Antje Meyer. 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* 137(2). 151–171.

Kennedy, Christopher. 1999. *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. Garland.

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, degree modification, and the semantics of gradable predicates. *Language* 81(2). 345–381.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 75(3). 522–551.

Lassiter, Daniel & Noah Goodman. 2014. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In Todd Snider, Sarah D'Antonio & Mia Weigand (eds.), *Semantics and Linguistic Theory (SALT)* 24, 587–610.

Leffel, Timothy, Miriam Lauter, Masha Westerlund & Liina Pylkkänen. 2014. Restrictive versus non-restrictive composition: An MEG study. *Language, Cognition, and Neuroscience* 29(10). 1191–1204.

Olejnik, Stephen & James Algina. 2003. Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods* 8(4). 434.

Potts, Christopher. 2008. Interpretive economy, Schelling points, and evolutionary stability. Manuscript, UMass Amherst.

Rips, Lance & William Turnbull. 1980. How big is big? Relative and absolute properties in memory. *Cognition* 8. 145–174.

Sedivy, Julie. 2002. Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language* 46(2). 341–370.

Sedivy, Julie. 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, Michael Tanenhaus, Craig Chambers & Greg Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71(2). 109–147.

Toledo, Assaf & Galit Sassoon. 2011. Absolute vs. relative adjectives: Variance within vs. between individuals. In Neil Ashton, Anca Chereches & David Lutz (eds.), *Semantics and Linguistic Theory (SALT)* 21, 135–154.

Timothy Leffel
National Opinion Research Center
at The University of Chicago
leffel-timothy@norc.uchicago.edu

Ming Xiang
Department of Linguistics
The University of Chicago
mxiang@uchicago.edu

Chris Kennedy
Department of Linguistics
The University of Chicago
ck@uchicago.edu