

Evaluation, Extent and Goldilocks

Rick Nouwen

Utrecht University

1 Evaluative polarity

I will call a word *evaluatively positive* if it expresses something that is good, or favourable or desirable. *Negative evaluation* is the opposite. So, for example, ‘*rude*’, ‘*stupid*’, ‘*ugly*’, ‘*disgusting*’ are evaluatively negative, while ‘*friendly*’, ‘*clever*’, ‘*beautiful*’ and ‘*tasty*’ are positive.

Evaluative meaning is often thought of as a form of *connotative meaning*. In the influential tradition of Charles Osgood’s semantic differential (Osgood, Suci, & Tannenbaum, 1957), evaluation is one of the cognitively universal dimensions that connect language and emotional experience. While Osgood saw connotation as fundamental to language and communication, connotative meaning plays a rather modest role in formal theoretical linguistic studies of meaning.¹ In formal linguistics, the focus is rather on the *denotative meanings*, which are generally void of affective content. In fact, work in the formal semantic tradition (Partee, 2004) tends to abstract away of *any* conceptual content, focusing instead on the logical structures that facilitate natural language understanding. A good example of this non-conceptual focus is formal semantic (and more broadly analytical philosophical) work on *vagueness*. The object is not to understand what it truly means for something to be ‘*rude*’, ‘*tall*’ or ‘*tasty*’, but the goal is rather to understand what the existence of such expressions means for fundamental assumptions underlying semantics, be they logical (e.g. Can we maintain bivalence?; Raffman, 2005), syntactic (e.g. What is the argument structure of adjectives?; von Stechow, 1984a, Kennedy, 2007) or pragmatic (e.g. How do vague meanings fit in a Gricean understanding of cooperative linguistic interaction?; Lassiter & Goodman, 2017).

¹In contrast, connotative meaning plays a huge role in NLP, especially in sentiment analysis. Also, evaluative polarity is discussed to some extent in cognitive semantics; see, for instance, Paradis, Van de Weijer, Willners, & Lindgren, 2012.

However, the area of vague adjectives is exactly where elements of connotative meaning, and in particular evaluation, may seem to start to become relevant to formal semantics. Antonyms of gradable adjectives express conceptual opposites in the connotative realm, and this opposition can be connected to entailment patterns between denotative meanings.

- (1) a. Scarlett is older than Oscar
 b. Oscar is younger than Scarlett

The comparatives in (1-a) and (1-b) entail each other, which in degree semantics is often taken to indicate that ‘*tall*’ and ‘*short*’ use the same scale, but with a different orientation (e.g. von Stechow, 1984b, Kennedy & McNally, 2005). There are many examples like this: ‘*old*’-‘*young*’, ‘*wide*’-‘*narrow*’, ‘*slow*’-‘*fast*’, etc. In antonym pairs like these, there’s always one member which is *marked*, in the sense that its use is more constrained and, when felicitous, leads to additional inferences (e.g. Lehrer, 1985). For instance, for dimensional adjectives like ‘*old*’-‘*young*’, only the unmarked member is felicitous in a multiplicative construction, as illustrated in (2). In degree questions, the marked member triggers an extra inference. For instance, while (3-a) reveals nothing in particular about the speaker’s assumptions, (3-b) makes apparent that the speaker is assuming Scarlett to be young.²

- (2) a. Scarlett is twice as old as Oscar.
 b. #Oscar is twice as young as Scarlett.
- (3) a. How old is Scarlett?
 b. How young is Scarlett?

Marked members of antonym pairs are often referred to as the *negative antonym*. It is tempting to think, then, that here we have a case where evaluative polarity is directly reflected in a formal semantic notion (like scale orientation and markedness). But this is too quick: evaluative polarity often does not match antonymic polarity. A good example is ‘*warm*’-‘*cold*’. Here ‘*cold*’ is the negative antonym: while (4-b) presupposes that it’s cold, (4-a) lacks a similar supposition.³

²Confusingly, the property displayed in (3-b) is sometimes referred to as *evaluative* in the semantic literature (Rett, 2014). Something is evaluative in this sense if it is norm-related. For instance, ‘*long*’ in ‘*The rod is long*’ is evaluative, but it is not in ‘*The rod is 10cm long*.’ In the remainder of this article, I will avoid this use of ‘evaluative’ and use evaluation consistently as a judgment on how good or bad, desirable or undesirable (etc.) something is.

³Note that we cannot perform the test in (2), because temperature does not behave as ratio scale, thus precluding multiplication (cf. Nouwen, 2018b).

- (4) a. How warm is it outside?
b. How cold is it outside?

The evaluative polarity of ‘*warm*’ and ‘*cold*’, however, is variable. For instance, ‘*cold*’ is usually a negative evaluation for soup and a positive one for beer. ‘*Warm weather*’ is good, but (again) nobody likes warm beer.

This is not to say that evaluative polarity is always variable. An adjective like ‘*beautiful*’ is inherently evaluatively positive and ‘*ugly*’ is clearly negative. Note, however, that this opposition does not track the antonymic relation. Both members of the ‘*beautiful*’-‘*ugly*’ pair behave as marked, for instance in the sense that both trigger speaker-related inferences when used in questions (Lehrer, 1985).

- (5) a. How beautiful is he?
b. How ugly is he?

In a nutshell, then, while polarity in connotation and in denotation sometimes overlap, there is little systematic interaction and, so, one might think, the formal semanticist can safely carry on ignoring connotation. Below, I will argue however that evaluative polarity *does* have an impact on denotational semantics. But this impact is not one of matching polarity, where for instance the directionality of the scale in the degree semantics of an adjective is determined by its evaluative polarity. Rather, I will argue that evaluation is a determining factor of *scalar extent* - the degree to which the property expressed by the adjective holds. I will provide two examples of phenomena where this link becomes apparent. In section 3, I will show that the intensifying effect of evaluative adverbs is partly determined by the evaluative polarity of the adverb and in section 4, I will show that some intensifying adverbs are sensitive to evaluative properties of the adjectives they combine with. Section 5 and 6 will discuss what needs to be in place to make sense of this connection between evaluation and denotation. My argumentation starts by showing that negative evaluation is naturally connected to *excessive extent*. To do this, I will use the parable of Goldilocks, to which I turn next.

2 Goldilocks Principles

In the story of Goldilocks and the three bears, two parent bears and their child bear get ready to eat their breakfast, but find the porridge they prepared to be too hot. They decide to go for a stroll through the woods, leaving their breakfast bowls to cool. In their absence an impish golden-haired girl,

Goldilocks, sneaks into their house. Among many other mischievous things she does, she tries out the bears' breakfast: Daddy's bowl of porridge is too hot, Mummy's bowl is too cold, but the little bear's porridge is *just right* and she proceeds to finish it off.

The Goldilocks fable does not excel in moral clarity. A common interpretation is that the story intends to show how the ill-natured selfish actions of Goldilocks affect the everyday life of the good-natured bears, where the girl's actions are particularly selfish because she takes only what is *just right*. The lasting legacy of the story, however, is simply this notion of *just rightness*. It turns out that it is handy to have a cultural reference point for all sorts of instantiations of the evaluatively ideal middle ground. The notion is similar to that of *the golden mean* from Aristotle's philosophy of virtue, but applicable much more broadly than to just the question of how to live life virtuously. In fact, the story of Goldilocks and the three bears has been a productive inspiration for quite a few scientific disciplines.

A typical *Goldilocks Principle* in science involves some ideal state of affairs that is positioned between states of affairs that are less ideal. Or stated differently, from the perspective of many different sciences, ideal situations tend to not be extreme situations, but rather moderate in some sense. One generally well-known example is from planetary science, where the so-called *Goldilocks zone* is the stretch of space around a star that allows liquid water to occur on a planet: it is sufficiently far from the sun to not let all water evaporate away but also not too far to only end up with ice. The details of how to establish which distance from the sun qualifies as a habital zone has subsequently become known as the *Goldilocks problem of climatology* (Kasting, Toon, & Pollack, 1988). In cognitive science, there's the *Goldilocks effect* (Kidd, 2012), the observation that the visual attention of infants tends to be highest when the stimulus is of intermediate complexity: while low complexity stimuli tend to already be familiar, too much complexity wastes valuable computational resources. In economy, there's the concept of a *Goldilocks economy*, which is an economy that has *Goldilocks growth*: growth that is prosperous but not so extreme that it causes excessive inflation.

Here, I will point out the relevance of Goldilocks to linguistic semantics. One way to understand Goldilocks principles is by seeing them as principles concerning the very fabric of *evaluation*: that evaluation is sensitive to *excess*. I can illustrate what I mean by this with my particular version of the Goldilocks concept. I will describe it in two separate Goldilocks principles:

The Goldilocks Principles of Evaluation

1. You can have too much of a good thing
2. You can't have so much of a bad thing that it becomes good again

These two principles together have exactly the intended effect. Being removed from the sun is *good* because it prevents the evaporation of water, but once you're too far, water will freeze and a good thing turns *bad*. And from there on, it will only get worse. Similarly, being close to the sun is *good* because it prevents the freezing of water, but, once again, being too close is *bad* because water will start to evaporate.

This connection between goodness, badness and excess can be applied much more generally, however, to the semantics of adjectives. Here's an example. Imagine some food that is really '*bland*' (=negative evaluative polarity). By adding salt the food will become tasty (=positive evaluative polarity), but if you keep on adding salt, the food will become '*salty*' (=negative evaluative polarity). And no extra added amount of salt is ever going to make it taste fine again. The same reasoning works the other way around: You made a recipe using loads of salt and it was inedible (=bad). Next, time you reduce the salt and it is now enjoyable (=good). The next time you reduced the salt even further, only to discover that the food is now bland (=bad). Because we can go through this reasoning in either direction, we need only the two principles stated here. (I could have added something like *You can have too little of a good thing*, but this is really not necessary. Not enough salt, is the same as having reduced the salt by too much.)

These principles are by no means universally valid. They *tend* to be true, but the world is not exclusively made up of Goldilocks-zones. For instance, sometimes more is better, ad infinitum, in breach with the first principle. A judge can't really ever be *too fair*. Ice cream can't really ever be *too tasty*. The strongest example of such a lack of excess is closeness to some ideal: if *G* is some ideal state, then obviously "being too close to *G*" is a non-sensical notion.

Crucial about my conception of Goldilocks principles is that they have to do with *excess*. The middle zone is ideal because it lacks excess: *goodness* occurs when excess doesn't. Conversely, *badness* involves any kind of excess, either of the extent to which some property holds or of the extent to which it does not hold. All this may seem very obvious, and I suppose it is. However, as I will show, these simple observations can explain why degree semantics

is sometimes sensitive to evaluative polarity. Take the example of ‘*warm*’-‘*cold*’ that I introduced above. Cold soup has negative evaluative polarity, since the soup has an excess of coldness. Cold beer does not suffer from a similar excess. Negative polar ‘*warm beer*’ is beer that is *too* warm, while the positive ‘*warm weather*’ does not involve excess.

In a nutshell then, there is a correspondence between evaluation and degree. In this paper, I will particularly connect that observation to adverbs of degree. I will seek to use Goldilocks to explain the degree of intensification that certain adverbs can bring about. In particular, Goldilocks explains why adverbs derived from negative valence adjectives tend to express high or extreme degree, while degree adverbs derived from more positive words express medium degree. Take for example (6) and (7):

(6) Scarlett is pretty tall.

(7) Scarlett is terribly tall.

There is a clear difference in the effect that the adverbs bring about in these two sentences. Scarlett is said to be quite a bit taller in (7) than in (6). As I will explain, this seems to be a general tendency, one that to my knowledge has received all but no attention in the literature.⁴

Given Goldilocks, the contrast between (6) and (7) makes intuitive sense. To be tall in a *terrible* way means to be excessively tall. To be tall in a positive way means that there is no excess involved. The remainder of this paper will make this idea more precise.

3 Evaluation and Scalar extent

When an intensifying adverb (Bolinger, 1972) modifies some gradable predicate, it expresses the extent to which that predicate holds. I will refer to this semantic effect as the *scalar extent*.⁵ For instance, *very* expresses high scalar extent, while *pretty* expresses a somewhat lower extent, etc. The suggestion I made above is that scalar extent is tied to evaluation. The idea is that intensifiers based on words with a positive connotation intensify to a lesser degree (are associated with lower scalar extent) than intensifiers based

⁴The only exception I know of is a short passage in Rissanen (2008).

⁵This is not to say that the semantics of such adverbs necessarily *is* their scalar extent. Rather, scalar extent is a shallow abstraction of the semantic effect of applying the adverb. For instance, one may adopt the influential proposal from Wheeler (1972) on the semantics of ‘*very*’, where ‘*being a very tall man*’ is given the semantics paraphrasable as ‘*being very tall for a tall man*’. The effect of this semantics will be that the scalar extent of ‘*very*’ is high degree. See section 5 for discussion.

on negative words. This idea is supported in English by adverbs of what I will call *medium* degree like ‘fairly’, ‘pretty’, ‘reasonably’ –all, arguably, of positive evaluative polarity– and adverbs of *high* degree like ‘terribly’, ‘terrifically’, ‘awfully’, etc. –all, arguably, of negative evaluative polarity. A quick look at German and Dutch yields similar intuitions. There, too, adverbs indicating medium degree typically derive from positive adjectives, while high degree adverbs derive from negative adjectives. For instance:

	Dutch	German
high degree (<i>negative valence</i>)	‘verschrikkelijk’ (<i>terrible</i>) ‘erg’ (<i>bad</i>) ‘ontzettend’ (<i>disrupting</i>)	‘schrecklich’ (<i>terrible</i>) ‘sehr’ (<i>painful</i>) ‘fürchterlich’ (<i>scarily</i>)
medium degree (<i>positive valence</i>)	‘aardig’ (<i>nice</i>) ‘best’ (<i>best</i>) ‘tamelijk’ (<i>fitting</i>)	‘recht’ (<i>right / just</i>) ‘leidlich’ (<i>tolerable</i>) ‘ziemlich’ (<i>fitting</i>)

It would be good to be able to get some sense of how general a pattern this is. To this end, I conducted a small study in which I compared measures of scalar extent to measures of valence associated to adverbs of degree. I hand-selected 24 adverbs that are derived from an adjective expressing a subjective, affective evaluation. I thereby avoided adverbs with specialized functions, such as ‘completely’ (end-of-scale) or ‘nearly’ (approximation). Also, I avoided modal evaluations involving surprise / expectation, like ‘surprisingly’ for reasons that will become apparent below.

I used Amazon Mechanical Turk to ask 61 participants to associate modifications of the adjective ‘warm’ to temperatures. Participants saw sentences of the form in (8) and were asked to indicate what temperature they thought it was, given that this sentence was produced on a spring day in New York City.

(8) It’s ADVERB warm today.

Participants had to provide the temperature they associated to the stimuli in degrees Fahrenheit by moving a slider to the desired position. The slider scale ranged from -10 to +110 degrees Fahrenheit. The aforementioned 24 adverbs took the place of ADVERB in (8) and there was an additional condition in which the adjective was left unmodified. Another manipulation was that the 25 sentences that are thus constructed were also offered to the participants with a negation added into them: ‘It isn’t ADVERB warm today’. All participants saw all 50 stimuli (corresponding to 50 conditions), which were

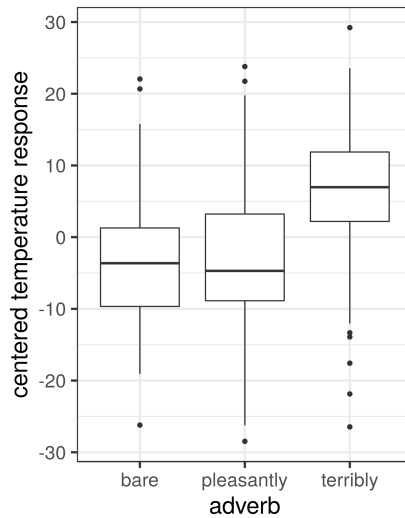


Figure 1: Example of the results for three conditions

presented in randomised order. There were no fillers.

I centered responses by converting each response of a participant to the distance to the mean of all responses by that participant. I removed all observations that were more than 2 standard deviations removed from the mean response per adverb condition. This removed exactly 100 of the 1525 observations.

The condition with negation was intended to test pragmatic strengthening effects of degree adverbs. Since no reliable differences were found for this condition, it is omitted from subsequent discussion.

The results of the positive condition in this survey give an indication of the scalar extent that adverbs are associated with.⁶ See figure 1 for an example.

The next step in this study is to connect these results to evaluative polarity. Luckily, since Osgood et al. (1957), factor analysis studies have confirmed time and time again that evaluative polarity is one of (usually) three factors that determine a word’s connotative meaning. This means that there is ample of work on establishing lexicons that record affective measures, including those related to evaluation. For the current study, I used the NRC valence-arousal-dominance lexicon (Mohammad, 2018) to obtain estimates of evaluative connotations associated to the adjectival stem of the adverbs

⁶Admittedly, the reliability of the responses is not optimal, given the artificial nature of the task. However, despite this, this kind of setup has been proven to be useful. See, for instance, Bennett and Goodman (2018) for a study involving a similar methodology.

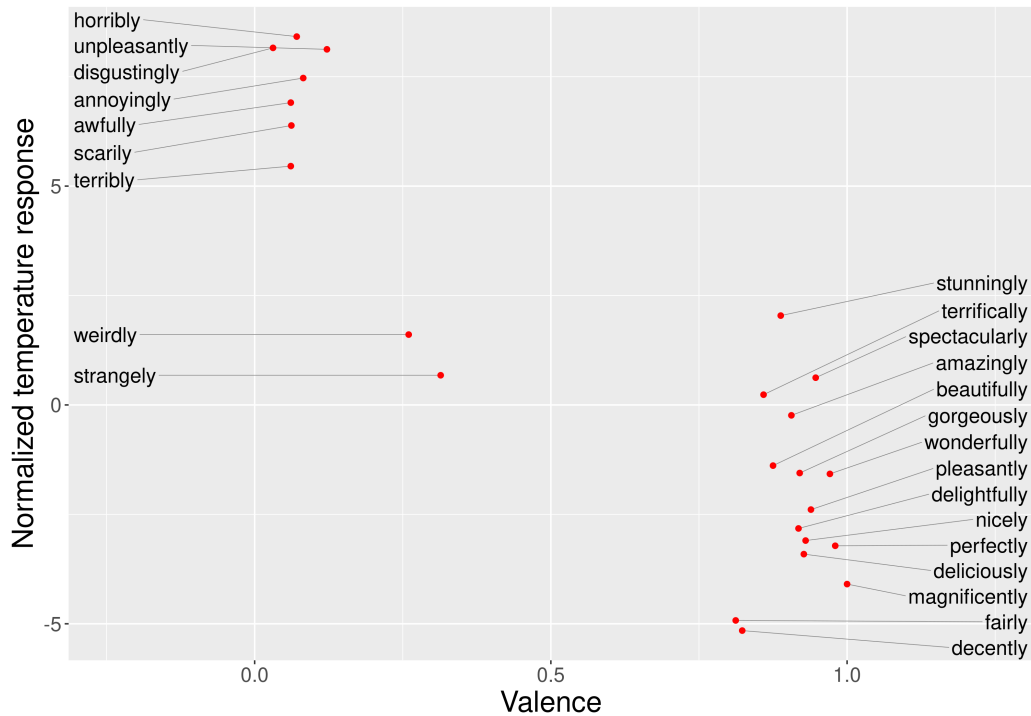


Figure 2: Mean centered temperature response versus valence

used in the mechanical turk survey. Here, *valence* corresponds to what I (following Osgood) have been calling evaluation. In what follows, I will use *valence* for the numerical score of how (evaluatively) positive or negative a word is, while using *evaluation* for the general underlying concept. In the NRC lexicon, valence is quantified on a 0 to 1 scale: 0 corresponding to extreme negative evaluation and 1 to extreme positive evaluation.⁷

Figure 2 plots the mean centered temperature response from the mechanical turk survey against the NRC VAD valence score of the adverbs. As can be seen, there is a clear negative correlation between the two.⁸ The higher the valence of the underlying adjectival predicate of an adverb, the lower the scalar extent that this adverb tends to express.

What is also clear from this plot is that the adverbs form two groups. In fact, what drives the correlation seems to be the fact that most adverbs

⁷The NRC VAD lexicon is available at: <https://saifmohammad.com/WebPages/nrc-vad.html>. This lexicon was built by asking annotators to rank four adjectives according to the relevant property (e.g. valence). The results is a reliably consistent annotation. See Mohammad (2018) for details on the methodology.

⁸Spearman’s rank correlation: $\rho = -0.73$, $p < 0.0001$. I’m using Spearman correlation, since neither variable is normally distributed.

have a valence close to either 1 or 0. The former tend to express relatively low degrees, the latter relatively high degrees. And, so, we see the adverbs of high degree on the left and the adverbs of medium degree on the right. This shows that scalar extent is to a large part determined by evaluative polarity, as would be expected from the Goldilocks connection between valence and excess.

4 A closer look at scalar extent

Languages tend to come with a diverse array of degree-modifying adverbs. These have specialized roles like expressing approximation (*‘almost’*, *‘barely’*), proportion (*‘completely’*, *‘totally’*, *‘half’*), or sufficiency and excess (*‘enough’*, *‘so’*, *‘too’*). Here, my main concern will be an often-made distinction between three kinds of adverbs of degree, that tracks scalar extent:

- Adverbs of low scalar extent, or *L-adverbs*.
e.g. *‘a bit’*, *‘somewhat’*, *‘slightly’*
- Adverbs of medium scalar extent, or *M-adverbs*.
e.g. *‘fairly’*, *‘pretty’*, *‘quite’*
- Adverbs of high scalar extent, or *H-adverbs*.
e.g. *‘very’*, *‘terribly’*, *‘extremely’*

Above, I showed some empirical support for this distinction between medium and high scalar extent modifiers. Not only did the sample of adverbs I looked at split in two distinct groups in terms of scalar extent, this division coincided with a split in evaluative polarity of the underlying adjective.

Adverb typologies like the one above are extremely frequent in the literature on degree. An early influential version of this way of thinking of degree adverbs originates in Stoffel (1901), who made a binary distinction between *intensives* (roughly, H-adverbs) and *downtoners* (roughly, M- and L-adverbs). A much more commonly seen way of dividing things up in the past few decades uses the three-way distinction into low, medium and high degree (Biedermann, 1969; Bolinger, 1972; Bäcklund, 1973; Gary, 1979; Van Os, 1989; Klein, 1998; Paradis, 1997).⁹ Such classifications are standardly made

⁹In fact, Klein follows van Os in assuming a separate category for degree modifiers expressing *extreme* high degrees. Note, furthermore, that the terminology used here is mine. I introduce these new terms mainly to avoid a fair amount of terminological confusion that exists in the literature. More traditional names include for instance *boosters* for H-adverbs, *moderators* or *compromisers* for M-adverbs and *diminishers* or *minimizers* for L-adverbs.

on the basis of scalar extent alone. Given that, the question arises what motivates a three-partite division. Why would we divide things up in *low*, *middle* and *high*, as if we have to model our adverbs in parallel to the breakfast bowls of Goldilocks' three bears? Why not a continuum? Classification only makes sense, if the type of scalar extent of an adverb coincides with some independent property these adverb classes have.

Of course, if the distinction between medium and high scalar extent is tracked by the evaluative polarity of the predicate underlying the adverb, as I suggested above, then this will give us a meaningful classification for M- and H-adverbs. Ultimately, however, I will question the value of this way of thinking about scalar extent. Before I do so, I turn to the part of the adverb classification that we haven't seen yet – L-adverbs – and show that they too provide evidence for a connection between evaluation polarity and degree semantics.

4.1 L-adverbs

It is well known that L-adverbs are sensitive to scale structure (e.g., Kennedy & McNally, 2005). In particular, they only combine with adjectives associated to a lower-closed scale. For instance, '*a bit*' is felicitous with an adjective like '*wet*', since the scale of *wetness* is lower bounded: it starts at completely dry and then involves increasing amounts of liquid. It is infelicitous with adjectives that lack a natural lower bound, such as for instance '*nice*'.

- (9) a. This towel is a bit wet.
 b. *This man is a bit nice.

In contrast to L-adverbs, M-adverbs and H-adverbs are not sensitive to scale structure:

- (10) a. This man is pretty nice.
 b. This man is very nice.

This would give us a way of predicting low scalar extent: adverbs compatible only with lower-bounded scales tend to express low scalar extent. Unfortunately, this particular way of distinguishing L-adverbs from M- and H-adverbs is not perfect. While most informants would agree there is a clear difference between (9-a) and (9-b), examples such as (11) are not so clearly infelicitous.

- (11) ?The towel is a bit thick.

Examples like (11) improve when an excessive reading is plausible (Paradis,

1997, p.76). That is, when acceptable, (11) tends to be interpreted as the towel being (slightly) too thick. This dependence on excess is the case more generally. While at first sight, a combination like ‘*a bit tall*’ may raise eyebrows, given the right context (namely, one involving excess) it becomes felicitous. Imagine, for instance, a context where two movie casting directors are trying to find someone who could play some historical figure who happened to be quite short. One of them suggests Scarlett Johanson. The other replies:

(12) Not sure. I think Scarlett Johanson is a bit tall.

and thus communicates that they are of the opinion that she may be too tall to play a short character. To make sense of this, first notice that excess is itself a lower-bounded scale, witness the acceptability of L-adverbs with the overt excess marker ‘*too*’, as in (13-a).

(13) a. This towel is a bit too thick.
 b. *This towel is a bit thick enough.

This means we made the following parallel observations: (i) L-adverbs combine only with adjectives associated to lower-bounded scales; (ii) excess behaves like such a lower-bounded scale; (iii) combinations of L-adverbs and adjectives that are not associated to lower-bounded scales can be felicitous when interpreted as expressing excess, and, as such, when interpreted with respect to a lower-bounded scale.

The link to evaluative polarity becomes apparent when we look at another common observation about the distribution of L-adverbs. L-adverbs tend to combine more easily with evaluatively negative adjectives than with adjectives expressing positive evaluation (e.g. Bylinina & Zadorozhny, 2012; Jaspers, van Craenenbroeck, & Wyngaerd, 2016; Nouwen, 2018a). Consider the contrasts in (14).

(14) a. That’s slightly odd / ?normal.
 b. He’s a bit tired / ?awake.
 c. My cat’s a bit sick / ?healthy.

This observation, too, makes sense once we see negative evaluation as involving excess. My reasoning about this observation is as follows: (i) L-adverbs have the distinguishing feature that they require bottom-closed scales; (ii) since negative evaluation involves excess and excess involves a scale minimum, L-adverbs are compatible with negative adjectives. As a result, I make the prediction that L-adverbs should co-occur more with negative valence adjectives than M/H-adverbs, since the latter lack the scalar sensitivity that

L-adverbs have. As I explained above, most adjectives have variable evaluative polarity. So, the prediction is rather that the *easier* it is for an adjective to receive a negative, i.e. excessive, connotation, the more often we should see that adjective pair with an L-adverb.

I conducted a corpus study to test this prediction. For this study, I used the hotel review dataset of Ganesan and Zhai (2012), which consists of roughly 259k English language hotel reviews.¹⁰ The rationale for using this dataset is that it includes a good deal of evaluative language. I extracted co-occurrences of a set of degree-related expressions (L-, M- and H-adverbs) with a adjectives.¹¹ Subsequently, I used the NRC valence-arousal-dominance lexicon (Mohammad, 2018; see above) to obtain valence scores for the adjectives in the hotel review data. Using the corpus and the lexicon, I then compared the sets of adjectives that different degree expressions combined with. The prediction is that the mean valence of adjectives that L-adverbs combine with is lower than the mean valence of adjectives that combine with degree expressions that are not constrained to combine with lower-closed scale adjectives. So, combinations with degree expressions that indicate medium or high degree are predicted to involve, on average, higher valence adjectives.

The results are in figure 3. As can be clearly seen, L-adverbs tend to co-occur with more negative adjectives, when compared to M-adverbs and H-adverbs. These latter two classes differ significantly from L-adverbs ($t=11.79$, $p<0.001$ and $t=15.44$, $p<0.001$, respectively).¹²

This study on L-adverbs shows further support for the idea that evaluation can have an impact on issues of denotative semantics. Here, the sensitivity of scale structure displayed by adverbs of low degree entails that such adverbs are also sensitive to evaluative aspects of the meaning of adjectives they combine with. This is the case since negative adjectives express excess and excess is lower-bounded. M- and H-adverbs show no such sen-

¹⁰The dataset is available at: <http://kavita-ganesan.com/entity-ranking-data/>.

¹¹This adverb set consisted of: ‘*a bit, slightly, somewhat, very, so, really, too, quite, as, more, extremely, pretty, little, most, rather, absolutely, rather, incredibly, relatively, perfectly, completely*’. This choice was based on the desire to have a wide range of different kinds of degree expressions and was further constrained by demanding significant frequency.

¹²Figure 3 also suggests that ‘*rather*’ has a relative preference for combinations with low valence words. This is, in fact, in line with a very early intuition in Stoffel (1901): “Suppose a man under treatment at an hydropathic establishment, about to take his morning-bath, were to put his hand into the water, and say to the attendant: «The water’s rather cold this morning», then he would be expressing unpleasant surprise at finding the water colder than he had expected. And if the attendant were to answer: «Yes, sir, it’s *pretty* cold, but then it’s the doctor’s orders», he would merely make an objective statement as to the temperature of the water.” (page 132).

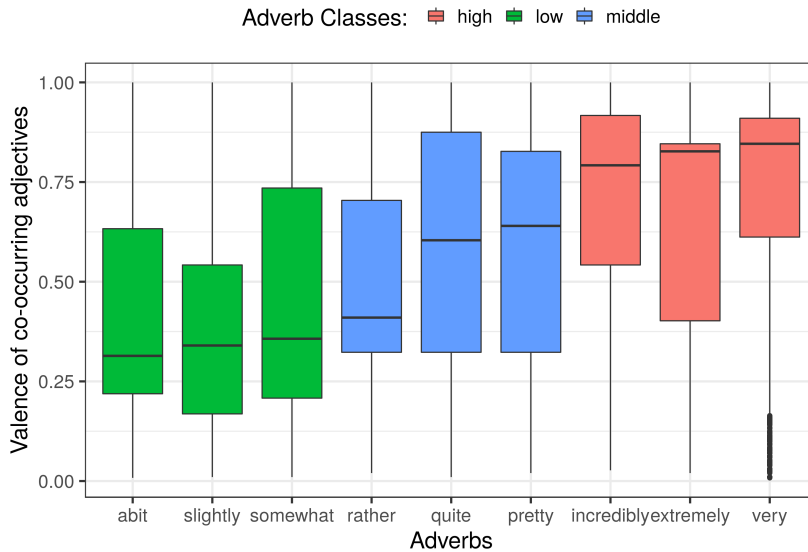


Figure 3: Mean valence of adjectives combining with degree expressions.

sitivity. Yet, they can be distinguished by the valence of their adjectival root. M-adverbs express medium degree because the underlying predicate is positive in valence. H-adverbs tend to be derived from negative predicates.

5 Bleaching

While I think that the above way of thinking about adverbial degree modification is insightful, I should hasten to add that it may not be so generally applicable. In fact, ultimately, I think that scalar extent classification of adverbs is ill-conceived. This is because degree adverb typologies like the one typically adopted in the literature assume that adverbs of degree come with some (lexically) fixed expression of scalar extent. This, I will argue now, is far too simplistic. There are many intensifiers that have variable – in fact, context-dependent – scalar extent. Take (15):

(15) It is *surprisingly* warm today.

Without specifying a context, one may imagine this sentence being said about a day that is much warmer than usual, so much warmer in fact that the speaker is surprised. But it would be wrong to assume that this means that *‘surprisingly’* is an adverb that indicates high degree. Imagine a different context in which the weather models predicted today to be extremely cold,

much colder than normal for the time of year. It turns out that the models, which are usually very reliable, were wrong. It *is* much colder than normal for the time of year, but it is still quite a bit warmer than the models predicted. In such a context, (15) makes sense. However, the adverb clearly cannot be said to express high degree.

A similar example is (16). This expresses merely that it warmer than I hoped it would: (16) does not entail that it was very warm. Think of a context: my 7-year old son hopes that the weather will be extremely cold, for only when the temperature reaches -10C will the school be closed. It turns out that it is actually very cold, but unfortunately for my son, it is -8C. In such a context, it is felicitous for my son to assert (16).

(16) It is *disappointingly* warm today.

This shows that it is important to distinguish intensifiers like ‘*very*’, that specialize in expressing high degree, from intensifiers like ‘*disappointingly*’ and ‘*surprisingly*’ that fulfill their degree function indirectly. Another way of saying this is that ‘*very*’ is *bleached*. All its meaning is directed at affecting scalar extent. In contrast, examples like (15) and (16) do not just express how warm it is, but also that the temperature is surprising / disappointing.

A good example of a bleached intensifier in English is ‘*terribly*’. In contrast to ‘*very*’, which has lost its original meaning of *true / real* centuries ago, for ‘*terribly*’ it is very clear what its non-degree meaning would be – after all, the stem ‘*terrible*’ is in use and has nothing to do with degree. But whenever we use ‘*terribly*’ as an adverb of degree, the meaning of this stem is completely absent from the resulting interpretation. For instance, there is no contradiction in (17-a) and nothing bad ends up being said about the man in question. In clear contrast, in (17-b) the man is said to have two contrasting (contradictory?) properties. The example in (17-c), in which an adverb that has not been bleached, ‘*disgustingly*’, modifies ‘*nice*’, equally conveys a contrast or even a contradiction.

- (17) a. He is *terribly* nice.
b. That man is both nice and *terrible*.
c. He is *disgustingly* nice.

In sum, the fact that not all adverbs of degree are bleached – adverbs like ‘*disgustingly*’, ‘*disappointingly*’, ‘*surprisingly*’ and many others are what I will call *unbleached* – makes it impossible to characterize intensifiers simply in terms of their *scalar extent*. Unbleached intensifiers do not straightforwardly express such an extent. Instead, they convey the property expressed by their adjectival stem and modify degree relative to that property.

Although I won't have anything deep to say about the diachronic process of bleaching, it seems evident to me that at least some part of that process is simply that bleached adverbs end up being associated to the scalar extent their initial unbleached version was regularly associated with. If an unbleached adverb, through the lexical content of its adjectival stem, ends up typically expressing high degree, then the unmediated expression of high degree is a natural candidate for the meaning of a bleached version of this adverb. For instance, the fact that '*terribly*' now expresses high degree suggests that it was typically associated to high degree when it was still unbleached. Similarly for '*pretty*', which is also bleached (witness the possibility of calling someone '*pretty ugly*'), and its tendency to express medium degree.

For this to make sense, however, it needs to be the case that the adjectival predicate underlying an adverb of degree has some role to play in what scalar extent is typically expressed by the adverb. The link between Goldilocks and evaluation that I developed above predicts that whenever the adverb is associated with some evaluative pole, the scalar extent of the adverb will tend to be in line with this value: medium degrees tend to be *good*, extreme degrees (because they involve excess) tend to be *bad*.

6 Connecting evaluation with degree

How does the unbleached evaluative content of adverbs interact with the degree semantics of adjectives? In this section, I will sketch what I think a formal prosopoeal would need to look like.

Wheeler (1972) proposes that unbleached degree modifiers are best analyzed as factive propositional operators. On his analysis, (18) is interpreted as *it is horrible that it is as warm as it is*. Morzycki (2008) argues against such an analysis on the basis that it would wrongly predict that the weather can be described as '*horribly warm*' also when it is very cold. On a freezingly cold day, most people would agree that it is indeed horrible that it is as warm as it is, but no-one would agree that it is '*horribly warm*'.

(18) Today is horribly warm.

Morzycki solves this by stipulating that evaluative intensifiers express extreme degrees: '*horribly warm*' means that it is horrible how extremely warm it is. This analysis in turn predicts that evaluative adverbs always express high degree. As I explained above, this is not the case. It can be '*disappointingly warm*' without being warm to a high degree. Moreover, evaluatively positive adverbs tend to express medium rather than high degree, as predicted by the Goldilocks approach. Morzycki's complaint about a Wheeler-

style analysis is valid, yet the explicit encoding of a particular scalar extent in the lexical content fails to account for the way the content of the adverbs’ underlying predicate determines the scalar extent. Wheeler’s analysis does lay such a connection. For that reason, I will now develop an account that builds on Wheeler’s idea without running into the objection raised by Morzycki. Before I can do so, let me introduce a couple of assumptions I make.

Let’s assume a standard degree-based semantics of adjectives. I assume that adjectives are associated to measure functions, which I will generally write as μ . For instance, the adjective ‘*tall*’ is associated to the measure function μ_{tall} , which takes an individual and returns a *degree* corresponding to that individual’s height.

I follow Kennedy (2007) in assuming that adjectives in their positive form are interpreted as exceeding a threshold θ . That is, ‘*Sue is tall*’ corresponds to:

$$\mu_{\text{tall}}(\text{Sue}) > \theta$$

The value of θ is such that subjects whose measurement exceeds the threshold *stand out* (Fara, 2000). The standard way of thinking about *standing out* is extensional in nature: the tallest child in her class is ‘*tall*’ because she literally *stands out* when compared to her peers. Standing out can also be thought of in intensional ways, though (e.g. Fara, 2000; Bylinina, 2014). When I decide that the queue in the supermarket that I am in is ‘*long*’, it doesn’t necessarily mean that it stands out when compared to the length of neighbouring queues. What I could mean is simply that the queue is significantly longer than what it normally is like, or than what I hoped it to be, etc. In general, all sorts of considerations may influence what in a certain context does and does not stand out.

Standing out is the interpretation given to an adjective’s positive form. But what happens when that positive form adjective is modified? It is not uncommon to think that modifiers may influence how the positive form is interpreted, that is, how the subject is said to be standing out. For instance, ‘*for*’ phrases could be seen as specifying the comparison class used to determine θ (Solt, 2011). A sentence like ‘*Sue is tall for a 4 year old*’ conveys that $\mu_{\text{tall}}(\text{Sue}) > \theta$ in such a way that Sue stands out when compared to 4 year olds.¹³ Wheeler’s (1972) analysis of ‘*very*’ can be seen in the same spirit. According to that proposal, ‘*Sue is very tall*’ is interpreted in just the same kind of way as ‘*Sue is tall*’ is, except that ‘*very*’ settles the

¹³This does not suffice to characterize the sentence’s meaning, since the sentence also conveys that Sue *is* a 4 year old. See Solt (2011) for discussion.

comparison class to be made up of individuals classified as ‘*tall*’. A similar idea is implemented in a probabilistic framework in Bennett and Goodman (2018). Bennett and Goodman use the Bayesian approach to positive form interpretation of Lassiter and Goodman (2017), which could be seen as a quantitatively explicit model of how *standing out* determines a θ threshold. Bennett and Goodman’s model takes intensifying adverbs to not contribute any actual meaning. Instead, they merely add to the cost of making the utterance. The pragmatic effect of that added cost is a narrower usage, corresponding to higher degrees of the property expressed by the adjective. Once more then, the modifier is seen as not altering the *way* the adjective is interpreted, it merely fills in a parameter of that interpretation.

For the cases we have looked at above, a claim that the lexical content of an intensifier is void does not make much sense. As I have shown above, the evaluative predicate that forms the stem of an unbleached intensifying adverb has a clear semantic effect. Moreover, this predicate (and in particular, its evaluative polarity) also determines the scalar extent of the intensifier, even when the adverb is bleached. Nevertheless, I think that the idea that modifications of positive form adjectives are often modifications of the *standing out* interpretation is correct for evaluative intensifiers, too. In fact, I think that the reason why ‘*horribly warm*’ means that it is *warmer* and not *colder* than what is normally considered to be pleasant is exactly because something is only ‘*horribly warm*’ when it stands out w.r.t. how warm it is: $\mu_{\text{warm}}(x) > \theta$. Extremely cold days *are* horrible, but they do not stand out on the warm scale, since they do not exceed θ .

To implement this, I will start with Wheeler’s semantics. I take the proposition in (19) to express what Wheeler had in mind for (26).

(19) Today is *horribly* / *pleasantly* warm.

(20) $\mu_{\text{horribly/pleasantly}}(\text{that today is as warm as it is}) > \theta'$

Note that (20) is a positive form interpretation of the adverb: that today is as warm as it is stands out w.r.t. how horrific / pleasant that is.

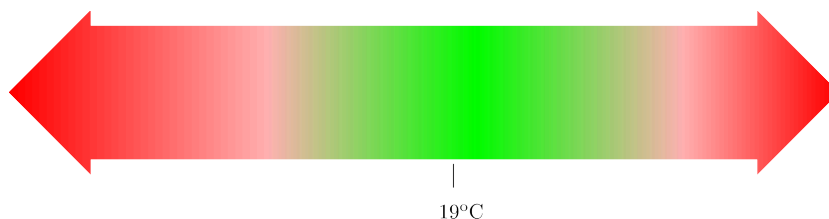
My proposal is now that (20) is only part of the interpretation of (19) and that the adjective in (19) still receives a positive form interpretation – which it does not in (20). That is, ‘*Today is horribly / pleasantly warm*’ is simply interpreted as (21), where θ is to be determined in part on the basis of the truth of (20). In other words, an evaluative intensifier contributes a proposition like (20) and by doing so influences the way in which the subject is said to be *standing out*.

(21) $\mu_{\text{warm}}(\text{today}) > \theta$

One way to think about the proposal is that the intensifier restricts the domain under consideration. Take (22) as an example.

(22) Today is *horrifically* warm

Let's depict the underlying scale as follows, where the green area indicates pleasant temperatures and the red unpleasant ones.



The sentence in (22) conveys both that (i) it is horrific that it is as warm as it is and (ii) that today stands out among days that satisfy (i). What kind of day does this sentence describe? According to (i), the day in question is either very cold or very warm. In other words, the adverbs restricts the scale to days with extreme temperatures:



The sentence expresses that today's temperature exceeds θ , as in (21), so we take this restricted context and look for an appropriate θ , one which makes days that make the sentence true *stand out*. Clearly, the extremely warm days stand out from the extremely cold ones. Since in positive form interpretation standing out is interpreted as *exceeding* a threshold, it is only the extremely warm days that can be described to be '*horrifically warm*'. In contrast, to be '*horrifically cold*' is to be extremely cold, since extremely cold days stand out because the measure of how cold they are exceeds a threshold.

An M-adverb works in the same way. '*Today is pleasantly warm*' conveys that (i) it is pleasant that it is as warm as it is and (ii) that today stands out among the days that satisfy (i). The restricted scale looks like this:



The result is that a pleasantly warm day is a day on the warm end of days that are pleasant because of the temperature. This is entirely in line with our intuitions.

What is attractive about this proposal is that there is no direct manipulation of the degree to which it is warm. That is, the scalar extent attributed to ‘*pleasantly*’ and ‘*horribly*’ comes about entirely by an inference based on the fact that our horror/pleasure at the current temperature *stands out*. As such, this approach is capable of accounting for the observations we made above. Scalar extent is tied to the content of the adverb and in particular to the evaluative polarity that its underlying predicate expresses.

7 Conclusion

I have shown that evaluative polarity is relevant for the denotative meanings of degree semantics. I have given two examples of this. On the one hand, the fact that negative evaluation involves excess makes it that negative predicates can turn into adverbs of high degree, while adverbs formed by positive predicates express medium scalar extent. On the other hand, I showed that evaluatively negative adjectives inherit a scalar property of excess in being lower-bounded.

I started this work by noting that connotative meaning like evaluative polarity is not often considered in formal linguistics. This is probably because other notions of polarity seem to impact more straightforwardly on denotative semantics, such as the presence or absence of a negation marker (‘*not*’, ‘*un-*’, etc.) or lexical antonymity. In contrast to these polar phenomena, evaluation is a much more illusive notion. The evaluative polarity of an adjective is not morpho-syntactically determined. I have given examples of adjectives like ‘*warm*’ and ‘*cold*’ that illustrate that they can receive both negative and positive readings, but even cases that appear to be much more straightforward, like for instance ‘*nice*’, turn out to be quite flexible. An example like (23) will readily be interpreted as saying something relatively *bad* about John.

(23) John is a bit *nice*.

And, so, here we see an example where the scalar requirements in the formal semantics of the adverb flip the connotative meaning of the adjective. This shows once more that connotation and denotation are not as irrelevant to one-another as is perhaps often thought.

References

- Bäcklund, U. (1973). *The collocation of adverbs of degree in english*. Acta Universitatis Upsaliensis.
- Bennett, E. D., & Goodman, N. D. (2018). Extremely costly intensifiers are stronger than quite costly ones. *Cognition*, 178, 147–161.
- Biedermann, R. (1969). *Die deutschen gradadverbien*. Universität Heidelberg.
- Bolinger, D. (1972). *Degree words*. Den Haag: Mouton.
- Bylina, L. (2014). *The grammar of standards: Judge-dependence, purpose-relativity, and comparison classes in degree constructions*. Netherlands Graduate School of Linguistics.
- Bylina, L., & Zadorozhny, S. (2012). Evaluative adjectives, scale structure, and ways of being polite. In (pp. 133–142). Springer.
- Fara, D. G. (2000). Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics*, 20, 45–81.
- Ganesan, K., & Zhai, C. (2012). Opinion-based entity ranking. *Information retrieval*, 15(2), 116–150.
- Gary, E. N. (1979). *Extent in english: A unified account of degree and quantity* (Unpublished doctoral dissertation). University of California, Los Angeles.
- Jaspers, D., van Craenenbroeck, J., & Wyngaerd, G. V. (2016). De syntaxis en semantiek van diminishers in het nederlands en het Duits. *Leuvense Bijdragen: Tijdschrift voor Germaanse Filologie*, 99–100, 87–108.
- Kasting, J., Toon, O., & Pollack, J. (1988). How climate evolved on the terrestrial planets. *Scientific American*(2), 90–97.
- Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable predicates. *Linguistics and Philosophy*, 30(1), 1–45.
- Kennedy, C., & McNally, L. (2005). Scale structure, degree modification and the semantics of gradable predicates. *Language*, 81(2), 345–381.
- Klein, H. (1998). *Adverbs of degree in dutch and related languages* (Unpublished doctoral dissertation). Rijksuniversiteit Groningen.
- Lassiter, D., & Goodman, N. (2017). Adjectival vagueness in a bayesian model of interpretation. *Synthese*, 194(10), 1–36.
- Lehrer, A. (1985). Markedness and antonymy. *Journal of linguistics*, 21(2), 397–429.
- Mohammad, S. M. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the annual conference of the association for computational linguistics (acl)*. Melbourne, Australia.

- Morzycki, M. (2008). Adverbial modification of adjectives: Evaluatives and a little beyond. In J. Dölling & T. Heyde-Zybatow (Eds.), *Event structures in linguistic form and interpretation*. Berlin: Mouton de Gruyter.
- Nouwen, R. (2018a). A note on positive and negative evaluative adjectives. In C. et al. (Ed.), *A coat of many colours*. <https://dj60.be/>.
- Nouwen, R. (2018b). A remark on conceptual metaphor and scalarity. *Unpublished manuscript*. Retrieved from <https://ricknouwen.org/>
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning* (No. 47). University of Illinois press.
- Paradis, C. (1997). *Degree modifiers of adjectives in spoken british english*. Lund University Press.
- Paradis, C., Van de Weijer, J., Willners, C., & Lindgren, M. (2012). Evaluative polarity of antonyms. *Lingue e linguaggio*, 11(2), 199–214.
- Partee, B. H. (2004). Reflections of a formal semanticist. *B. Partee, Compositionality in Formal Semantics: Selected Papers by Barara H. Partee*.
- Raffman, D. (2005). Borderline cases and bivalence. *The Philosophical Review*, 114(1), 1–31.
- Rett, J. (2014). *The semantics of evaluativity*. OUP Oxford.
- Rissanen, M. (2008). From ‘quickly’ to ‘fairly’: On the history of rather. *English Language & Linguistics*, 12(2), 345–359.
- Solt, S. (2011). Notes on the comparison class. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *Vagueness in communication*. Berlin: Springer. doi: 10.1007/978-3-642-18446-8-11
- von Stechow, A. (1984a). Comparing semantic theories of comparison. *Journal of Semantics*, 3, 1–77.
- von Stechow, A. (1984b). My reaction to cresswell’s, hellan’s, hoeksema’s and seuren’s comments. *Journal of Semantics*, 3.
- Stoffel, J. (1901). *Intensives and down-toners*. Heidelberg: Carl Winter.
- Van Os, C. (1989). *Aspekte der intensivierung im deutschen*. Narr Tübingen.
- Wheeler, S. (1972). Attributives and their modifiers. *Nous*, 6, 310–334.