Proceedings of SALT 32: 481–497, 2022

Incremental pragmatic interpretation of gradable adjectives: The role of standards of comparison*

Stavroula Alexandropoulou
Osnabrück University, University of Potsdam

Marisha Herb
University of Potsdam

Henrik Discher
Humboldt University of Berlin

Nicole Gotzner
Osnabrück University, University of Potsdam

Abstract While for relative gradable adjectives the value on the underlying measurement scale that serves as a standard of comparison is contextually determined, for absolute gradable adjectives this is typically taken to be a fixed, context-invariant value (Rotstein & Winter 2004; Kennedy & McNally 2005). The present study investigates how lexical-semantic factors, such as the type of standard of comparison invoked by gradable adjectives, affect the incremental computation of scalar implicatures triggered by such adjectives. Our study shows that the incremental computation of scalar implicatures is facilitated by the immediate visual context but only for relative adjectives. Minimum standard absolute adjectives, which impose a lower bound on their corresponding measurement scales, robustly trigger upper-bounded interpretations independently of the availability of contrastive visual information. Our findings indicate that different kinds of scalar meaning are computed incrementally and potentially in parallel. Overall, these findings shed new light on theories of scalar implicatures and highlight the need for a model of adjective meaning that incorporates semantic and pragmatic factors (see also Gotzner 2021; and for related ideas in the domain of quantifiers see Franke & Bergen 2020 and Cremers, Wilcox & Spector 2022, and Magri (2017) for Hirschberg scales).

Keywords: gradable adjectives, comparison standards, scalar implicature, scalar diversity

* We thank Helena Aparicio, Jakub Dotlačil, Louise McNally, Rick Nouwen, Ira Noveck, Chao Sun, Bob van Tiel, and the audiences of HSP 2022, AMGL42, SALT 2022, and XPRAG 2022 for valuable comments on this work. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the Emmy Noether project awarded to Nicole Gotzner (Grant Nr. GO 3378/1-1).

1 Introduction

The meaning of gradable adjectives is well understood in semantics and typically modeled via so-called measurement scales comprising of sets of totally ordered degrees (Cresswell 1976; Kennedy & McNally 2005; Kennedy 2007; see Solt 2015 for an overview). At the same time, considerably less is known about how pragmatic interpretations involving pairs of gradable adjectives (e.g., <*warm, hot*>) are derived (but see Doran, Ward, McNabb, Larson & Baker 2012; van Tiel, van Miltenburg, Zevakhina & Geurts 2016; Gotzner, Solt & Benz 2018b; Leffel, Cremers, Gotzner & Romoli 2019). The standard assumption is that scalar implicatures triggered by weak adjectival scale-mates are computed via the negation of stronger alternatives just as the scalar implicatures triggered by quantifiers and other Horn scale-mates (so-called uniformity assumption, see especially van Tiel et al. 2016). Yet, recent research has suggested that the rich semantic structure of adjectives may be crucial to how their scalar inferences are derived (see especially Gotzner, Solt & Benz 2018a; Gotzner et al. 2018b; Leffel et al. 2019).

Semantic theories of adjective meaning distinguish different types of adjectives based on the way context affects their interpretation. While the meaning of relative gradable adjectives is computed on the basis of the comparison class instantiated in a given context (e.g., *warm*, whose standard varies for different domains), absolute gradable adjectives have fixed, context-invariant standards (e.g., *breezy* denotes a non-zero degree of wind).[1] Previous work has found that such properties of measurement scales taken to underly the meaning of gradable adjectives affect the inferences derived from pairs of adjectival Horn scale-mates differing in informational strength such as *warm* and *hot* (e.g., Gotzner et al. 2018b).

The present study addresses two main research questions: (i) How are standards determined during incremental interpretation and (ii) how do the different semantic properties of relative and absolute adjectives interact with pragmatic processing? We focus on the role of context (as instantiated by a Contrast object) in the computation of a potential scalar/upper-bounded inference for relative (*warm* ⇝ 'warm but not hot') and minimum standard absolute adjectives (*breezy* ⇝ 'breezy but not windy').

We build on the visual world eye-tracking studies by Aparicio, Xiang & Kennedy (2016); Aparicio, Kennedy & Xiang (2018), who found that the processing of relative adjectives hinges on the visual presence of an object that helps set up the relevant comparison class (so-called referential contrast effect, building on seminal work by Sedivy, Tanenhaus, Chambers & Carlson 1999) while the processing of minimum-standard absolute adjectives relies solely on linguistic information. If the differential role of context (comparison classes vs. absolute standards) also factors

1 It is an open debate whether these distinctions are semantic in nature or ultimately an effect of the priors that differ for different types of adjectives (see Lassiter & Goodman 2014).

into the computation of scalar implicatures, we expect to find differential referential contrast effects for relative and minimum-standard adjectives during incremental interpretation. We show by means of an incremental decision task that this prediction is borne out. Overall, our findings have repercussions for theories of implicatures and the role of comparison classes and standards in adjective interpretation.

The present paper is structured as follows: Section 2 presents the theoretical and experimental background relevant for our experimental study and Section 3 details our experiment and the obtained results. Section 4 discusses our main finding and its implications, and Section 5 concludes.

2 Theoretical and experimental background

2.1 Relative/absolute distinction: Different standards, different inference rates

The two types of gradable adjectives we will be focusing on in the present study are relative and absolute adjectives, and specifically the minimum-standard absolute adjectives for the latter case. Relative and absolute gradable adjectives differ in the type of standard they involve: For relative adjectives (e.g., *long, warm*), the standard of comparison or threshold of application is a context-dependent value on the underlying measurement scale and is determined on the basis of a given comparison class. For absolute adjectives, on the other hand, the standard of comparison is typically the endpoint of the measurement scale they encode, which makes it a fixed and context-invariant value (Kamp & Rossdeutscher 1994; Rotstein & Winter 2004; Kennedy & McNally 2005; Kennedy 2007). More specifically, minimum-standard absolute adjectives like *bent, breezy* require that an individual possess a minimal, non-zero degree of the measured property (e.g., a non-zero degree of bend) in order to qualify as such (*bent*), that is, the relevant standard value is the minimal value, the endpoint of the corresponding scale. Figure 1 illustrates the different types of standards invoked by the two adjective classes under consideration, which will be crucial in our study. The green line on the left pointed to by the arrow (target line) counts as long given the contextual lines (comparison class) in the first scenario, which is not the case in the second scenario below, where the contextual lines present a greater degree of length compared to that of the target line. On the contrary, the target line on the right counts as bent in either scenario and independently of the contextual lines, as it meets the requirement of presenting a non-zero degree of bend. The differential context-dependence in setting the standard of comparison has been attested experimentally (for example, Schmidt, Goodman, Barner & Tenenbaum 2009; Solt & Gotzner 2012; Lassiter & Goodman 2014; Qing & Franke 2014).

Properties of measurement scales, such as the type of standard value, have recently been shown to affect the availability of different pragmatic inferences of

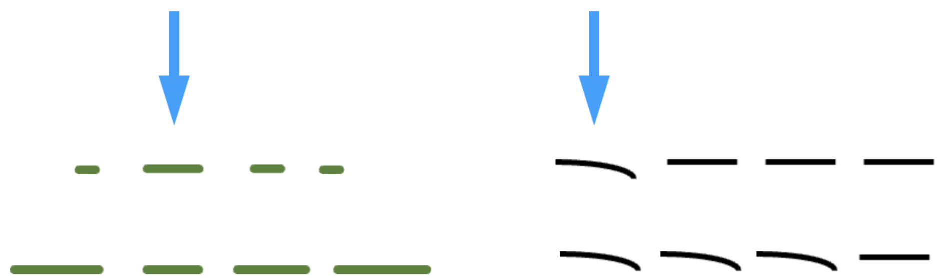


Figure 1 An illustration of the context-dependent vs. context-invariant comparison standards of the relative adjective *long* and of the minimum-standard absolute adjective *bent*, respectively.

gradable adjectives participating in Horn lexical scales (Gotzner et al. 2018b; see also Leffel et al. 2019). Gotzner et al. find lower rates of scalar implicatures for relative adjective scales (<*large, gigantic*>) than for absolute adjective scales and especially for scales where the stronger scale-mate is endpoint-denoting (e.g., <*possible, certain*>; see also van Tiel et al. 2016). Their explanation for scalar implicatures of relative adjectives being not computable in all contexts is that participants are uncertain about where the threshold of application of the weak scale-mate is as well as that of the stronger scale-mate (*large* vs. *gigantic*), making it unlikely to draw a scalar implicature, which negates the applicability of the stronger scale-mate. On the other hand, scales whose stronger scale-mate denotes an endpoint are more robust triggers of scalar implicatures, because the adjectives referring to endpoints are probably "particularly salient alternatives for the purposes of scalar implicature calculation" (Gotzner et al. 2018b: 10).

2.2 Incremental processing of adjectives

Sedivy et al. (1999) investigated the incremental interpretation of relative adjectives like *tall* in a visual world eye-tracking study using the referential task, namely, consisting of an auditory instruction sentence (e.g., *Pick up the tall glass*) and a visual display with 4 potential referents of the sentence (Target, Competitor, Distractor 1, Contrast/Distractor 2 objects). The instruction was temporarily ambiguous between the Target object, which satisfied both the adjective and the noun information of the instruction, and the Competitor object (e.g., a tall pitcher) that only satisfied the adjective information. Their crucial manipulation was the presence or absence of the so-called Contrast object (a short glass), which satisfied the noun information but not the adjective information. Sedivy et al. showed that listeners interpret sentences containing adjectives like *tall* incrementally, taking rapidly into account the contextually-defined contrast. Their main finding was that, given the temporarily ambiguous instruction, listeners were faster at converging on the Target (tall glass) when presented with a Contrast object (short glass) in the immediate visual context than without such an object. This is known as the referential contrast effect (RCE), which was crucially observed before the noun information was presented auditorily. The RCE has been given a pragmatic explanation, according to which speakers only utter an adjective when this helps the listener choose between two objects that belong to the same category, e.g., two glasses (a tall and a short one).

Building on Sedivy et al.'s (1999) study, Aparicio et al. (2016, 2018) investigated to what extent the RCE may be driven by aspects of the lexical semantics (e.g., comparison standard) of the adjective used in the instruction sentence. They replicated Sedivy et al.'s (1999) experiment adding the type of adjective (relative vs. minimum-standard adjective) as a second manipulation.[2] Aparicio et al. found for relative adjectives (*tall*) the RCE attested by Sedivy et al. (but see Leffel, Xiang & Kennedy 2016; Qing, Lassiter & Degen 2018), while they did not find such an effect for minimum-standard adjectives (e.g., *spotted*). When processing minimum standard adjectives, participants used only linguistic information in order to identify the Target referent (Aparicio et al. 2018). Aparicio et al. suggest that this difference may result from the different lexical semantic properties of the adjective classes they tested: Relative adjectives (e.g., *tall*) need contextual support in order to fix the value of their threshold on the basis of a contextually salient comparison class. In particular, it seems that the Contrast object forms a comparison class with the Target object, on the basis of which the Target is classified as *tall* (see also Leffel et al. 2016). On the contrary, this is not needed for minimum standard adjectives that have a fixed, context-invariant threshold, i.e., the minimum, endpoint value on the relevant measurement scale. Hence, part of the story of RCEs, characteristic of the incremental interpretation of adjectives in a referential task, should also take into consideration the lexical semantics of the adjectives (semantic account of RCEs).

3 Current study

3.1 Research questions and design

The present study aims to investigate how scalar meaning is resolved during incremental interpretation. In particular, we focus on whether context affects the computation of a potential scalar implicature/upper-bounded interpretation for relative (*warm* $\rightsquigarrow$ 'warm but not hot') and minimum standard adjectives (*breezy* $\rightsquigarrow$ 'breezy but not windy') during incremental processing.

2 They also tested maximum-standard absolute adjectives (e.g., *straight*) and color adjectives (e.g., *yellow*), whose results will not be reported in the present paper, as they are not directly relevant.

There are two key assumptions that enter into the hypotheses for the current study. First, lexical-semantic factors such as the type of standard invoked by gradable adjectives (context-dependent vs. context-invariant) affect the availability of upper-bounded interpretations of such adjectives participating in Horn scales: unlike absolute adjectives, relative adjectives like *warm* do not trigger scalar implicatures (SIs) in all contexts, because there is uncertainty as to where the threshold value and scalar boundary for the stronger scale-mate lies (based on Gotzner et al. 2018b). Second, context affects the incremental processing of gradable adjectives differentially: the processing of relative adjectives is facilitated by the context (and the visual presence of a Contrast object in particular), whereas that of minimum standard adjectives relies solely on linguistic information (based on Aparicio et al. 2016, 2018). Hypothesizing that these differences with respect to the effect of context on adjective interpretation also factor in the incremental computation of upper-bounded interpretations, we predict that the context will affect the SI computation differentially for different types of adjectives. We operationalize context in terms of presence of a Contrast object on a visual display as in Sedivy et al.'s (1999) and Aparicio et al.; Aparicio et al.'s (2016; 2018) studies.

Our experiment manipulated two factors, with two levels each: (i) Adjective Type: minimum-standard absolute adjectives vs. relative adjectives, and (ii) Contrast: presence of a Contrast object on the visual display (contrast condition) vs. no presence of a Contrast object on the visual display (no-contrast condition). So our experiment had a 2×2 design.

The specific prediction we set out to test given this design is that there will be differential RCEs for relative and minimum-standard adjectives during incremental interpretation. This should be reflected in a significant interaction of the two factors, i.e., Adjective Type*Contrast. Our hypothesis and prediction as well as the participant exclusion criterion, the experimental procedure, and statistical analysis of our data that follow were all pre-registered and are available on the Open Science Framework.[3]

3.2 Methods

3.2.1 Participants

We recruited 120 participants on the Prolific platform.[4] The data of one participant were not registered nor did the participant fill in the completion code of the experiment. The remaining number of participants was 119 (81 female, 33 male, 4 diverse, and 1 preferring not to say, mean age: 30.41, age range: 18-75). Participants

3 This is the link to our pre-registration: https://doi.org/10.17605/OSF.IO/E82JD.
4 https://www.prolific.co

were US residents and were also screened for their native language. They were only included in the analysis if their self-reported first native language was English. No participants were excluded on the basis of this criterion. The experiment lasted about 4 minutes and all participants were paid 0.57 USD in compensation.

3.2.2 Materials

The experimental materials consisted of the following adjectival scales:

(1) Minimum-standard adjective scales: <*sleepy*, *asleep*>, <*breezy*, *windy*>, <*drizzly*, *rainy*>

(2) Relative adjective scales: <*warm*, *hot*>, <*chubby*, *fat*>, <*content*, *happy*>

We borrowed these lexical scales from van Tiel & Pankratz (2021). We also borrowed and adapted from van Tiel & Pankratz (2021) the images that accompanied these scales in their Control-True, Control-False, and Target conditions. All these materials were pretested by van Tiel & Pankratz in two experiments.

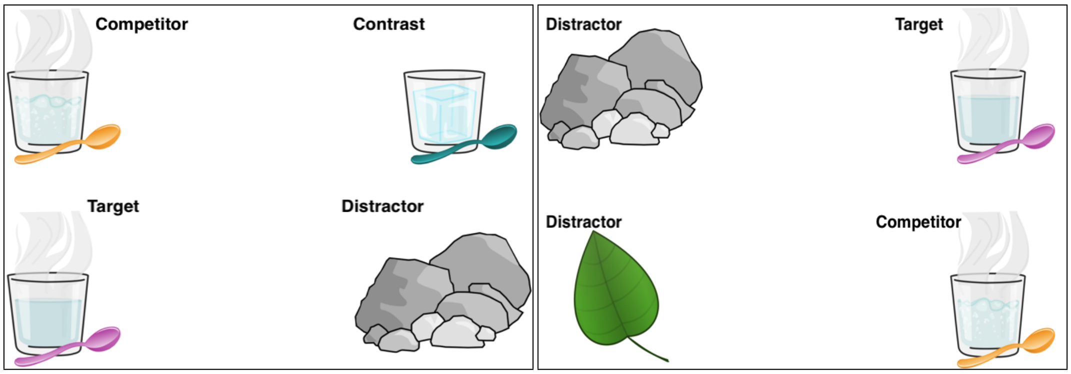
The classification in (1) and (2) was made on the basis of the following two diagnostic tests (see Kennedy & McNally 2005; Kennedy 2007) given COCA co-occurrences and English native speaker's informal judgements: (i) the licit use of a given adjective with adverbs like *very* and *extremely* would indicate a relative gradable adjective, (ii) the licit use of a given adjective with modifiers like *slightly, a bit, a little* indicates a minimum-standard gradable adjective.[5]

Every trial consisted of a visual display of 4 pictures of objects and an instruction sentence appearing incrementally over the visual display. The instruction sentence contained the weak scale-mate of the scales in (1)/(2) (critical adjective) and instructed participants to click on the picture that matches a certain description, e.g., *Click on the picture of the breezy weather with the yellow flag*. For each such sentence, three images were constructed, adapting those from van Tiel & Pankratz (2021): An image that made the instruction sentence up to the noun modified by the critical adjective (e.g., *Click on the picture of the breezy weather*) literally true but that made the scalar inference associated with the adjective ('breezy but not windy') false (Competitor image), an image that made the whole instruction sentence unambiguously true (Target image), and an image that made the instruction sentence up to the critical adjective unambiguously false (Contrast image).[6] Note that the

5 We also took extra care to identify and exclude from consideration special uses of an adjective with a fixed, absolute standard that would allow for modification by *very*. Such uses are special because typical absolute adjectives are coerced into an atypical relative-like interpretation (see Kennedy & McNally 2005).

6 The Competitor image corresponded to van Tiel & Pankratz's (2021) Target condition, the Target image corresponded to their Control-True condition, and the Contrast object corresponded to their Control-False condition.

Instruction: [**Click on the picture of the**]$_{base}$ [**warm**]$_{adj}$ [**water**]$_{noun}$ [**with the purple spoon.**]$_{disamb}$



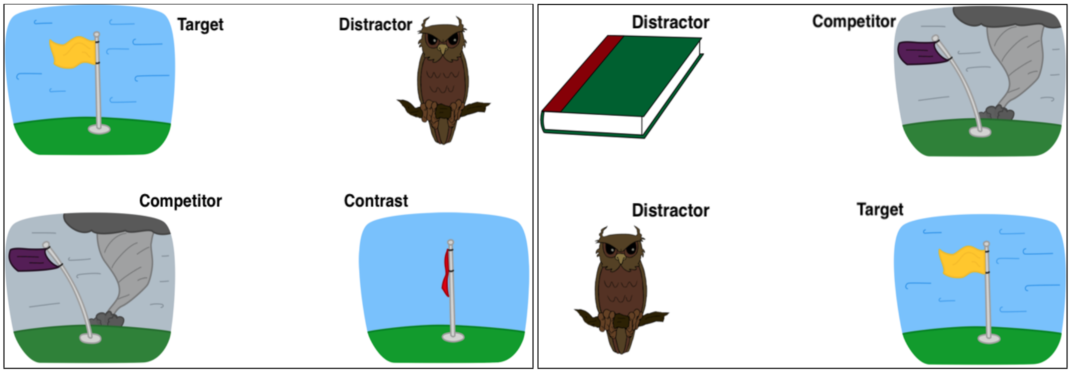
(a) Contrast condition. (b) No-contrast condition.

Figure 2 Example item of the relative adjective scale <*warm*, *hot*>. Bracketing in the instruction sentence represents the segments of the instruction revealed incrementally.

Contrast object (e.g., windless weather) satisfied the noun information (weather) but not the critical adjective information (breezy). The fourth object of the display was always a distractor, namely, an object that satisfied neither the critical adjective information nor the noun information (cf. an owl). The displays in the no-contrast trials had a second distractor in place of the Contrast object. Figures 2 and 3 show an example trial of a relative and a minimum-standard adjective, respectively, in the contrast and the no-contrast conditions.

The Competitor object always instantiates a higher degree of the property encoded by the critical adjective, representing the stronger scale-mate of a given adjective scale (pretested by van Tiel & Pankratz 2021). The instruction sentences differed from van Tiel & Pankratz's test sentences in that the critical adjective is in attributive prenominal position as opposed to predicate position (e.g., *The weather is breezy*) as well as in the addition of the lexical material *Click on the picture of the* at the beginning and of the *with*-phrase at the end of the sentence. The *with*-phrase of the instruction always named a color attribute and a noun and disambiguated the sentence, forcing the selection of the Target referent. The deviation of the present experimental design from that of Sedivy et al.'s (1999) and Aparicio et al.; Aparicio et al.'s (2016; 2018) is that our Competitor object belongs to the same category as the Target object (e.g., they are both pictures of weather), making the instruction sentence ambiguous for a larger time window compared to Sedivy et al.'s (1999) and Aparicio et al.; Aparicio et al.'s (2016; 2018) studies. That is to say, the instruction

Instruction: [**Click on the picture of the**]$_{base}$ [**breezy**]$_{adj}$ [**weather**]$_{noun}$ [**with the yellow flag.**]$_{disamb}$



(a) Contrast condition. (b) No-contrast condition.

Figure 3 Example item of the minimum standard absolute adjective scale <*breezy*, *windy*>. Bracketing in the instruction sentence represents the segments of the instruction revealed incrementally.

sentence in the current study is temporarily ambiguous given the visual display up to the noun modified by the critical adjective rather than up to the critical adjective. The key idea behind this set-up is that a way to disambiguate the sentence is by drawing the scalar implicature associated with the critical adjective (e.g., 'breezy but not windy weather'). This leaves no choice but to click on the Target object given that the Competitor object represents the stronger scale-mate negated in the scalar reasoning. Hence, higher proportions of Target selections during the temporarily ambiguous sentence would indicate the computation of the relevant upper-bounded interpretation of the critical adjective. Importantly, the weak scale-mate of each adjective scale was found by van Tiel & Pankratz (2021) to trigger SIs when judged against a Competitor picture in a picture verification task.

We had 6 critical trials, 3 for each Adjective Type (i.e., 3 relative adjectives and 3 minimum-standard adjectives), each appearing in both contrast and no-contrast conditions. Critical trials were distributed across two lists such that one list included 3 trials in the contrast condition of relative adjectives and 3 in the no-contrast condition of minimum standard adjectives, and the other list had 3 trials in the no-contrast condition of relative adjectives and 3 in the contrast condition of minimum standard adjectives.

We also included 18 filler trials in every list, which were adapted from van Tiel & Pankratz's (2021) materials too. We had 3 different types of fillers to ensure a counterbalanced design: (i) 6 filler trials where the picture representing the strong scale-mate of a given scale (e.g., *hot, windy*) was the Target object referred to by the instruction sentence (3 trials had 3 objects belonging to the same category (e.g., glass of water) and 3 had 2 objects of the same category, namely those representing the weak and the strong scale-mate of a given scale), (ii) 6 filler trials where the picture representing the Contrast object or antonym for a given scale (e.g., *cold water, windless weather*) was the Target referent of the instruction sentence (3 trials had 3 objects belonging to the same category and 3 had 2 objects of the same category, namely those representing the weak scale-mate and the antonym of a given scale), and (iii) 6 filler items where a distractor object was being referred to as the Target of the instruction (3 trials had 3 similar objects of the same kind and 3 had 2 objects of the same kind, namely, the pictures representing the weak and the stronger scale-mates of a given scale, respectively).

Each of the four image types in the critical trials (Target, Competitor, Distractor 1, Contrast/Distractor 2) was rotated through the four positions of the visual display per condition while overall each type of image appeared equal number of times ($N = 6$) at the four different positions of the display.

Each stimulus list had 24 test trials in total and we created 10 pseudorandomized orders of those for each list (i.e., 20 pseudorandomized lists in total). Participants were randomly assigned to a list and were rotated through the 20 pseudorsndomized lists.

In order to familiarize participants with the task, we also created six practice trials resembling the test trials: 2 of those had 3 objects of the same type on the visual display, another 2 had 2 objects of the same type, and 2 had 4 distinct objects on the visual display. Contrary to the test trials, the critical adjective in the practice trials was a material, a pattern or a color adjective, and the Target referent was uniquely identifiable either early (critical adjective) or late (*with*-prepositional phrase).

3.2.3 Procedure

Participants first answered demographic questions and then read the instructions (adapted from Sun, Pankratz & van Tiel 2021). They were informed that they would see sequences of displays of 4 pictures and that a written sentence would appear incrementally at the top of each display. This incrementally revealed sentence instructed them to click on one of the pictures of the display and gave a description of that picture. Participants' task was to guess which of the 4 pictures (every revealed segment of) the sentence was referring to. To make a guess, they had to click on one of the 4 pictures. When they made a guess by clicking on a picture, more words of the sentence appeared. Participants were also told that sometimes it would not be possible to determine which picture a revealed segment of the sentence is referring to (ambiguity windows), and that they would have to make a guess in these cases.

The instruction sentence was broken down into 4 segments/windows, revealed incrementally: the baseline window (*Click on the picture of the*), the critical adjective window (*warm/breezy*), the critical noun window (*water/weather*), and the disambiguation window (*with the purple spoon/with the yellow flag*), see Figures 2 and 3. The critical adjective and noun windows were the two ambiguity windows.

After they read the instructions, participants proceeded to the practice phase. During the practice phase, participants would receive feedback if they clicked on an incorrect referent for the instruction sentence once the disambiguation window was revealed. After completing the practice phase, participants were free to start the actual experiment.

3.3 Results and discussion

The data of one participant were removed because they took the study on their smartphone. In addition, following Degen, Kursat & Leigh (2021), we excluded participants ($N = 11$) with accuracy lower than 95% in selecting the correct referent in the last window of disambiguation (in the critical trials), where the correct referent was unambiguously the Target one. We further removed trials on which participants selected the wrong referent in this last window. The remaining data of 107 participants were included for further analysis.[7] Figure 4 shows proportions of choices of each of the four visual objects per condition per window. Error bars represent 95% CIs.

The data were analyzed using R (version 4.0.5). We fit mixed-effects logistic models for each of the two ambiguity windows (Adjective, Noun), predicting choices of Target over Competitor in terms of Adjective Type and Contrast (sum-coded fixed effects), including the maximal converging random-effect structure justified by the experimental design.

The analysis revealed a significant interaction effect in the Noun window ($\beta = -0.37$, $SE = 0.17$, $t = -2.21$, $p < 0.05$) and a significant main effect of Adjective Type ($\beta = 0.89$, $SE = 0.45$, $t = 1.98$, $p < 0.05$), probably driven by the interaction effect. Table 1 shows the model we ran as well as its output. No effect in the Adjective window reached significance.

Crucially, the attested significant interaction in the Noun window indicates that the presence of a Contrast object affects the incremental processing of scalar implicatures differentially for different types of adjectives, confirming our prediction. Specifically, it is suggested that scalar implicatures of minimum standard adjectives are computed independently of the context while for relative adjectives the computation of scalar implicatures is aided by the context, and the presence of a Contrast

7 Here are the demographic data of these participants: 73 female, 29 male, 4 diverse, and 1 preferring not to say, mean age: 30.68, age range: 18-75.

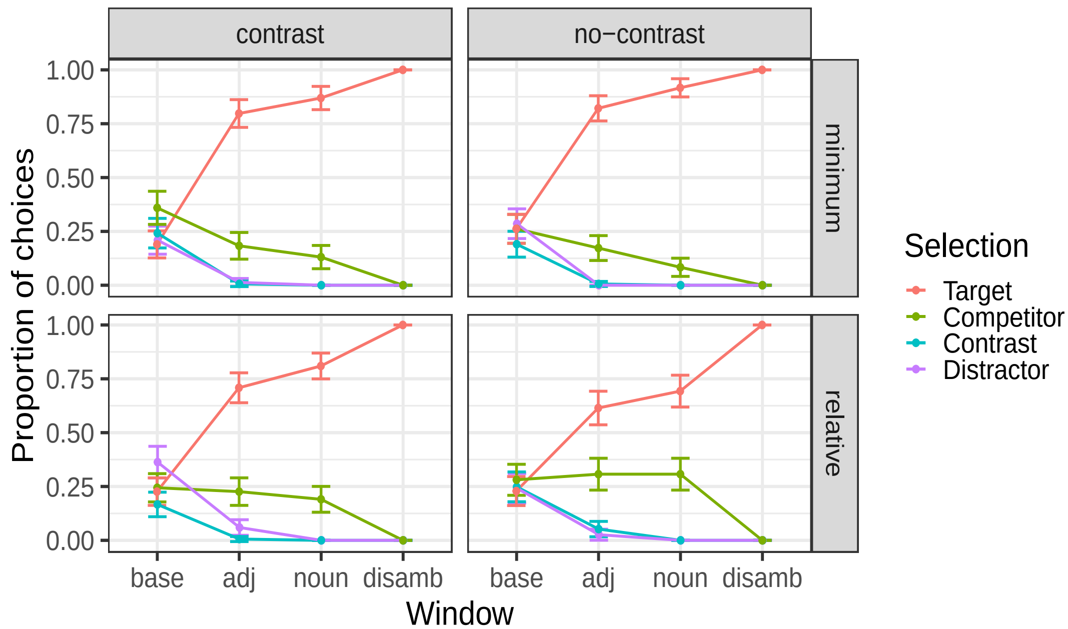


Figure 4 Proportions of choices of each of the four visual objects per condition per window. Error bars represent 95% CIs.

object in particular.

4 General discussion

The main finding of our study is that context (as instantiated by a Contrast object) affects the pragmatic interpretation of relative and absolute adjectives differentially, as hypothesized. Our findings indicate that semantic thresholds or lower bounds on a scale are computed incrementally with upper-bounded interpretations. We explain the observed differences between relative and absolute adjectives as follows.

Since minimum standard absolute adjectives instantiate a lower bound, listeners do not need to inspect the Contrast object in order to resolve their meaning. Relative adjectives, in turn, necessarily depend on contextual information to resolve their meaning.

There are different ways to explain the role of contrastive contextual information in the processing of relative adjectives. One possibility is that the Contrast object has the effect that the threshold for the weak term (e.g, *warm*) is set lower since the comparison class now includes lower degrees of temperature (see Barner & Snedeker

	Estimate	SE	z-value	p-value
Contrast	0.006367	0.135436	0.047	0.9625
AdjectiveType	0.886920	0.448350	1.978	0.0479 *
Contrast:AdjectiveType	-0.370029	0.167505	-2.209	0.0272 *

Table 1 Output of logistic mixed-effects regression model (Noun window): glmer(Target∼Contrast*AdjectiveType + (1+AdjectiveType|Participant) + (1|AdjectiveScale), data =dataNoun, family = binomial(link = logit))

2008; Schmidt et al. 2009; Solt & Gotzner 2012 for such comparison class effects). As a result, the lower bound of the weak term is further away from the Competitor (which instantiates a higher degree of temperature) and thus participants are more likely to choose the Target in the contrast condition compared to the no-contrast condition. Alternatively, it could be that participants are less likely to construct a stronger alternative in the no-contrast condition for relative adjectives so that the SI is likely to be computed.

Overall, our findings are in line with a hybrid model in which semantic and pragmatic meaning is resolved incrementally. The results may be taken to cast doubt on the assumption that a standard mechanism involving exhaustification with respect to alternatives underlies the interpretation of adjectival Horn scales. An alternative account of adjective meaning might be able to capture interpretative differences as a result of different priors (e.g., Lassiter & Goodman 2014) or based on different intervals on an underlying measurement scale – without postulating that comprehenders access a stronger alternative (see Gotzner 2021). Overall, we take this work to contribute to the growing literature indicating that to capture scalar inferences, a hybrid model incorporating both pragmatic principles and grammatical structure is needed (see especially Franke & Bergen 2020 and Cremers et al. 2022 in the domain of quantifiers).

5 Conclusion

Our study investigated how scalar meaning is resolved during incremental interpretation, focusing on the incremental processing of different types of gradable adjectives and of their upper-bounded/scalar inferences in particular. Specifically, we aimed to examine the role of context in the incremental upper-bounded/scalar interpretation of different types of gradable adjectives, i.e., relative (*warm*) and minimum-standard absolute adjectives (*breezy*). Our hypotheses were based on the following recent findings: (i) upper-bounded/scalar inferences of relative adjectives (‘warm but not hot’) are not derived in all contexts, contrary to absolute adjectives, as there is uncertainty as to the threshold of application of the weak (*warm*) and strong scale-mates (*hot*), and (ii) the processing of relative adjectives but not of minimum-standard absolute adjectives is facilitated by the context. Based on this, we hypothesized that the immediate visual context manipulated in our study (i.e., presence or absence of a Contrast object) will affect the scalar implicature computation differentially for relative (*warm*) and minimum-standard absolute adjectives (*breezy*).

As predicted, we found that the immediate visual context facilitates the processing of scalar implicatures triggered by relative adjectives ('warm but not hot'), whereas for minimum-standard adjectives scalar implicatures ('breezy but not windy') are computed independently of the context. Given this finding, we conclude that, unlike minimum-standard absolutes adjectives, inferences of relative adjectives are context-dependent. More generally, it is concluded that lexical-semantic factors, such as the type of standard invoked by a given adjective, affect the processing of the scalar implicature associated with the adjective. Overall, comprehenders resolve scalar meaning incrementally, during which semantic and pragmatic processing appear to be highly intertwined.

References

Aparicio, Helena, Chris Kennedy & Ming Xiang. 2018. Perceived informativity and referential effects of contrast in adjectivally modified NPs. In Elena Castroviejo, Louise McNally & Galit Weidman Sassoon (eds.), *The Semantics of Gradability, Vagueness, and Scale Structure*, 199–220. Springer Cham. doi:10.1007/978-3-319-77791-7_8.

Aparicio, Helena, Ming Xiang & Chris Kennedy. 2016. Processing gradable adjectives in context: A visual world study. In *Semantics and Linguistic Theory (SALT)* 25, 413–432. doi:10.3765/salt.v25i0.3128.

Barner, David & Jesse Snedeker. 2008. Compositionality and statistics in adjective acquisition: 4-year-olds interpret *tall* and *short* based on the size distributions of novel noun referents. *Child Development* 79(3). 594–608. doi:10.1111/j.1467-8624.2008.01145.x.

Cremers, Alexandre, Ethan G. Wilcox & Benjamin Spector. 2022. Exhaustivity and anti-exhaustivity in the rsa framework: Testing the effect of prior beliefs. doi:10.48550/ARXIV.2202.07023. https://arxiv.org/abs/2202.07023.

Cresswell, Max J. 1976. The semantics of degree. In B. Partee (ed.), *Montague Grammar*, New York: Academic Press. doi:10.1016/B978-0-12-545850-4.50015-7.

Degen, Judith, Leyla Kursat & Daisy Dorothy Leigh. 2021. Seeing is believing: Testing an explicit linking assumption for visual world eye-tracking in psycholinguistics. In *Annual Conference of the Cognitive Science Society (CogSci)*, vol. 43, 1500–1506.

Doran, Ryan, Gregory Ward, Yaron McNabb, Meredith Larson & Rachel E. Baker. 2012. A novel paradigm for distinguishing between what is said and what is implicated. *Language* 88. 124–154. doi:10.1353/lan.2012.0008.

Franke, Michael & Leon Bergen. 2020. Theory-driven statistical modeling for semantics and pragmatics: a case study on grammatically generated implicature readings. *Language* 96(2). e77–e96. doi:10.1353/lan.2020.0034.

Gotzner, Nicole. 2021. The use and interpretation of gradable adjectives. Invited talk at the Bochum Language Colloquium, Ruhr University Bochum, 28.1.2021.

Gotzner, Nicole, Stephanie Solt & Anton Benz. 2018a. Adjectival scales and three types of implicature. In *Semantics and Linguistic Theory (SALT)*, vol. 28, 409–432. doi:10.3765/salt.v28i0.4445.

Gotzner, Nicole, Stephanie Solt & Anton Benz. 2018b. Scalar diversity, negative strengthening, and adjectival semantics. *Frontiers in Psychology* 9. 1659. doi:10.3389/fpsyg.2018.01659.

Kamp, Hans & Antje Rossdeutscher. 1994. DRS-construction and lexically driven inference. *Theoretical Linguistics* (20). 165–235. doi:10.1515/thli.1994.20.2-3.165.

Kennedy, Chris. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30. 1–45. doi:10.1007/s10988-006-9008-0.

Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81. 345–381. http://www.jstor.org/stable/4489896.

Lassiter, Daniel & Noah D. Goodman. 2014. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In Todd Snider (ed.), *Semantics and Linguistic Theory (SALT)*, vol. 23, 587–610. CLC Publications. doi:10.3765/salt.v23i0.2658.

Leffel, Timothy, Alexandre Cremers, Nicole Gotzner & Jacopo Romoli. 2019. Vagueness in implicature: The case of modified adjectives. *Journal of Semantics* 36(2). 317–348. doi:10.1093/jos/ffy020.

Leffel, Timothy, Ming Xiang & Christopher Kennedy. 2016. Imprecision is pragmatic: evidence from referential processing. In *Semantics and Linguistic Theory (SALT)*, vol. 26, 836–854. doi:10.3765/salt.v26i0.3937.

Magri, Giorgio. 2017. Blindness, Short-Sightedness, and Hirschberg's Contextually Ordered Alternatives: A Reply to Schlenker (2012). In S. Pistoia-Reda & F. Domaneschi (eds.), *Linguistic and Psycholinguistic Approaches on Implicatures and Presuppositions* Palgrave Studies in Pragmatics, Language and Cognition, 9–54. Cham: Palgrave Macmillan. doi:10.1007/978-3-319-50696-8_2.

Qing, Ciyang & Michael Franke. 2014. Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In Todd Snider, Sarah D'Antonio & Mia Weigand (eds.), *Semantics and Linguistic Theory (SALT)*, vol. 24, 23–41. LSA and CLC Publications. doi:10.3765/salt.v24i0.2412.

Qing, Ciyang, Daniel Lassiter & Judith Degen. 2018. What do eye movements in the visual world reflect? A case study from adjectives. In *Annual Conference of the Cognitive Science Society (CogSci)*, vol. 40, 2297–2302.

Rotstein, Carmen & Yoad Winter. 2004. Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics* 12. 259–288. doi:10.1023/B:NALS.0000034517.56898.9a.

Schmidt, Lauren A., Noah D. Goodman, David Barner & Joshua B. Tenenbaum. 2009. How tall is tall? Compositionality, statistics, and gradable adjectives. In *Annual Conference of the Cognitive Science Society (CogSci)*, vol. 31, 2759–2764.

Sedivy, Julie C., Michael K. Tanenhaus, Craig G. Chambers & Gregory N. Carlson. 1999. Achieving incremental semantic interpretations through contextual representation. *Cognition* 71. 109–147. doi:10.1016/s0010-0277(99)00025-6.

Solt, Stephanie. 2015. Measurement scales in natural language. *Language and Linguistics Compass* 9(1). 14–32. doi:10.1111/lnc3.12101.

Solt, Stephanie & Nicole Gotzner. 2012. Experimenting with degree. In *Semantics and Linguistic Theory (SALT)*, vol. 22, 353–364. doi:10.3765/salt.v22i0.2636.

Sun, Chao, Elizabeth Pankratz & Bob van Tiel. 2021. Polarity modulates the processing cost of scalar implicatures. Short talk at the conference "Architectures and Mechanisms for Language Processing" (AMLaP), Université de Paris, France, 04.09.2021.

van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina & Bart Geurts. 2016. Scalar diversity. *Journal of Semantics* 33(1). 107–135. doi:10.1093/jos/ffu017.

van Tiel, Bob & Elizabeth Pankratz. 2021. Adjectival polarity and the processing of scalar inferences. *Glossa: A Journal of General Linguistics* 6(1). 32. doi:10.5334/gjgl.1457.

Stavroula Alexandropoulou
Institute of Cognitive Science
Osnabrück University
Wachsbleiche 27
49069 Osnabrück
stavroulaale@gmail.com

Marisha Herb
University of Potsdam
Am Neuen Palais 10
14469 Potsdam
herb@uni-potsdam.de

Henrik Discher
Humboldt University of Berlin
Unter den Linden 6
10117 Berlin
henrik.discher@web.de

Nicole Gotzner
Institute of Cognitive Science
Osnabrück University
Wachsbleiche 27
49069 Osnabrück
nicole.gotzner@googlemail.com