Evaluative adjectives, scale structure, and ways of being polite

Lisa Bylinina[1] and Stas Zadorozhny[2]

[1] Institute for Linguistics OTS, Utrecht
e.g.bylinina@uu.nl
[2] Yandex LLC
zador@yandex-team.ru

1 Introduction

This study is an attempt to add new facts (and new conclusions one could draw from them) to an old and settled topic. The evidence we will be using is of a quantitative sort, based on a corpus study. However, there will be nothing particularly quantitative or corpus-based in conclusions we draw from the data, as well as the semantic decisions we make given the results of the study.

At the same time the quantitative side of this work is of particular importance. This is an attempt to conduct a pilot corpus study from scratch – it includes collecting a corpus, basic pre-processing, calculating raw statistics, and formulating and testing certain types of linguistic hypotheses. This pipeline can be useful beyond the topic of this particular study. It does not require access to paid resources and allows to control various parameters to fit the needs of the study, and yet it is not very demanding technically. We hope to reuse and improve this set-up for other studies, the current paper being an example of the possible ways to use this pipeline. The overall goal is to make quantitative semantic tasks solved in more systematic way.

The topic we investigate here is the semantics of evaluative adjectives (EAs) – both positive (*charming*, *industrious* etc.) and negative (*lazy*, *ugly* etc.). (Bierwisch 1989) describes EAs as a subclass of gradable adjectives along with 'dimensional adjectives' (DAs) like *tall*, *short*, *big*, *small* etc.

EA vs. DA distinction is arguably not a clear-cut one, but there are several reasons for still making this distinction. We discuss them in further detail below, but they boil down to EAs and DAs referring to their underlying scales in different ways. (Bierwisch 1989) is still the only comprehensive study of how EAs and DAs are different, and no new substantial facts were added to Bierwisch's description since 1989.

We concentrate on further distinctions within the EA class – namely, how positive EAs are different from negative EAs in terms of scale structure. (Bierwisch 1989) treats them in exactly the same way, but observe the following contrast:

(1) a. Clyde is slightly stupid.
b. Clyde is slightly lazy.

(2) a. ??Clyde is slightly smart.
b. ??Clyde is slightly industrious.

We will explore this and related contrasts further and see what they suggest. The study is limited to several degree modifiers and a pre-selected list of EA and DA adjectives.

Section 2 sets the theoretic stage for the study and formulates the problem; section 3 describes the corpus study. The results will be made sense of in section 4, and section 5 concludes.

2 Evaluative adjectives (Bierwisch 1989)

2.1 Gradable adjectives: background and conventions

For our purposes we do not need to takes sides in the ongoing debates about nature of vagueness and gradability in natural language. We do not find it crucial for the case at hand to, say, choose between degree and delineation semantics for vagueness.

We will assume degree semantics of gradability for no particular reason. Here is an implementation of this view that analyzes gradable adjectives as measure functions: functions of type $\langle e, d \rangle$ from the domain of individuals to degrees on a certain scale (Bartsch and Vennemann 1973, Kennedy 1997, 2007):

(3) $||$ tall $|| = \lambda x.$ **tall** (x)
where **adj**(x) is 'the degree on the appropriate scale that represents x's measure of adjective-ness'

Measure functions are converted into properties of individuals by degree morphology (comparative morphemes, intensifiers etc.). For the unmarked positive form (*John is tall*) a null POS morpheme is introduced, with a denotation along the lines of (4), where $\mathbf{d}_s$ is 'contextually appropriate standard of comparison, whatever that is' Kennedy (2007):

(4) $||_{Deg}$ POS $|| = \lambda g \lambda x. g(x) \succeq \mathbf{d}_s$

2.2 Scale structure and degree modifiers

Under any view on vagueness/gradability, gradable adjectives are used for reasoning about domains that are or can be structured in a particular way, namely, (partially) ordered with respect to a certain dimension. For many purposes it would suffice to deal with the orderings on the domains of individuals and not involve a separate notion of scales as distinct objects, c.f. (Bale 2008). However, there are arguably facts that are not easy to capture if one does not assume a richer notion of scales. The obvious candidate is the absolute/relative distinction within the class of gradable adjectives (but see (van Rooij 2010)). This distinction manifests itself in a number of ways, for example, degree modifier distribution:

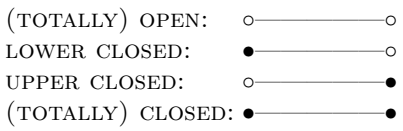
(5) a. ??perfectly/??slightly {tall, deep, expensive, likely}
b. ??perfectly/??slightly {short, shallow, inexpensive, unlikely}

(6) a. ??perfectly/slightly {bent, bumpy, dirty, worried}
b. perfectly/??slightly {straight, flat, clean, unworried}

(7) a. perfectly/??slightly {certain, safe, pure, accurate}
b. ??perfectly/slightly {uncertain, dangerous, impure, inaccurate}

(8) a. perfectly/slightly {full, open, opaque}
b. perfectly/slightly {empty, closed, transparent}

These contrasts have been argued to signal the differences in structure of scales associated with different gradable adjectives, sketched below (Rothstein and Winter 2004, Kennedy and McNally 2005, Kennedy 2007):



Adjectives that make use of a scale that is closed at least at one end are absolute adjectives, adjectives with totally open scales are relative ones. There have been several attempts to give an explanation of why distribution of degree modifiers needs to correlate with existence of a scale bound; some of them challenge the general picture presented above (Sassoon and Toledo 2011, Solt 2011, McNally 2011).

To give the simplest possible insight about why low degree modifiers like *slightly* are only possible with scales with a minimum, it would suffice to say that the positive form of relative adjectives requires of an entity to **significantly** exceed a contextual standard, i.e. **stand out** in a relevant respect; positive forms of absolute adjectives, on the other hand, do not include the significance component. We illustrate this by directly including **significantly** into semantics of POS_{rel} combining with relative adjectives vs., say, POS_{min} combining with adjectives with lower bound scales (Kennedy 2007):

(9) a. $|| \text{POS}_{rel} || = \lambda g \lambda c \in D_{\langle e,t \rangle} \lambda x.g(x) \not\succ \text{norm}\ (c)(g)$
c = comparison class, g = gradable property, $\not\succ$ = significantly exceed
b. $|| \text{POS}_{min} || = \lambda g \lambda x.g(x) \succ \textbf{min}(\text{SCALE}(g))$

Thus, relative standards cannot be addressed when reasoning about small differences, while absolute standards can. Whatever the right compositional analysis of degree modifiers is, this observation needs to hold for them as well as for POS.

In addition, *very* is usually taken to specifically target standards of relative adjectives: 'in normal usage, absolute adjectives reject modification by *very*' (Kennedy and McNally 2005, 370):

(10) a. ??I always leave the door to my office very open.
b. ??That drug is currently very available.

2.3 Dimensional vs. evaluative adjectives

The distinction between evaluative (EAs) and dimensional adjectives (DAs) in (Bierwisch 1989) is motivated by two observations: 1) the antonymous pairs of EAs have a less obvious relation to each other than DA antonu spairs; sometimes it is hard to tell whether the two given EA items form a pair or not; 2) inference judgements for EAs vary a great deal inter- and intra-individually.

The intuition behind observation 1 is the following: '*Hans ist klein* (Hans is short) assigns to Hans a certain degree of height, while *Hans ist faul* (Hans is lazy) does not mean that Hans has a certain degree of industriousness. Put somewhat differently, even a negative DA always specifies a positive value on the scale of its antonym, whereas this does not apply to a negative EA: even a short person has height, but a lazy person cannot be to any extent industrious' (Bierwisch 1989, 88). Here is a more theoretic way to formulate the same thought:

(11) (Bierwisch 1989, ix)
Antonymous DAs refer to the same scale of a given dimension and differ in the ordering on the scale, antonymous EAs refer to different scales or parts of scales.

Thus it is not obvious that every EA has a single antonym, at least in the sense that DAs do:

(12) DAs:
a. tall ↔ short
b. heavy ↔ light
c. hot ↔ cold

(13) EAs:
a. brave, bold, courageous ↔ cowardly, timid, fearful
b. clever, bright, shrewd, intelligent, brilliant ↔ stupid, idiotic, foolish
c. pretty, beautiful, gorgeous, handsome ↔ ugly, hideous, grotesque
d. lazy, indolent, unproductive ↔ hard-working, industrious, workaholic

The second observation concerns the following kinds of inference patterns and the fact that speakers do not have a consensus on whether these inferences hold:

(14) a. How hard-working/lazy is Hans?
→ Hans is hard-working/lazy.
b. Hans is as hard-working/lazy as Eva
→ Hans is hard-working/lazy.

c. Hans is more hard-working/lazier than Eva
   → Hans is hard-working/lazy.

d. Hans is the most hard-working/laziest
   → Hans is hard-working/lazy.

e. Hans is too hard-working/lazy for that
   → Hans is hard-working/lazy.

f. Hans is hard working/lazy enough for that
   → Hans is hard-working/lazy.

g. Hans is less hard-working/lazy than Eva
   → Hans is hard-working/lazy.

This motivates a conclusion made in (Bierwisch 1989) that EAs are underlyingly not gradable at all, though they **can** used as gradable with the help of a type shift or coercion that results in a scale with a derived zero (=lower-bound). (Morzycki 2011) notes that this solution is 'immediately worrying, since non-dimensional adjectives straightforwardly form comparatives and occur with degree modifiers.' Our main goal is not to evaluate the idea of coercion from non-gradable to gradable per se, but to check whether the conclusion Bierwisch makes is a fully informed one.

The most important point in this solution for us is that positive and negative EAs are treated the same and are ultimately assigned the same scale structure. However, as we have seen in the introduction, there are contrasts between positive and negative EAs that look like differences in scale structure (1-2).

There is a clear need for a more wide-ranged study, first of all, to be able to make a claim about the data regarding low degree modification in EAs, and then see whether the data points in a direction of a particular analysis.

3 Corpus study

To collect the corpus we used BootCaT toolkit (Baroni and Bernardini 2004) – software which allows to bootstrap corpora from web with a small initial set of seed words as its input. BootCat uses these seeds to query search engine (Google, Yahoo) and download urls from search results page to build initial corpora. These initial corpora are then used for expansion of seed words. The procedure repeats iteratively until the desired number of documents is reached.

Manipulating initial seed words is a way to control various parameters of the corpus, such as the language, genre or topic of downloaded documents. This is an extremely useful feature, though we did not make the best of it in the current study. In our work we used the general list of most frequent English words to obtain general corpora without bias towards specific genre or topic.

As result of applying the BootCaT procedure we have collected a corpus consisting of 156117 unique documents. These data were normalized (lowercasing, character trash removal etc.) and split into sentences using simple heuristics and set of standard UNIX tools. Then we collected basic statistics for bigrams (all pairs of consecutive words) in sentences – bigram frequency and marginal

frequency (the frequency of first or second component of a bigram among all pairs).

We were interested in particular events – namely, co-occurrence of certain degree modifiers with certain adjectives. The modifiers we are interested in are low degree modifiers *a bit*, *a little bit*, *slightly* and *somewhat* (=LOW), and a standard-booster *very* (=VERY). The adjectives we are interested in are, first of all, positive EAs (=POS) and negative EAs (=NEG). Also we use data from relative adjectives (=REL) and lower-bound adjectives (=MIN). We draw the lists of adjectives from various sources – theoretic papers discussing gradability and scale structure (Kennedy and McNally 2005, Kennedy 1997, 2007, Rothstein and Winter 2004, Sassoon and Toledo 2011, McNally 2011) as well as works that include quantitative studies (Solt 2011, Sassoon 2011), and also various descriptive grammars of English that introduce semantic divisions within adjectives. We thank Stephanie Solt, Galit Sassoon and Chris Kennedy for sharing their lists and thoughts on possible ways to extend the lists we are using.

So we collect only events of the form $\langle x, y \mid x \in \{$ LOW, VERY $\}, y \in \{$ POS, NEG, MIN, REL $\}\rangle$. For each of the events we calculate the following frequencies: $f(x, y), f(x, *), f(*, y), f(*, *)$, where $*$ stands for any word. Random sample of collected data:

x	*y*	*f(x,y)*	*f(x, *)*	*f(*, y)*
abit	responsible	2	17582	25617
abit	black	1	17582	54527
abit	bright	2	17582	8687
very	creepy	12	194306	669
very	brave	58	194306	2058
very	helpful	1515	194306	18699
very	angry	230	194306	4980
very	funny	231	194306	7197
somewhat	protective	2	10091	4039
very	inexpensive	131	194306	2749

Cross-tabulation of data:

class/mod	NEG	POS	MIN	REL	total
a bit	350	92	79	218	739
a little bit	66	12	8	19	105
slightly	114	42	202	57	415
somewhat	212	191	101	98	602
LOW	742	337	390	392	1861
VERY	3032	21130	1299	15091	40522
total	3774	21467	1689	5483	

We also filtered out events with frequency below 3 due to unreliability of probability estimates for low-frequency events, as discussed in (Evert 2005).

Our particular interest is how certain classes of modifiers are related to certain classes of gradable adjectives. The main goal of the paper is to know more

about scale structure of positive and/vs. negative EAs, so the obvious pairs of distributions to compare are the following:

$$P(x = low, y = neg) \text{ vs. } P(x = low, y = pos) \quad (1)$$

$$P(x = very, y = pos) \text{ vs. } P(x = very, y = neg) \quad (2)$$

As we are using quite a lot of theoretic assumptions on how degree modifiers are sensitive to the scale structure, it would also be good to check these assumptions against our data before drawing any conclusions from results on (1) and (2). Contrasting the following pairs of distributions is aimed at proving these assumptions right or wrong on our data:

$$P(x = low, y = min) \text{ vs. } P(x = low, y = rel) \quad (3)$$

$$P(x = very, y = rel) \text{ vs. } P(x = very, y = min) \quad (4)$$

Finally, we try to see whether there are reasons to group positive or negative EAs with lower-bound or relative DAs based on co-occurrences with low degree modifiers:

$$P(x = low, y = neg) \text{ vs. } P(x = low, y = min) \quad (5)$$

$$P(x = low, y = pos) \text{ vs. } P(x = low, y = rel) \quad (6)$$

$$P(x = low, y = neg) \text{ vs. } P(x = low, y = rel) \quad (7)$$

$$P(x = low, y = min) \text{ vs. } P(x = low, y = pos) \quad (8)$$

The same for *very*:

$$P(x = very, y = neg) \text{ vs. } P(x = very, y = min) \quad (9)$$

$$P(x = very, y = rel) \text{ vs. } P(x = very, y = pos) \quad (10)$$

$$P(x = very, y = rel) \text{ vs. } P(x = very, y = neg) \quad (11)$$

$$P(x = very, y = min) \text{ vs. } P(x = very, y = pos) \quad (12)$$

First of all, we should establish a measure of relatedness between a modifier and an adjective and then compare the resulting distributions. We use pointwise mutual information (PMI) as a measure that was created precisely for these purposes (Fano 1961, Church and Hanks 1990):

$$PMI(x,y) = log\frac{P(x,y)}{P(x)P(y)} \tag{13}$$

In our case $P(x,y)$ is the probability of a bigram, $P(x)$ is the probability of a modifier preceding any word and $P(y)$ is the probability of an adjective following any word. Informally, mutual information compares the probability of observing x and y together (the joint probability) with the probabilities of observing x and y independently (chance). If there is a genuine association between x and y, then the joint probability *P(x,y)* will be much larger than chance *P(x)P(y)*. Thus, the large positive values of PMI indicate stronger association between events; if it takes zero value, then x and y are independent; negative values indicate that x and y are in complementary distribution.

We can estimate the probabilities required for calculation of PMI using maximum-likelihood estimation and previously collected bigram frequencies:

$$P(x,y) = \frac{f(x,y)}{f(*,*)} \tag{14}$$

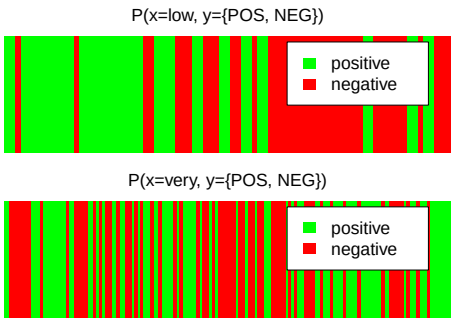
$$P(x) = \frac{f(x,*)}{f(*,*)} \tag{15}$$

$$P(y) = \frac{f(*,y)}{f(*,*)} \tag{16}$$

So maximum-likelihood estimation of PMI (13) will be:

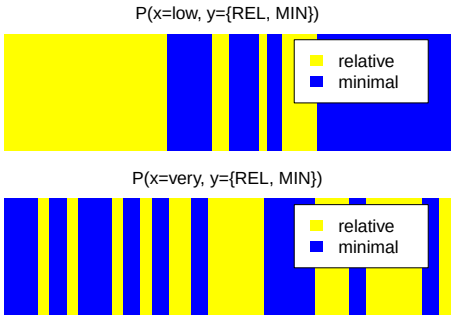
$$\overline{PMI}(x,y) = log\frac{f(x,y)f(*,*)}{f(x,*)f(*,y)} \tag{17}$$

To illustrate possible differences in distribution, we plot pairs of groups (1-12) by ordering them by PMI. Each color represents a class of adjectives and PMI value for each adjective in two groups increased from left to right.

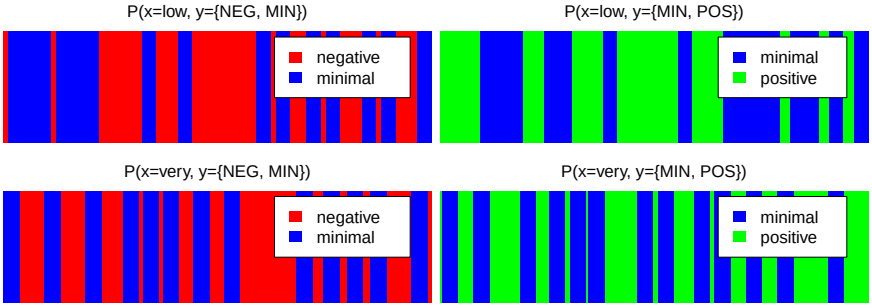


These two figures display distributions of positive and negative EAs with respect to their PMI with low degree modifiers (the first figure) and *very* (the second figure). Positive EAs (*beautiful*, *smart* etc.) are green, negative EAs (*ugly*, *stupid* etc.) are red. The figures visualize that negative EAs seem to group to the right of the low degree modifier figure (which speaks of more association), while the *very* picture looks free of any obvious tendency.

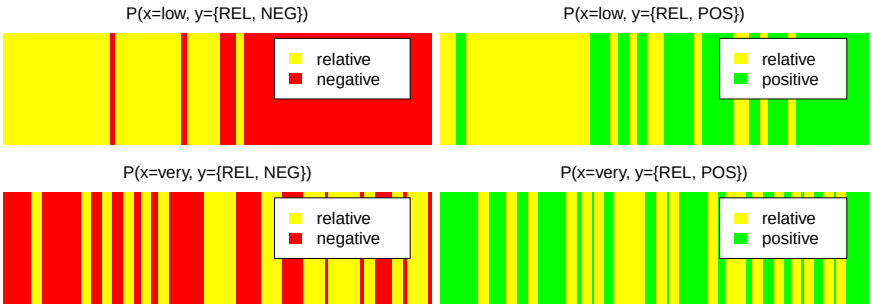
This pair illustrates how well our data support the theoretic assumptions about scale structure sensitivity of degree modifiers. It is rather obvious from these figures that lower-bound adjectives (painted blue) have stronger association with low degree modifiers than relative adjectives do (painted yellow). However, the picture with *very* is not obvious and needs more attention. Theory predicts it to be a mirror image of the low degree modifier picture, but it is not.



Including the following four figures is motivated by the suggestion made in (Bierwisch 1989) that both positive (painted green) and negative EAs (painted red) are structurally similar to lower-bound adjectives (painted blue). It is not obvious from these pictures alone whether there is any tendency, but it can be seen rather well that these three classes are not strikingly different from each other:



Comparing negative and positive EAs in the same way to relative adjectives (yellow) gives quite different results. Again, *very* is less informative than we would expect it to be, but there is a clear tendency for both positive and negative EAs to co-occur with low degree modifiers much more often than relative adjectives do:



To qualitatively asses the observed differences, we performed series of statistical tests. We used Wilcoxon rank-sum test (MWW) (Mann and Whitney 1947, Hollander and Wolfe 1999), which allows assessing whether the two groups have a tendency to have different values (two-sided test) or whether one of the two groups tends to have greater values then the other group (one-sided test). The choice of this test was motivated by unknown distribution of samples (MWW is a non-parametric test).

For each pair of distributions (1-12) we formulate the following null hypothesis and an alternative hypothesis:

(15) a. H_0: There is no difference in association score between two compared classes of adjectives with respect to a given modifier

b. H_1: There is a significant difference in association score between two given classes of adjectives with respect to a given modifier

Based on our expectations from what theories of gradable adjectives predict, for some pairs we also formulate a different alternative hypothesis:

(16) H_1: The association score tends to be larger for one class then for the other.

The following table represents results of MWW statistical test for distributions (1-12) with significance level 0.01 and frequency threshold 2 for low degree modifiers:

H_1	accept/decline	p-value	W
$NEG \neq POS$	accept	3.62e-05	1924
$NEG > POS$	accept	1.81e-05	1924
$MIN \neq REL$	accept	1.35e-06	707
$MIN > REL$	accept	6.8e-07	707
$NEG \neq MIN$	decline	0.10503787	1194
$NEG > MIN$	decline	0.05251893	1194
$POS \neq REL$	accept	5.25e-06	879
$POS > REL$	accept	2.62e-06	879
$NEG \neq REL$	accept	0	2863
$NEG > REL$	accept	0	2863
$MIN \neq POS$	decline	0.15485217	425
$MIN > POS$	decline	0.07742609	425

Results for *very* with the same settings:

H_1	accept/decline	p-value	W
$POS \neq NEG$	decline	0.0534381	4634
$POS > NEG$	decline	0.02671905	4634
$REL \neq MIN$	decline	0.09397819	176
$REL > MIN$	decline	0.0469891	176
$NEG \neq MIN$	decline	0.73924513	471
$NEG > MIN$	decline	0.63470126	471
$REL \neq POS$	decline	0.13602659	1248
$REL > POS$	decline	0.0680133	1248
$REL \neq NEG$	accept	0.00339799	1099
$REL > NEG$	accept	0.00169899	1099
$MIN \neq POS$	decline	0.43602203	580
$MIN > POS$	decline	0.7845581	580

Results shows that we should reject statement about same distributions and may accept alternative hypothesis about significant difference for pairs (1, 2, 4, 5, 11).

Described procedure has potential problems that are worth mentioning. The bigram set can contain bigrams that are not actually one syntactic unit but belong to different constituents, as we used raw text only, without any syntactic information.

4 Interpreting the results

This section evaluates the existing theories of gradable adjectives and their scale structure against the results of the corpus study presented in the previous section. We formulate expectations of the standard theories and sum up whether they are borne out by our study in the following list:

Expectation 1 *Very* goes with relative adjectives and not with absolute ones → **Wrong**

Expectation 2 Low degree modifiers go with lower-bound adjectives and not with relative ones → **Right**

Expectation 3 Neither POS nor NEG EAs go with *very* as well as relative adjectives do → **Wrong**

Expectation 4 Both POS and NEG EAs go with low degree modifiers, quite like lower-bound adjectives do → **Right**

Expectation 5 There is no difference between POS and NEG EAs in how well they go with low degree modifiers → **Wrong**

Expectation 6 There is no difference between POS and NEG EAs in how well they go with *very* → **Right**

Expectation 5 comes from (Bierwisch 1989) and is actually in a way expected to be wrong given examples (1-2) that were the initial motivation for our study. With only this fact in mind (now supported by statistics), one would be tempted to say that while negative EAs do have a scale with a minimum, positive EAs do not. When the correct Expectation 6 is added to the picture, the conclusion

to draw from that would be something like the following (we did draw this conclusion in an earlier version of this paper):

Preliminary analysis Positive EAs have a totally open scale; negative EAs have a totally open scale as well, but they can also be interpreted in a different way that involves a scale with a minimum. Positive EAs lack this secondary interpretation.

This analysis crucially relies on what existing theories say about the way different classes of adjectives co-occur with different classes of degree modifiers. However, if we include relative adjectives and lower-bound adjectives in our study and compare them to each other and to positive and negative EAs pairwise, the picture changes. First, *very* as a test for relative adjectives does not seem to work at all. Second, when we look at low degree modifiers, we see that original hypothesis from (Bierwisch 1989) seems very close to what we observe. Both positive and negative EAs behave like lower-bound adjectives, and both positive and negative EAs behave differently from relative adjectives. This strongly suggests that all EAs have a scale with a minimum, quite as predicted. But this poses a problem for the observed differences between positive and relative EAs in how well they combine with low degree modifiers. We need to resolve this inconsistency.

Why the positive/negative EA asymmetry? Negative EAs generally express or implicate judge's negative feeling or attitude related to a certain object. As part of a general tendency to attenuate negative judgements, similar to what is described in (Baumeister et al. 2001, Rozin and Royzman 2001) a.m.o., judgements with negative EAs are quite often attenuated. The reason is that, basically, 'bad is stronger than good'. One could think of various ways to make sense of this intuition within a semantic theory of gradability. For instance, one could require of any mapping between positive and negative scales to be mediated by a certain coefficient, so that an interval on a positive scale would correspond to a bigger interval on a negative one. However this is implemented, the result should be that positive scales are in some sense more dense than negative ones, and this is a property of scales that is different from having or lacking a minimum or a maximum. We can now reformulate our analysis:

Analysis All EAs have lower-closed scales, but positive EAs have more dense scales than negative ones.

5 Conclusion

We conducted a corpus study of evaluative adjectives, both positive and negative. The particular goal of this study was to establish their scale structure and how positive and negative EAs are different from each other. As a result, we found that the original hypothesis of (Bierwisch 1989) is supported by our data, but there are still contrasts that are not captured by this hypothesis. We believe that pragmatic factors play a role, namely, that negative judgements have more 'strength' than positive ones, which explains massive low degree modification of negative but not positive EAs.

Bibliography

Bale, A.: 2008, A universal scale of comparison, *Linguistics and Philosophy* **31**(1), 1–55.

Baroni, M. and Bernardini, S.: 2004, Bootcat: Bootstrapping corpora and terms from the web, *Proceedings of LREC 2004*, pp. 1313–1316.

Bartsch, R. and Vennemann, T.: 1973, *Semantic Structures: A Study in the Relation between Syntax and Semantics*, Frankfurt.

Baumeister, R., Bratslavsky, E., Finkenauer, C. and Vohs, K.: 2001, Bad is stronger than good, *Review of General Psychology* **5**, 323–370.

Bierwisch, M.: 1989, The semantics of gradation, *in* M. Bierwisch and E. Lang (eds), *Dimensional Adjectives: Grammatical Structure and Conceptual Interpretation*, Springer-Verlag, Berlin, pp. 71–261.

Church, K. W. and Hanks, P.: 1990, Word association norms, mutual information, and lexicography, *Computational Linguistics* **16**, 22–29.

Evert, S.: 2005, *The statistics of word cooccurrences : word pairs and collocations*, PhD thesis, Universitt Stuttgart, Stuttgart.

Fano, R.: 1961, *Transmission of Information: A Statistical Theory of Communications*, The MIT Press, Cambridge, MA.

Hollander, M. and Wolfe, D.: 1999, *Nonparametric statistical methods*, Wiley series in probability and statistics: Texts and references section, Wiley.

Kennedy, C.: 1997, *Projecting the adjective: the syntax and semantics of gradability and comparison*, PhD thesis, UCSD.

Kennedy, C.: 2007, Vagueness and grammar: The semantics of relative and absolute gradable predicates, *Linguistics and Philosophy* **30**, 1–45.

Kennedy, C. and McNally, L.: 2005, Scale structure, degree modification, and the semantics of gradable predicates, *Language* **81**(2), 345–381.

Mann, H. and Whitney, D.: 1947, On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Stat.* **18**, 50–60.

McNally, L.: 2011, The relative role of property type and scale structure in explaining the behaviour of gradable adjectives, *in* R. e. a. Nouwen (ed.), *Papers from the ESSLLI 2009 Workshop on Vagueness in Communication*, Vol. 6517 of *Lecture Notes in Artificial Intelligence*, Springer, Berlin, pp. 151–168.

Morzycki, M.: 2011, The several faces of adnominal degree modification, *Proceedings of the 29th West Coast Conference on Formal Linguistics (WCCFL)*, Cascadilla Press, Somerville, Mass.

Rothstein, S. and Winter, Y.: 2004, Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers, *Natural Language Semantics* **12**, 259–288.

Rozin, P. and Royzman, E.: 2001, Negativity bias, negativity dominance, and contagion, *Personality and Social Psychology Review* **5**, 296–320.

Sassoon, G.: 2011, A slightly modified economy principle: Stable properties have non-stable standards, *Sinn und Bedeutung 16*, Utrecht.

Sassoon, G. and Toledo, A.: 2011, Absolute vs. relative adjectives: Variance within vs. between individuals, *Proceedings of SALT21*.

Solt, S.: 2011, Comparison to fuzzy standards, *Sinn und Bedeutung 16*, Utrecht.

van Rooij, R.: 2010, Vagueness and linguistics, *in* G. Ronzitti (ed.), *The Vagueness Handbook*, Springer.