

Chapter 1

Perceived Informativity and Referential Effects of Contrast in Adjectivally Modified NPs

Helena Aparicio, Christopher Kennedy and Ming Xiang

Abstract Referential Effects of Contrast (RECs) involving reference resolution of adjectivally modified NPs (e.g. *the tall glass*) have been attributed to pragmatic reasoning based on the informativity of modification (Sedivy et al. 1999; Sedivy 2003, 2004, a.o.). Recently, it has been claimed that informativity alone cannot account for all the attested interactions between adjectival meaning and context and that factors related to efficiency in the search of a referent also play an important role (Rubio-Fernández 2016). Building on Aparicio et al. (2015), this paper demonstrates that perceived informativity plays an important role in RECs, but lexical semantic properties of different adjective classes are also relevant. We present results from a Visual World eye-tracking study which shows that adjective classes differ in whether they introduce RECs, and results from an offline judgment task which show that this difference correlates to some extent with the perceived informativity of members of these classes. Color adjectives, relative adjectives and maximum standard absolute adjectives were rated as overinformative when used as modifiers in the absence of contrast, and gave rise to RECs; minimum standard absolute adjectives were not rated as overinformative when used as modifiers in the absence of contrast, and did not give rise to RECs. Taken together, our results show that perceived informativity plays an important role in RECs. We also discuss additional differences between the adjective classes which suggest that differences in lexical semantics can further contribute to differences in RECs.

Helena Aparicio,
Massachusetts Institute of Technology, e-mail: haparici@mit.edu
Christopher Kennedy, Ming Xiang,
University of Chicago, e-mail: {ck, mxiang}@uchicago.edu

1.1 Perceived Informativity and Referential Effects of Contrast

There exists ample evidence that listeners process linguistic input incrementally (Crain and Steedman 1985; Altmann and Steedman 1988; Eberhard et al. 1995, among many others), and that pragmatic information pertaining to different sources is quickly integrated during online processing (Hanna et al. 2003; Hanna and Tanenhaus 2004; Grodner and Sedivy 2011). For instance, in a classic eye-tracking study, Tanenhaus et al. (1995) showed that contextual visual information, introduced by the manipulation of the visual display, was immediately adopted by the listeners to guide their online parsing decisions. This experimental paradigm, which later came to be known as the Visual World (VW) paradigm, has proven especially sensitive in detecting effects of context during online processing. In VW eye-tracking experiments, participants' eye-movements are tracked as they look at arrays of objects while listening to an auditory instruction that typically requires them to visually identify an object in the display in order to perform the experimental task. Eye-movements are a particularly good measure of language processing in reference-resolution tasks because eye-fixations reflect with millisecond granularity what objects in the visual context are being considered as potential referents of the linguistic input (Cooper 1974; Eberhard et al. 1995; Tanenhaus et al. 1995; Pyykkönen-Klauck and Crocker 2016). Therefore, eye-movement patterns can be used to make inferences about whether and at what point of linguistic processing the information of the visual context becomes relevant.

Within Visual World studies, a hallmark of this rapid online integration of pragmatic information comes from *Referential Effects of Contrast* (henceforth RECs). The effect was initially reported by Sedivy et al. (1999) in a study investigating how properties of the visual context influenced the processing of NPs containing an attributive prenominal adjective 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. A **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. All trials contained a competitor object that presented a higher degree of the property in the instruction when compared to the target, but could not be felicitously described by the adjective (e.g. a pitcher that was taller than the glass, but was itself not tall for a pitcher, see Figure 1.1).

The main finding of the experiment was that participants' fixations converged on 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 at a point in which the head noun had not yet been processed. Therefore, this decision was performed at a time in which the linguistic instruction was still ambiguous between the two objects that could be described by the adjective in the instruction

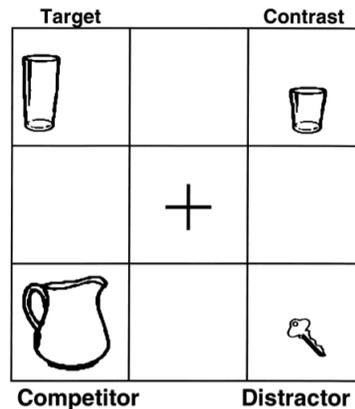


Fig. 1.1 Experimental set up used by Sedivy et al. (1999) (Contrast Condition).

(i.e. the target and the competitor), suggesting that the presence of the contrasting object was used very early.

Despite the fact that RECs have been consistently replicated with adjectivally modified NPs (Sedivy et al. 1999; Sedivy 2004; Weber et al. 2006; Grodner and Sedivy 2011; Wolter et al. 2011; Aparicio et al. 2015; Leffel et al. to appear), the exact mechanisms underlying these effects are not fully understood, and it remains an open question whether all the RECs reported in the literature are born equal (cf. Sedivy 2003, 2004). The crucial difference between the Contrast condition and the No-Contrast condition is that in the former, the visual display includes objects that contrast only with respect to the information provided by a noun modifier, not with respect to the information provided by the head noun; while in the latter all objects in the display contrast with respect to the information provided by the noun. This makes the use of a modifier non-contrastive or “redundant,” since the head noun alone suffices to distinguish the intended referent from the other objects in the display. A referential contrast is observed when visual target identification takes place significantly faster in the Contrast condition compared to the No-Contrast condition. Such effects receive a natural pragmatic explanation in terms of the interaction of the Gricean Maxims of Quantity and Manner (Grice 1975). Since a definite description with a restrictive modifier is both more complex and more informative than a corresponding description without a modifier, a speaker’s use of a modified form provides an indication that she intends to refer to an object that contrasts relative to the modifier but not the noun, which in turn facilitates referential fixation in the Contrast condition but not in the No-Contrast condition.

A naive version of the Gricean account of RECs would lead to the expectation that (cooperative) uses of modifiers should be restricted to contexts involving contrast; i.e., contexts in which the modifier is not redundant, in the sense described above. However, there is evidence that speakers frequently use modifiers in referential NPs even in the absence of contrast (Pechmann 1989; Nadig and Sedivy 2002;

Sedivy 2003; Maes et al. 2004; Sedivy 2004; Koolen et al. 2011). Certain patterns seem to emerge in the use of such apparently redundant adjectives. Experimental production tasks have consistently shown that color adjectives are more likely to be used redundantly than other classes of adjectives like dimensional or material adjectives (Pechmann 1989; Belke and Meyer 2002; Nadig and Sedivy 2002; Sedivy 2004). Several factors have been found to be good predictors of when a speaker is more likely to use a redundant adjective. For instance, color adjectives that denote a stereotypical property of the object (e.g., a *yellow* banana) are less likely to be used redundantly (Sedivy 2003), while atypical color adjectives are more likely to be used redundantly (Westerbeek et al. 2015). A second factor affecting the production of redundant adjectives in referential communication tasks is the amount of variation present in the visual scene. Speakers are more likely to utter an overspecified description when the visual scene contains color variability, i.e. the visual display is polychrome, than when it does not, i.e., the visual display is monochrome (Koolen et al. 2013; Rubio-Fernández 2016).

The fact that speakers not only often choose to include overspecified adjectives as part of their utterances, but also do so in systematic ways is unexpected in the context of the naive Gricean view, in which all redundant adjectives are suboptimal from an informativity point of view. Rubio-Fernández (2016) suggests that overspecification should be recast in terms of efficiency rather than informativity, as modifiers may facilitate target identification by helping the hearer optimize the visual search of the target object (see Paraboni et al. 2007; Arts et al. 2011 for similar claims). In this respect, efficiency can be regarded as a pragmatic cooperative phenomenon. Assuming that hearers are sensitive to the systematicities in the production patterns of redundant adjectives, different adjective classes could in principle be associated with different expectations regarding the probability that a given adjective will be used contrastively. This is relevant for VW experiments such as the ones discussed above, as it leads to a more nuanced prediction than the naive Gricean view, namely that only those adjective classes for which a redundant adjective is *perceived* as providing too much information in the context should give rise to such effects, i.e. there should be a correlation between perceived overinformativity and strength of referential contrast. The resulting picture, like the naive Gricean one, remains rooted in reasoning about (over-)informativity of a complex form, but allows for variation in classes of modifiers based on the extent to which they are independently perceived as over-informative or not.

To test this hypothesis we conducted two experiments to explore the relation between RECs and perceived informativity. In Experiment 1 (§1.2), we extend a prior study of RECs in so-called “relative” vs. “absolute” adjectives by Aparicio et al. (2015) to the class of “minimum standard” absolute adjectives. We show that minimum standard absolute adjectives fail to trigger RECs, in contrast to the relative and maximum standard absolute adjectives analyzed by Aparicio et al., as well as to color adjective controls. In Experiment 2 (§1.3), we compare all four classes of adjectives for perceived informativity, and show that minimum standard adjectival modifiers differ from all the other classes of adjectival modifiers in not being perceived as overinformative in the absence of contextual support for contrastive inter-

pretations, in support of the perceived informativity-based view of RECs described above. However, among the other three classes of adjectives, we also found that the magnitude of the perceived (over)informativeness does not completely map to the size of the RECs reported in Aparicio et al. (2015). We conclude with discussion of the role that lexical semantic factors may play in driving perceived informativity and variable RECs.

1.2 Experiment 1: Variable RECs across Adjective Classes

In a VW study modeled after Sedivy et al.'s (1999) design, Aparicio et al. (2015) examined RECs in definite descriptions containing modifiers from three classes of adjectives: relative adjectives, maximum standard absolute adjectives and color adjectives. (For general discussion of these adjectives and their semantic and pragmatic properties, see Unger 1975; Pinkal 1995; Rotstein and Winter 2004; Kennedy and McNally 2005; Kennedy 2007; McNally 2011.) Aparicio et al.'s decision to examine these adjectives was based on an interest in the potential role that different kinds of context dependence play in the interpretation of adjectives generally, and in the generation of RECs in particular. Relative adjectives (RelAs) such as *big*, *small*, *tall* and *short* are inherently context-sensitive, because their "threshold" for application can change across contexts. For example, the threshold for determining what individuals fall in the extension of the predicate '*tall*' will be significantly higher in a discussion about basketball players (who tend to be taller than average) than in a discussion about jockeys (who tend to be shorter than average). The set of objects or individuals used to determine the threshold of relative adjectives, e.g. basketball players vs. jockeys, is usually referred to as the comparison class, and is one of the parameters that plays a role in fixing the extension of a relative adjective in context.

Maximum standard absolute adjectives (MaxAAs) like *full*, *empty*, *straight* and *flat* manifest a different type of context dependence. Unlike RelAs, MaxAAs have context *independent* uses that are true of an object just in case it manifests a maximal degree of the relevant property. In such an use, '*empty*' is true of a cookie jar, for example, just in case it contains no cookies at all. MaxAAs also have uses that tolerate deviation from a maximal degree, however: in many contexts, a cookie jar containing just one or two cookies could be felicitously described as empty (especially if the goal is to get someone to fill it again). A question of current research is whether such uses of MaxAAs arise from the same semantic principles that regulate context dependent interpretations of RelAs, or whether they involve a pragmatic phenomenon of "imprecise" uses of expressions with context invariant denotations (see e.g. Sassoon and Toledo 2011; Lassiter and Goodman 2013, 2015; Qing and Franke 2014; Leffel et al. to appear).

Although the study in Aparicio et al. (2015) did not address this question directly, it provided a baseline examination of the processing of RelAs vs. precisely interpreted MaxAAs used as modifiers in definite descriptions, with color adjectives

(ColAs) as a control.¹ Following Sedivy et al. (1999), two critical kinds of visual displays were tested, illustrated in Figure 1.2. 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); 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. Aparicio et al. found that all three adjective types displayed RECs, though there were differences in the time-course of the effects: for ColAs and RelAs, RECs appeared before information about the head noun was available to participants. However, in the case of MaxAAs the REC was delayed and did not obtain until the noun window. This led the authors to conclude that lexical processing can also play an important role in further shaping RECs, a point to which we return in Section 1.4.

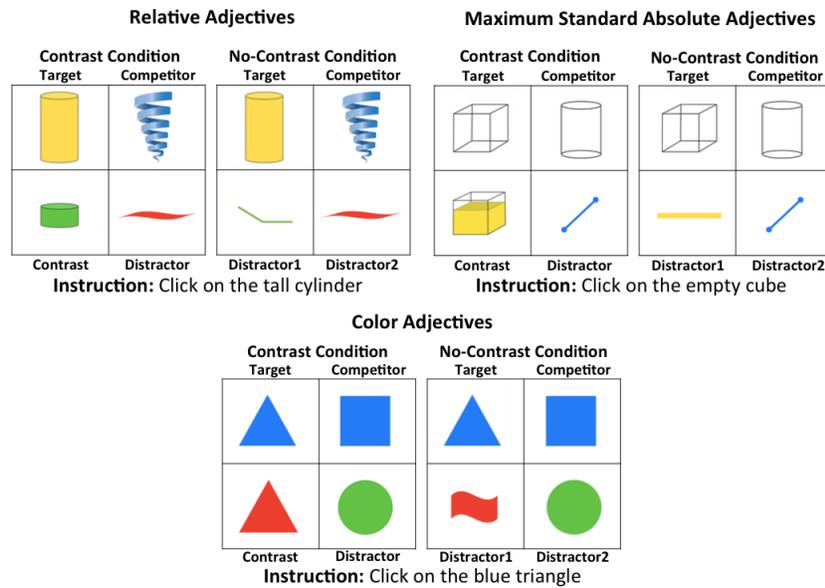


Fig. 1.2 Item Examples (Aparicio et al. 2015).

Our experiment extends the Aparicio et al. design to a second class of absolute adjectives: “minimum standard” absolute adjectives (MinAAs) such as *bent*, *spot-*

¹ Although color adjectives are both context dependent and vague, they are sensitive to different kinds of contextual parameters from RelAs and MaxAAs. See Rothschild and Segal (2009); Kennedy and McNally (2010); Clapp (2012) for discussion.

ted, *bumpy* and *striped*. Like MaxAAs, MinAAs have context invariant uses, but unlike MaxAAs, they merely require their arguments to have greater than a minimum degree of the relevant property. A *bent rod*, for example, is a rod with some degree of bend; and a *spotted shirt* is a shirt with some number of spots.² Our goal in examining MinAAs was both to fill out the empirical picture of RECs in relative vs. absolute adjectives that was only partially provided in the Aparicio et al. study, and to identify potential differences in REC effects among natural classes of adjectival modifiers.

1.2.1 Design

Following Aparicio et al. (2015), we used geometric shapes to construct the visual stimuli with the goal of controlling for potential effects of world-knowledge about artifacts on adjective interpretation. Six MinAAs were included in one experiment, which are listed in Table 1.1.

Min. St. Absolute Adjective	Noun
bent	line
bumpy	square/triangle
curved	line
open	circle
spotted	square/circle
striped	square/triangle

Table 1.1 Adjective-Noun pairs tested in Experiment 1.

² Several linguistic tests diagnose whether an absolute adjective makes use of a maximum vs. minimum vs. relative standard. For instance, Kennedy (2007) points out that these three classes give rise to different entailment patterns when used in comparatives. In comparatives of the form *X is more A than Y*, MinAAs entail that X is A (i); MaxAAs entail that B is not A (ii); and (unmarked) RelAs entail neither that X is (not) A nor that Y is (not) A (iii).

- (i) a. The red towel is wetter than the blue towel. \Rightarrow
b. The red towel is wet.
- (ii) a. The red towel is drier than the blue towel. \Rightarrow
b. The blue towel is not dry.
- (iii) a. The red towel is bigger than the blue towel. \Rightarrow
b. The red towel is (not) big.
c. The blue towel is (not) big.

The distribution of modifiers like *slightly* and *completely* are also often described as tests for MinAA and MaxAA status, respectively, but strictly speaking, these modifiers test for minimum and maximum scalar endpoints, respectively, which are independent of — though generally correlated with — maximum and minimum standards.

Two conditions were tested (see Figure 1.3). In the Contrast condition, the visual display contains: 1) a TARGET object (e.g., a spotted circle) that participants are requested to click on; 2) a COMPETITOR that shares the target property but presents a different shape (e.g., a spotted triangle); 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 circle with no spots); 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 short spiral). The No-Contrast condition was created by substituting the contrasting object with a second distractor. None of the shapes in the visual array shared color.

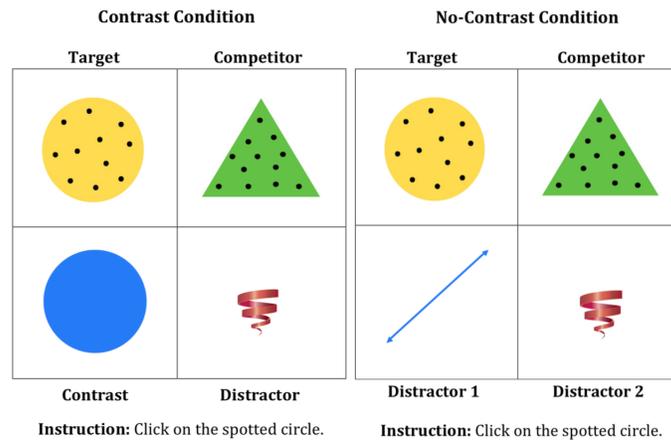


Fig. 1.3 Item example for Experiment 1.

Ten experimental items were constructed.³ Conditions were 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. 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.

As in Aparicio et al. (2015), six different types of fillers (10 trials per type) were constructed (see Figure 1.4). 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 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

³ See supplementary materials to this chapter for a full list of the experimental items used in Experiment 1.

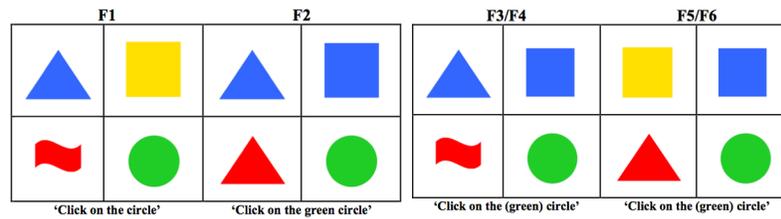


Fig. 1.4 Fillers (Experiment 1).

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.

1.2.2 Materials

1.2.2.1 Visual Stimuli

Pictures used in experimental trials as targets, contrasts and competitors (a total of 29 pictures) were normed in a series of three description-picture matching studies on Mechanical Turk. 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 studies here. In addition, 18 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.

1.2.2.2 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. The mean duration of the adjective for all trials was 503 ms (SD = 76.09). None of the adjectives bore pitch accent or rising tone.

1.2.3 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.

1.2.4 Procedure

Participants saw a visual display with four pictures. Their eye movements were tracked while listening to instructions such as '*Click on the spotted circle*'. 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.

1.2.5 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, $M = 20.57$).

1.2.6 Results

Analyses were performed on two consecutive windows (W1 and W2) of 150 ms starting from the onset of the adjective, such that the right boundary of W2 coincided with the onset of the head noun (set at 703 ms after offsetting the adjective window by 200 ms to adjust for the time required to plan and launch an eye-movement). A third window (W3) of 150 ms starting at the onset of the head noun was also analyzed. W1 and W2 contain fixations reflecting the processing of the adjective, whereas W3 contains fixations reflecting the processing of the head noun. Analyses were run on the aggregated proportion of fixations in each of the three windows (see Figure 1.5). One adjective-noun combination was removed from the data analysis, since the stimuli was found to not appropriately represent the adjectival property.

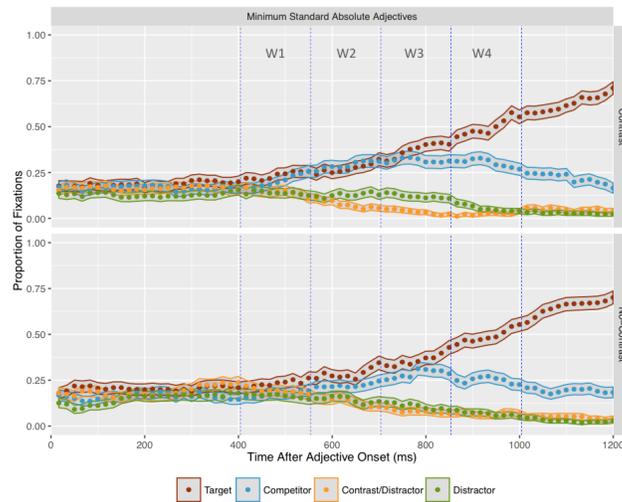


Fig. 1.5 Proportions of fixations to each of the four objects in the display over time starting at the adjective onset for each adjective type. The vertical dashed blue lines mark the boundaries of the four windows defined for data analysis, with the noun onset coinciding with the right boundary of W2 (703 ms from the onset of the adjective).

Figure 1.5 contains the proportions of fixations to each of the four objects in the visual display for each condition. Eye fixations to the target and the competitor objects were analyzed. In order to determine whether target vs. competitor disambiguation occurred faster in the Contrast than in the No-Contrast condition, a two-way ANOVA using OBJECT TYPE (target vs. competitor) and CONDITION (Contrast vs. No-Contrast) as factors was run in each window. Results did not reveal any significant main effect of CONDITION in any of the time windows examined (all $F_s(1, 39) > 0.5$, $p_s > 0.1$). W1 and W2 did not show a significant main effect of OBJECT TYPE ($F_s(1, 39) > 1.88$, $p_s > 0.1$). Even though the main effect of OBJECT TYPE reached significance in W3 ($F(1, 39) = 4.12$, $p < 0.05$), pair comparisons be-

tween target vs. competitor for the Contrast and No-Contrast conditions separately did not yield any significant results ($p_s > 0.1$). No interactions between OBJECT TYPE and CONDITION (all $F_s(1, 39) > 0.01$, $p_s > 0.3$) were observed in any of the three windows. To verify whether there were any RECs in even later time windows, a fourth 150 ms window (W4) spanning from 853-1003 ms was examined. As in W3, a two-way ANOVA showed a main effect of OBJECT TYPE ($F(1, 39) = 31.00$, $p < 0.00001$), but no significant main effect of CONDITION ($F(1, 39) = 1.31$, $p > 0.2$), or OBJECT TYPE x CONDITION interaction ($F(1, 39) = 0.47$, $p > 0.4$) was observed. A one-way ANOVA with OBJECT TYPE as factor revealed a significant difference between the two levels for both the Contrast ($F(1, 39) = 12.59$, $p < 0.002$) and the No-Contrast condition ($F(1, 39) = 26.43$, $p < 0.00001$) such that participants fixated significantly more on the target object than the competitor object.

In addition to the ANOVA analysis reported above, a second analysis using logistic mixed effects models was also performed. The goal of this analysis was to determine whether there were significant differences in the rate at which the proportions of fixations to the target objects in the Contrast and the No-Contrast conditions increased as a function of time. Figure 1.6 plots the proportion of fixations over time to the target objects in the two conditions tested. The existence of a significant difference, such that the target object in the Contrast Condition received a higher proportion of looks earlier than the target object in the No-Contrast condition would be indicative of a REC. A window spanning from the onset of the adjective to the end of W3 (853 ms) was defined for data analysis. The factors CONDITION and TIMEPOINT were included as main effects, with SUBJECTS and ITEMS factored in as random effects.

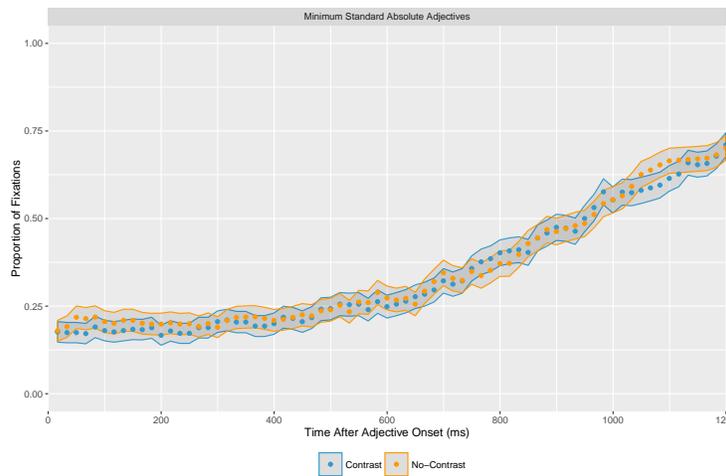


Fig. 1.6 Proportions of fixations over time to the target objects in the Contrast and the No-Contrast condition. The plotted window starts at the adjective onset and spans for 1200 ms.

As in the previous analysis, no significant interaction between CONDITION: TIME-POINT was found ($\beta = -0.0004208$, $p > 0.1$), confirming that MinAAs did not trigger RECs.

1.2.7 Discussion

Our results clearly show that MinAAs do not give rise to RECs, since target vs. competitor disambiguation times did not differ significantly across conditions. The same results were achieved when the proportions of looks to the target objects in the Contrast and the No-Contrast conditions were compared. Therefore, information about the visual context was not used by participants during the adjective window to make predictions about potential referents at a point in which the linguistic instruction was ambiguous given the visual context. Rather participants only relied on the linguistic information available to them to narrow down the set of potential referents in the visual display as the auditory instruction unfolded. The current results contrast with the findings reported by Aparicio et al. (2015), who found RECs for each of the three adjectives tested, i.e. RelAs, ColAs and MaxAAs. Taken together, these two sets of results show that not all prenominal adjectives are equally context-sensitive, even when there is contextual support for a contrastive interpretation.

A important question is whether all the differences in the availability and properties of the observed RECs result from pragmatic reasoning—as modulated by the informativity considerations discussed in §1.1 regarding the use of overspecified prenominal adjectives—or whether RECs are also affected by grammatical factors related to the lexical-semantic properties of each adjective class. Experiment 2 seeks to address this question by quantifying how informative each of these adjective classes are perceived to be when used restrictively vs. redundantly.

1.3 Experiment 2: Perceived Informativity

Experiment 2 addresses the question of whether all the adjective types tested by Aparicio et al. (2015) and the current eye-tracking experiment (see §1.2) are perceived as equally informative when the display contains a contrastive object (Contrast condition), compared to displays that do not (No-Contrast condition). With this goal in mind, Experiment 2 consisted of an offline judgement task, where participants were instructed to rate whether the instructions used in the eye-tracking experiments provide a sufficient amount of information to confidently identify the target object in the relevant visual display.

If the online eye-tracking effects reported by Aparicio et al. (2015), as well as the results reported above for Experiment 1, are shaped by differences in the perceived informativity, we predict the following patterns of results for Experiment 2: First, since MinAAs are the only type of adjective that do not give rise to RECs, we don't

expect to find any differences in perceived informativity between the Contrast and the No-Contrast conditions. All other adjectives should show a significant difference between these two conditions such that the No-Contrast condition is perceived as more overinformative than the Contrast condition. Second, based on the timing of the RECs observed for each adjective type, we would expect that the magnitude of the overspecification penalty should be greater for MaxAAs than for ColAs and RelAs.

The same lists and adjectives (RelAs=9, MaxAAs=4, MinAAs=6, ColAs=4) used in the eye-tracking studies were tested with a total of 60 experimental items (20 containing RelAs, 10 containing MaxAAs, 10 containing MinAAs and 20 containing ColAs). Conditions were 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 (see Figure 1.7). The same 60 filler trials used in Experiment 1 were included (see §1.2).

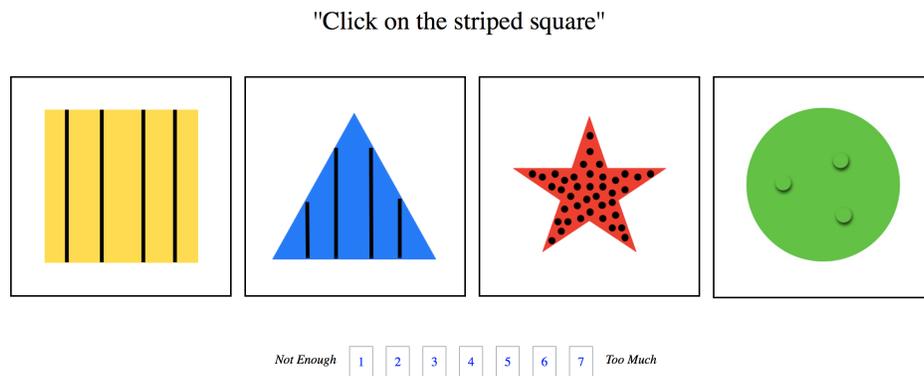


Fig. 1.7 Item example for Experiment 2.

1.3.1 Methods

1.3.1.1 Materials

Stimuli consisted of the same visual displays used by Aparicio et al. (2015), a total of 100, plus the 20 visual displays tested in the eye-tracking experiment reported in §1.2. The auditory instructions used in both eye-tracking experiments were transcribed and accompanied the visual displays.⁴

⁴ See supplementary materials to this chapter for a full list of the experimental items used in Experiment 2.

1.3.1.2 Procedure

Participants saw displays of four pictures on a computer screen coupled with a written statement such as ‘*Click on the striped square*’. For each of the displays, participants were instructed to rate whether the instruction provided a sufficient amount of information to identify the right target. Judgments were indicated on a 1-7 scale, where 1 corresponded to *Not enough information* and 7 corresponded to *Too much information*. At the beginning of the experiment, participants had three practice trials to help them become familiar with the task.

1.3.1.3 Participants

Participants were 32 native speakers of English between the ages of 18-35 (12 females; mean age = 30) recruited through the website Amazon Mechanical Turk. Three subjects were removed from data analysis because they were not between 18-35 leaving a total of 29 (10 females; mean age = 29). All participants were paid \$3.

1.3.2 Results

Means were obtained for all adjective types. Visual inspection of the left plot in Figure 1.8 reveals that the No-Contrast condition received higher ratings compared to the Contrast condition for ColAs, RelAs and absolute adjectives (AAs). For the class of AAs, data from MaxAAs and MinAAs were combined. The ratings in the Contrast condition were used as the baseline comparison against the ratings in the No-Contrast condition, as the former represents ratings pertaining to the condition containing the optimal amount of information, since target identification would not be possible in the absence of the adjective. Paired t-tests confirm that the differences between the two conditions were statistically significant (ColAs: $t(28) = -5.78, p < 0.0001$; RelAs: $t(28) = -3.20, p < 0.01$; AAs: adjectives $t(28) = -3.85, p < 0.001$). However, closer inspection to the two subclasses of AAs (central plot, Figure 1.8) shows that the difference between conditions observed for AAs is mostly driven by MaxAAs, which present the higher ratings in the No-Contrast condition. A paired t-test confirmed that this difference was highly significant ($t(28) = -5.89, p < 0.0001$). MinAAs, on the other hand, showed a non-significant difference across conditions ($t(28) = -0.91, p > 0.3$).

A 2-way ANOVA using ADJECTIVE TYPE and CONDITION as factors was run on the three classes of adjectives that showed significant differences between the two conditions, i.e. ColAs, RelAs and MaxAAs. A significant interaction for ADJECTIVE TYPE x CONDITION was detected ($F(2, 56) = 7.64, p < 0.008$), showing that the magnitude of the effect was different across the three adjective types. In order to further explore this interaction, a 2-way ANOVA was run in three different

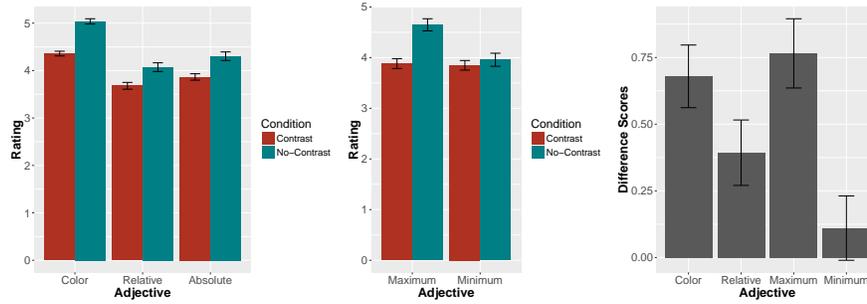


Fig. 1.8 **Left.** Rating means for color, relative and absolute adjectives; **Central.** Rating means for maximum and minimum standard absolute adjectives; **Right.** Difference Scores between the Contrast and the No-Contrast condition for each adjective type.

subsets of the data. The interaction remained significant for the subset containing RelAs and MaxAAs ($F(1, 28) = 10.70, p < 0.002$), and the subset containing RelAs and ColAs ($F(1, 28) = 13.10, p < 0.001$), while it did not reach significance for the data subset containing only ColAs and MaxAAs ($F(1, 28) = 0.7, p > 0.4$). This suggests that the magnitude of the effect was comparable for ColAs and MaxAAs (see right panel of Figure 1.8 containing the difference scores obtained by subtracting the Contrast condition from the No-Contrast condition for each adjective type), and that the ADJECTIVE TYPE \times CONDITION interaction detected for the full data set was driven by differences between ColAs and MaxAAs on the one hand and RelAs on the other.

1.3.3 Discussion

For ColAs, RelAs and MaxAAs, the No-Contrast condition received significantly higher ratings than the Contrast condition. This means that participants perceived a difference between the optimally informative baseline in the Contrast condition and the No-Contrast condition, which they judged to contain more information than necessary. Interestingly, no parallel effect was found for MinAAs, suggesting that participants did not perceive differences between the degree of informativity of the two conditions tested. Our results also revealed that the magnitude of the effect of perceived informativity was not the same for ColAs, RelAs and MaxAAs. The results from the 2-way ANOVA interaction and the t-tests indicate that the effect was bigger for ColAs and MaxAAs than it was for RelAs, while no significant difference in perceived informativity was found between ColAs and MaxAAs. The main conclusion that can be extracted from these results is that perceived informativity is indeed modulated by adjective class. In the general discussion (§1.4), we address the relationship between perceived informativity and RECs.

1.4 General Discussion

Out of the four adjective classes tested in Experiment 1 and in Aparicio et al.'s (2015) study, we were able to detect RECs for ColAs, RelAs and MaxAAs. However, MinAAs failed to display a REC, as target vs. competitor disambiguation took place in the same time window, i.e. W4, for both the Contrast and the No-Contrast condition (see Figure 1.9). An important finding of Aparicio et al.'s (2015) is that there exist non-trivial timing differences in the RECs of ColAs and RelAs on the one hand, and MaxAAs on the other. For the former, the effect took place in W2, during the adjective window, whereas for the latter the effect did not occur until W3, a window that already reflects processing of the head noun. In the case of ColAs and RelAs, participants committed to the target object at a point in which the linguistic input was still ambiguous between two objects in the visual display (i.e., target and competitor), whereas for MaxAAs, target identification was facilitated in the Contrast condition, but was nevertheless significantly delayed, as participants did not discriminate between target and competitor until information about the head-noun was available to them.

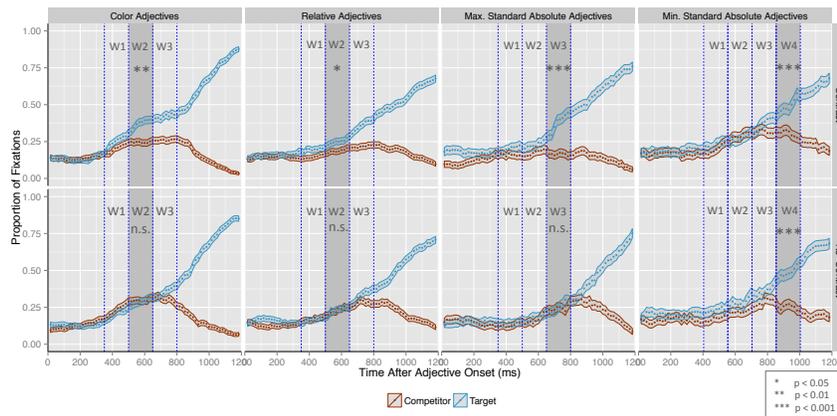


Fig. 1.9 Proportions of fixations to target vs. competitor over time starting at the adjective onset. Data belonging to ColAs, RelAs and MaxAAs are reproduced from Aparicio et al. (2015). All windows are 150 ms long. For each adjective, the right boundary of W2 coincides with the onset of the head noun. The grayed time windows correspond to the first window in which a significant difference was found.

Experiment 2 also revealed important asymmetries in the effect of perceived informativity across adjective types. MinAAs were the only class of adjectives that did not display differences in perceived informativity between the Contrast and the No-

Contrast condition. Interestingly, MinAAs were also the only adjective class that did not give rise to RECs. However, ColAs, RelAs and MaxAAs did show an overspecification penalty, as indicated by the significantly higher ratings obtained for these three adjective classes in the No-Contrast condition, which was not compatible with a contrastive interpretation of the adjective.

Taken together, the previous results reported by Aparicio et al. (2015), as well as the results from Experiment 1 and 2 suggest that informativity is an important factor in RECs, as shown by the relation between RECs and the offline measure of perceived informativity: adjectives that showed an overspecification penalty (ColAs, RelAs and MaxAAs) also gave rise to RECs, whereas adjectives that did not show an overspecification penalty (MinAAs) did not display RECs. However, the timing differences observed in the RECs of ColAs, RelAs and MaxAAs could not be uniquely attributed to the overspecification penalties detected by Experiment 2 for these three types of adjectives. As discussed above, the magnitude of the perceived (over)informativeness was different across the three adjective types with RelAs showing a significantly smaller effect compared to ColAs and MaxAAs, for which the size of the effect was comparable. If perceived informativity was the only source of RECs we would expect ColAs and MaxAAs to pattern alike with respect to the timing of their RECs, showing earlier effects compared to RelAs. However, this is not what Aparicio et al.'s (2015) results show, with MaxAAs being delayed with respect to ColAs and RelAs. We therefore conclude, that informativity cannot be the only factor driving RECs.

Based on these results, we would like to suggest that there exist at least two non-mutually exclusive sources of the RECs. The first one pertains to perceived informativity considerations related to quantity and manner-based pragmatic reasoning about referential contrast triggered by the mention of the prenominal adjective. Second, RECs are also modulated by differences in lexical processing incurred by distinct lexically encoded types of context-dependence. The differences in the timing of the REC of RelAs and MaxAAs can be explained in this way. While relative adjectives like *tall* resort to context in order to fix the value of their *semantic* threshold (typically computed with respect to a contextually salient comparison class), MaxAAs like *empty* have been argued to only interact with context in order to fix a *pragmatic* threshold of imprecision (Kennedy 2007; Syrett et al. 2009; van Rooij 2011; Burnett 2014; Qing and Franke 2014; Leffel et al. to appear). If lexical context-sensitivity is an important component of the timing resolution of RECs, it is conceivable that RelAs could trigger RECs with a different time course from MaxAAs. But the exact mechanism that relates context-sensitivity to the time course of RECs still remains a question for future research. Another question that remains to be explored is whether the early REC attested for ColAs also results from facilitated lexical processing (though see Aparicio et al. (2015) for an argument against this view). In principle, the adjectival threshold of ColAs is not assumed to depend on a contextually salient comparison class for its resolution (Kennedy and McNally 2005). This may mean that other high level perceptual factors such as the visual saliency of color might underlie the timing resolution of the REC for ColAs.

Given the abundance of results showing that speakers have a greater tendency to use ColAs redundantly than any other class of adjectives (see Pechmann 1989; Belke and Meyer 2002; Nadig and Sedivy 2002; Sedivy 2004, among many others), it is somehow unexpected that Experiment 2 showed such a clear penalty for overspecified uses of ColAs. If hearers are sensitive to the probabilities of use of overspecified adjectives, ColAs would be expected to give rise to the lowest overspecification penalty among all the adjectives tested in Experiment 2. It is possible that the nature of the stimuli used in our experiment had an effect on how overinformative ColAs were perceived to be. In a production experiment, Rubio-Fernández (2015) shows that the rates of overspecification of ColAs vary depending on the nature of the object. Rubio-Fernández found lower rates of color overspecification with geometric shapes in polychrome displays than in displays containing garments, a type of object for which color is a more central feature. A final important issue is the question of why MinAAs did not show differences in perceived informativity in the two conditions tested. At this point, we do not have an explanation for the lack of sensitivity to the visual context displayed by this adjective class. Further research will have to determine why this class of adjective does not seem to be associated with an expectation of contrastive use.

1.5 Conclusion

The experiments presented in this paper had the goal of determining whether informativity-based reasoning about the use of a prenominal modifier is the sole driver of Referential Effects of Contrast involving adjectivally modified NPs. By examining four different classes of adjectives, we have shown that perceiving the use of a particular class of adjective as overinformative when used redundantly is related to whether such adjective class should give rise to a REC. However, while pragmatic reasoning is an important source of these effects, it cannot alone account for the variety of attested patterns of RECs. We conclude that lexical semantic factors determining how context-sensitive a given adjective class is further contributes to the temporal resolution of such effects.

Acknowledgements: We would like to thank Aaron Hill, Jackson Lee, Gabriel Aparicio and Katie Franich for their assistance during various stages of the eye-tracking experiment reported in this paper, as well as Michelle Namkoong and Sonia Juan Rubio for help in the creation of part of the visual stimuli. Finally, we would also like to thank the audiences at the workshop on *Gradability, Scale Structure, and Vagueness: Experimental Perspectives*, and the 22nd AMLaP Conference. This project was supported by a NSF grant (BCS 1227144) to C. Kennedy and M. Xiang.

References

- Altmann, G. and Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3):191–238.
- Aparicio, H., Xiang, M., and Kennedy, C. (2015). Processing gradable adjectives in context: a Visual World study. In *Proceedings of Semantics and Linguistic Theory 25*.
- Arts, A., Maes, A., Noorman, K., and Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, 43:361–374.
- Belke, E. and Meyer, A. S. (2002). Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during “same”–“different” decisions. *European Journal of Cognitive Psychology*, 14(2):237–266.
- Burnett, H. (2014). A delineation solution to the puzzles of absolute adjectives. *Linguistics and Philosophy*, 37(1):1–39.
- Clapp, L. (2012). Indexical color predicates: Truth conditional semantics vs. truth conditional pragmatics. *Canadian Journal of Philosophy*, 42(2):71–100.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6:84–107.
- Crain, S. and Steedman, M. (1985). On not being led up the garden path: the use of context by the psychological parser. In Dowty, D., Karttunen, L., and Zwicky, A., editors, *Natural Language Parsing*. Cambridge University Press, Cambridge, MA.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., and Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24(6):409–436.
- Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J., editors, *Syntax and Semantics Volume 3: Speech Acts*. Academic Press, New York.
- Grodner, D. and Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In Pearlmuter, N. and Gibson, E., editors, *The processing and acquisition of reference*, pages 239–272. MIT Press, Cambridge, MA.
- Hanna, J. E. and Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, 28:105–115.
- Hanna, J. E., Tanenhaus, M. K., and Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49:43–61.
- Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45.
- Kennedy, C. and McNally, L. (2005). Scale structure and the semantic typology of gradable predicates. *Language*, 81(2):345–381.
- Kennedy, C. and McNally, L. (2010). Color, context, and compositionality. *Synthese*, 174(1):79–98.

- Koolen, R., Gatt, A., Goudbeek, M., and E., K. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, 43:3231–3250.
- Koolen, R., Goudbeek, M., and E., K. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, 37:395–411.
- Lassiter, D. and Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In Snider, T., editor, *Proceedings of Semantics and Linguistic Theory 23*, pages 587–610, Ithaca, NY. CLC.
- Lassiter, D. and Goodman, N. D. (2015). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, pages 1–36.
- Leffel, T., Xiang, M., and Kennedy, C. (to appear). Imprecision is pragmatic: Evidence from referential processing. In *Proceedings of Semantics and Linguistic Theory 26*.
- Maes, A., Arts, A., and Noordman, L. (2004). Reference management in instructive discourse. *Discourse Process*, 37:117–144.
- McNally, L. (2011). The relative role of property type and scale structure in explaining the behavior of gradable adjectives. In Nouwen, R., van Rooij, R., Sauerland, U., and Schmitz, H.-C., editors, *Vagueness in Communication*, volume 6517 of *Lecture Notes in Computer Science*, pages 151–168. Springer, Berlin Heidelberg.
- Nadig, A. S. and Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children’s on-line reference resolution. *Psychological Science*, 13(4):329–336.
- Paraboni, I., van Deemter, K., and Masthoff, J. (2007). Generating referring expressions: Making referents easy to identify. *Computational Linguistics*, 33:229–254.
- Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, 27(1):89–110.
- Pinkal, M. (1995). *Logic and lexicon: the semantics of the indefinite*. Kluwer, Dordrecht.
- Pykkönen-Klauck, P. and Crocker, M. W. (2016). Attention and eye movement metrics in visual world eye tracking. In Knoeferle, P., Pykkönen-Klauck, P., and Crocker, M. W., editors, *Visually situated language comprehension*, pages 67–82. John Benjamins.
- Qing, C. and Franke, M. (2014). Gradable adjectives, vagueness, and optimal language use: a speaker-oriented model. In Snider, T., D’Antonio, S., and Weigand, M., editors, *Proceedings of Semantics and Linguistic Theory 24*, pages 23–41, Ithaca, NY. CLC.
- Rothschild, D. and Segal, G. (2009). Indexical predicates. *Mind and Language*, 24(4):467–493.
- Rotstein, C. and Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics*, 12(3):259–288.
- Rubio-Fernández, P. (2015). Redundancy is efficient –and effective, too. In *Paper presented at the XI Conference on Architectures and Mechanisms for Language Processing (AMLaP)*.
- Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? an efficiency-based analysis of color overspecification. *Frontiers in Psychology*, 7(153).

- Sassoon, G. and Toledo, A. (2011). Absolute and relative adjectives and their comparison classes. Unpublished manuscript.
- Sedivy, J. 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, J. C. (2004). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In Trueswell, J. C. and Tanenhaus, M. K., editors, *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*, pages 345–364. MIT Press, Cambridge, MA.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., and Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2):109–147.
- Syrett, K., Kennedy, C., and Lidz, J. (2009). Meaning and context in children’s understanding of gradable adjectives. *Journal of Semantics*, 27(1):1–35.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, 268(5217):1632–1634.
- Unger, P. (1975). *Ignorance: a case for scepticism*. Clarendon Press, Oxford.
- van Rooij, R. (2011). Vagueness and linguistics. In Ronzitti, G., editor, *Vagueness: A Guide*, chapter 6, pages 123–179. Springer, Dordrecht.
- Weber, A., Braun, B., and Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, 49(3):367–392.
- Westerbeek, H., Koolen, R., and Maes, A. (2015). Stored object knowledge and the production of referring expressions: the case of color typicality. *Frontiers in Psychology*, 6(935).
- Wolter, L., Skovbroten Gorman, K., and Tanenhaus, M. K. (2011). Scalar reference, contrast and discourse: Separating effects of linguistic discourse from availability of the referent. *Journal of Memory and Language*, 65:299–317.