Tests of scale structure theory in dimensional and multidimensional adjectives

Galit Weidman Sassoon
[1] Bar Ilan University, Ramat Gan, Israel.
galitadar@gmail.com

Abstract

This paper addresses the need to pay attention to the multiplicity of possible interpretations of adjectives when applying to them the standard tests of scale structure and standard (Kennedy & McNally 2005). In particular, the paper considers simple-dimensional, complex-dimensional, and multidimensional interpretations of multidimensional adjectives (Sassoon in progress).

1 Introduction

Standard theories of gradability associate gradable adjectives with a single scalar dimension per context, like height or health (Bierwisch, 1989; Kennedy, 2007; Kennedy & McNally, 2005; Rotstein & Winter, 2004). Often, the dimension is used to compose a relation between individuals and degrees – a denotation at type <d, <e,t>>. For example, *tall* denotes a relation between degrees d and entities x which are at least d tall. A null morpheme, *pos*, introduces a membership norm c called *standard* into the logical form and truth conditions of positive forms. For example, (1) is true iff Ann is at least as tall as the norm (von Stechow, 2007).

(1) Ann is pos_c tall.

Degrees can reflect a single measurement, like height (a basic dimension), or the output of a function over degrees in a set of measurements, like f_F with weights w_F $f(f_{F1}...f_{Fn}, w_{F1}...w_{Fn})$ (a complex dimension; Bylinina, 2013; Kennedy, 2013; McNally & Stojanovic, 2014; Umbach, 2016; Solt, 2018). For example, the optimism of an entity x can be modeled as the *weighted sum* $\mathbf{\sum w_F f_F(x)}$ of x's degrees in various measurements, and x's health can be modeled as a *weighted product* $\mathbf{\Pi w_F f_F(x)}$. Weighted products capture the intuition that, for example, any life-threatening disease reduces one's average health below any plausible standard, no matter how healthy one is otherwise. When degrees are multiplied, a low degree in a single dimension strongly reduces the overall product (for example, $1 \cdot ... \cdot 1 \cdot 0.5 = 0.5$) and thus reduces the classification probability. In contrast, when degrees are added, a few low degrees hardly affect the overall sum (for example, $1 + ... + 1 + 0.5$ is almost the maximal sum possible). This is useful to model cases in which the contributions of the different dimensions are independent, as is characteristic of traits like *optimistic* (Murphy 2002; Pothos & Wills 2011).

Indeed, recent work adopts the intuitive view that speakers weigh the dimensions of multidimensional adjectives by importance (Kennedy 2013; McNally & Stojanovic, 2014), and sum up the degree of the entities in those dimensions factored by their weights (Gärdenfors 2004; Bylinina, 2013; Solt, 2018). For example, Kennedy (2013) argues for uncertainty about how the dimensions involved in standard calculation are weighted in different situations. Kennedy thereby explains faultless disagreement in the presence of multiple dimension, namely, the fact that no side can be proven wrong when speakers disagree about whether the application of an adjective like *typical, beautiful* or *safe* to a given entity is truthful or not. Furthermore, as illustrated above, the cognitive literature often uses averaging functions to collapse the set of dimensional degrees and weights into a single degree. While more complex functions than averaging are also possible, this does not affect the arguments in this paper.

Thus, for example, on any of its **simple dimensional interpretations**, *healthy*, denotes a relation, R_F, between entities x and their levels d of health with respect to a contextually given dimension F (e.g., cholesterol, diabetics, flu or chickenpox; Bartsch 1984; Kennedy 2013). By contrast, on its **complex dimensional interpretation**, *healthy* denotes a relation, $R'_{healthy}$, between entities x and their 'averaged' health levels d (e.g., the weighted product, $\mathbf{\Pi w_F f_F(x)}$, of the degrees of entities x in the contextually relevant health indices F). With a maximum standard for *healthy* and a minimum standard for *sick*, entities are predicted to count as *pos healthy* iff they are maximally healthy in every respect, and as *pos sick* otherwise (Bylinina 2013).

However, besides adjectives whose scales are based on simple or complex dimensions of the sort illustrated above, Bartsch & Vennemann (1972) and Bartsch (1984) represented certain adjectives as multidimensional. Sassoon (in progress) and Sassoon & Fadlon (2017) argued that these adjectives have, in addition to any dimensional interpretation, also a multidimensional interpretation, where the scale is based on dimension counting.

For example, on its **multidimensional interpretation**, *healthy* denotes a relation, $R_{healthy}$, between entities x and the number n of dimensions with respect to which they are healthy. In this interpretation, the standard of *healthy* standard represents the minimal number of dimensions with respect to which entities have to be healthy in order to count as *pos healthy*. By virtue of a maximum standard, *pos healthy* conveys being healthy in every respect. *Sick* denotes a relation, R_{sick}, between entities x and the number n of dimensions with respect to which they are sick, and by virtue of a minimum standard, *pos sick* conveys being sick in at least one respect.

The multidimensional interpretation of an adjective differs from its dimensional and complex-dimensional interpretations in—often subtle but—important respects (see discussion in Sassoon in progress). The default multidimensional interpretation of a positive adjective like *safe*, for example, ('safe in every respect') asymmetrically entails any interpretation based on a specific dimension (for example, frequency of cases of robbery or rape). Evidence for its presence is the intuition that a neighborhood can be considered *not pos safe* even if the only piece of information available is that some big enough danger exists, namely there is a respect in which the neighborhood is not safe. No further information is needed about the nature of that danger (the respect being violated). This much information does not suffice to falsify an interpretation of *pos safe* based on a particular dimension like robbery.

By contrast, the multidimensional interpretation of *pos safe* (safe in every respect) is asymmetrically entailed by any **maximum-standard** complex (e.g., averaging-based) interpretation of *pos safe*. The two interpretations differ because a neighborhood which is not maximally safe in every respect (namely, not *pos safe* in the complex-dimension sense) can still be safe in every respect (namely, *pos safe* in the multidimensional sense). Evidence for the multidimensional interpretation comes from the intuition that a neighborhood can be considered *pos safe* even when some degree of danger of some sort exists, for example, robbery occurs but rarely enough that the neighborhood counts as safe in this respect. In general, a multidimensional interpretation is sensitive to the standards of the dimensions and does not necessitate maximal standards (relative adjectives may well constitute dimensions). By contrast, a complex dimensional interpretation is not sensitive to the standards of the dimensions, and, with a maximum standard reduces to a quantificational interpretation like "adjective in every respect", only assuming the dimensions have a maximum standard too.

In sum, intuitively, in some contexts, a neighborhood is considered (*pos*/*perfectly*) *safe* iff the neighborhood is safe in every respect, and *not* (*pos*/*perfectly*) *safe* otherwise. No account in terms of a unique dimensional interpretation (either simple or complex) captures both these truth condition and falsity condition simultaneously (Sassoon, in progress).

Section 2 briefly reviews some of the motivations for dimension-counting interpretations, and clarifies the distinctions between them and other interpretations of adjectives that involve counting scales (Sassoon, in progress). This multiplicity of interpretations gives rise to a need to make the standard tests of scale structure of adjectives more precise. In particular, section 3 reviews some of the scale structure tests. Some tests are based on judgments of inference patterns or contradictions between sentences with a given adjective. While judging whether an inference follows or a contradiction holds, it is important to control for the type of interpretation of each ambiguous or context-sensitive word in the premises and conclusions or in the potentially contradicting sentences. Section 3 suggests that to better understand the results of the tests of scale-structure theory in the presence of counting-based (quantificational) interpretations, the application of a supplementary test, based on exceptive phrases, is needed.

2 Motivations for the dimension-counting hypothesis

Exception phrases indicate universal generalizations as opposed to existence statements, as shown in the contrast in examples (3a,b) vs. (3c,d) (Hoeksema, 1995; Moltmann, 1995; von Fintel, 1994).

(1) a. Everyone arrived except for Mary.
b. No one arrived except for Mary.
c. #Someone arrived except for Mary.
d. #Not everyone arrived except for Mary.

Thus, the higher acceptability of exception phrases in examples like (4a) rather than (4c) seems to stem from a higher tendency to interpret positive forms of positive adjectives like *healthy*

as involving universal quantification over dimensions. In addition, the higher acceptability of exception phrases in examples like (4b) than those like (4d) seems to stem from a higher tendency to interpret positive forms of negative antonyms like *sick* as involving existential quantification over dimensions. Negation reverses the quantificational force, resulting in universal quantification in (4b) and existential quantification in (4d) (Hoeksema, 1995).

(2) a. Mary is **healthy** except for high cholesterol (ch)
b. Mary is **not sick** except for the flu
c.# Mary is **sick** except for normal cholesterol (ch)
d.# Mary is **not healthy** except for (normal) cholesterol (ch)

Judgment studies support the acceptability contrasts indicated in (2) (see review in Sassoon, in progress) and corpus studies reveal distributional patterns reflecting these judgments. Sassoon (2013) considered 1300 naturally occurring examples of the form 'Adj. except' with 8 antonym pairs in positive vs. negated contexts. Positive adjectives manifested mainly interpretations involving universal quantification over dimensions, while negative adjectives manifested mainly interpretations involving existential quantification.

Sassoon (in progress) argued that the basis for these trends is a scale based on dimension counting with a tendency toward a maximum standard in positive adjectives, as opposed to a minimum standard in negative ones. One motivation for this proposal was the observation that comparison constructions may have an interpretation in which dimension-cardinalities are directly compared. For instance, example (3), in addition to having an **access reading**, as in (3a), and a **quantificational reading**, as in (3b), can also have a **dimension-counting reading**, as in (3c).

(3) Ann is more successful than Bill.
a. Ann is more successful than Bill is in some **salient respect** (their math studies).
b. Ann is more successful than Bill is **in n-many (for example, most) respects**.
c. Ann is pos successful **in more respects** than Bill is.

In the dimension-counting proposal, *successful* denotes the dimension-counting relation that holds between degrees d and entities who are successful in at least d respects. Thus, comparisons like (3) can convey that there is a number d, such that Ann is successful in at least d respects, while Bill isn't, namely reading (3c). In addition, in this proposal, *(un)successful* denotes the dimension-counting relation that holds between degrees d and entities who are (un)successful in at least d respects. Thus, (4) is correctly predicted to have a dimension-counting interpretation, conveying that there are more dimensions in which Bill is successful than there are dimensions in which he is unsuccessful (Sassoon, in progress).

(4) Bill is more successful than unsuccessful.

In (5) (from 2009's academic section of the corpus of contemporary American English, Davis 2010), the number of dimensions in which reading and spelling are alike seem to compare to the number of dimensions in which they are different. The contextually supplied dimensions are language skills. Reading and spelling count as similar with respect to a given skill if both require it. Thus, reading and spelling count here as similar in the multidimensional sense of the dimension-counting proposal because (and to the extent that) they require the same skills.

(5) "Reading and spelling require the same language skills (Moats, 2005), have a strong correlation (Ehri, 2000), and support the development of each other (Snow, Griffin, &; Burns, 2005). **Reading and spelling are more similar than different**..."

Furthermore, intuitively, degree modified adjectives like, for example, *{perfectly, mostly, very, somewhat} happy,* may contribute information about the number of dimensions (e.g., all, most, many, some, respectively) whose norms their argument exceeds. Again, dimension-counting scales predict the availability of such readings. In sum, the dimension-counting account captures the wider set of interpretations of positive, comparative and degree-modified forms of multidimensional adjectives as compared with dimensional adjectives (Sassoon, in progress).

Moreover, while dimensional adjectives by definition do not have readings involving quantification over dimensions, they may have readings involving quantification over other types of objects. Such readings can also be diagnosed using exception phrases. For example, we can describe a crowded classroom using (6a), but we cannot describe a sick child using (6b).

(6) a. The classroom is full/empty except for one chair.
b. #The child is warm expect for one degree.

Arguably, these judgments stem from the tendency of *full* towards interpretations with a maximum standard as opposed *warm*, which does not have this tendency. In the context of the utterance in (6a), *full* is associated with a chair-counting scale ranging between 0 and the maximum number of chairs in the given classroom. Thus, *full* denotes the relation between degrees d and locations containing chairs x, which holds iff the number of occupied chairs in x is d. Since *full* tends toward a maximum standard, *pos full* truly apply to a location x iff every chair in x is occupied.

By contrast, *warm* in a context suitable for the utterance in (6b), has a conventional mid-scale standard of 36˚. Thus, *The child is pos warm* conveys that the child's temperature is warmer than 36˚. Since these truth conditions do not reduce to universal quantification over temperatures, an exceptive is not licensed. Again, we see a connection between exceptive licensing and maximal standards. Thus, exceptive licensing can form a test for maximal standards of cardinality scales, including dimension-counting or other object-counting scales.

In fact, Yoon (1994) argued for adjectival readings involving universal and existential quantification over individuals or subparts. Many of Yoon's examples are also multidimensional adjectives. For example, for a table or table part to be dirty it has to be dusty, stained, oily, crumby, or dirty in some other way, while for it to be clean it has to be clean in every way. Thus, positive forms like (7a,b) may involve quantification over parts, dimensions, or both.

(7) a. The tables are clean/not dirty (except for the one on the left, which is dusty).
b. The table is clean/not dirty (except for one part, which is slightly dusty).

The evaluation of whether a table/part is, e.g., *rather dirty*, *fairly clean,* or *cleaner* than another table/part depends on what precisely is being counted.

With this in hand, we move on to scale structure theory's tests.

3 Standard types predicted by the different hypotheses

Gradable predicates divide by whether their unique scale has a minimum, a maximum, both or neither, and whether their standard is identified with the scale maximum, minimum, or neither (Kennedy, 2007; Kennedy & McNally, 2005; Rotstein & Winter, 2004; Van Rooij, 2010; Syrett, 2007). In relative adjectives, like *interesting,* the standard is context relative, while in absolute adjectives the standard is a lexicalized scale-endpoint. For instance, one indication that *clean* has a maximum standard is the intuition that (8a), unlike (8b), is a contradiction, as the symbol $\#_C$ indicates. The source of the contradiction in (8a) is the inference in (9a). The consistency of (8b) correlates with the inference failure indicated in (9b).

(8) a. $\#_C$This table is clean, but that one is cleaner.
b. This table is beautiful, but that one is more beautiful.
(9) a. $\#_C$That table is cleaner than this one.
=> This table is not maximally clean. => This table is not clean.
b. That table is more interesting than this one.
=> This table is not (maximally) interesting. $\not\Rightarrow$ This table is not interesting.

An indication that *different* has a minimum-standard is the intuition that (10a) is a contradiction, unlike (10b). The source of the contradiction in (10a) is the inference in (11a). The consistency of (10b) correlates with the inference failure in (11b).

(10) a. $\#_C$This chair is not different from mine, but is more different than that one is.
b. This chair is not beautiful, but is more beautiful than that one is.
(11) a. $\#_C$This chair is more different from mine than that one is.
=> This chair is at least minimally different from mine.
=> This chair is different from mine.
b. This chair is more interesting than that one is.
=> This chair is at least minimally interesting. $\not\Rightarrow$ This chair is interesting.

These tests of standard type yield clear results in adjectives that are **not** multidimensional, like those used in much of the experimental work on scale structure theory (e.g., Syrett, Kennedy, & Lidz, 2009). This work addressed mostly basic scales of dimensional adjectives like

straight-bent, *full-empty* and *transparent-opaque*. To illustrate, examples (12a,b) are clearly inconsistent, indicating a maximum- and minimum-standard for *straight* and *bent*, respectively.

(12) a. $\#_C$ This rod is straight, but that one is more straight.
b. $\#_C$ This rod is not bent, but is more bent than this one.

However, when the interpretation of adjectives involves quantification over dimensions, the test results might be affected by the force of the quantifier over dimensions. In particular, the tests don't yield clear results when the default standards of the multidimensional interpretation, dimensional interpretation, and dimensions' interpretations are not identified with the same scale point. To illustrate, the tests demonstrate that the interpretations of *similar* or *familiar* are typically associated with **midpoint or minimum-standards**. The consistency of (13a,b) suggests that their standard is not identified with the scale-maximum. The fact that (13c,d) is judged as inconsistent, suggests that the standard is identified with the scale minimum, while them being judged as consistent may suggest a midpoint standard.

(13) a. This version is similar to the original draft, but that one is more similar.
b. This fruit looks familiar, but that one looks more familiar.
c. $?_C$ This version is not similar to the original, but it is more similar than that one.
d. $?_C$ This fruit does not look familiar, but it looks more familiar than that one.

However, such an application of the tests ignores the fact that *similar* and *familiar* have multidimensional interpretations. For example, the paper versions talked about in (13a) can be similar or not in font type, font size, line spacing, length, wording, topics, depth, or strength of argumentation. Thus, (13a) can relate to font size only, to a uniform complex dimension based on weighted summing over different respects, or to a scale based on counting different respects in which the two versions are similar. Similarly, a fruit can be familiar or not with respect to shape, size, color, taste, having seeds or not, having stripes or not, or serving certain purposes or not. Thus, again, *familiar* as applied to fruits has a variety of interpretations.

Moreover, we have shown that the exception-phrase tests of universality and existentiality over dimensions illustrated in (2) indicate whether a multidimensional interpretation tends to have a maximum or minimum standard. The corpus results reviewed in section 2 suggest that positive forms of positive adjectives often involve universal quantification over dimensions. In particular, the distribution of exception phrases suggests that *pos familiar* often conveys being familiar in all relevant respects. Since, according to the standard theory, the test in (13b) indicates that *familiar* **does not have a maximum standard**, the proposal that *familiar* is associated with a complex (e.g., averaging-based) dimension with a minimum or mid-point standard cannot explain *familiar*'s tendency toward universal quantification over dimensions. Following this proposal, fruits are predicted to count as familiar iff they are familiar to at least some non-maximal degree in a single (basic or complex) dimensional scale. Thus, it does not follow that familiar fruits are familiar in every respect.

However, this tendency can be captured assuming a multidimensional (dimension-counting) interpretation for this adjective that does involve **a maximum standard**. Given the empirical

generalization that positive adjectives tend to be universal over their dimensions (see section 2), being positive, the default interpretation of *pos familiar* is familiar in every respect. Given this universal interpretation, exceptives are correctly predicted to be licensed.

In fact, when the adjectives in Sassoon's (2013) corpus study were divided into adjectives with default maximum and minimum standards by the standard tests of scale structure theory, the two adjective-sets did not differ significantly in their universality vs. existentiality index (exceptive frequencies). The reason was precisely that some maximum standard adjectives were negative and thus more existential than universal over their dimensions (like *unfamiliar*), while some midpoint or minimum standard adjectives were positive and thus more universal than existential over their dimensions (like *familiar*).

Hence, the proposal that multidimensional adjectives have an interpretation based on dimension-counting, together with the generalization that multidimensional interpretations tend toward a maximum standard in positive adjectives and a minimum standard in negative adjectives, is needed to capture the corpus data and judgments in (13). Based on the proposal and generalization, the multidimensional truth conditions of positive forms of, for example, *familiar*, require that for **all** familiarity respects F, the individual talked about would be at least **somewhat** familiar with respect to F. The consistency of (13b) follows, because other individuals may be even more familiar than the fruit talked about. They may, for instance, be **very** familiar in the given respects.

As for (13d), it is only predicted to be inconsistent assuming the standard hypothesis that *familiar* is associated with a complex (e.g., averaging-based) dimension, by which *not pos familiar* means not being even a bit familiar (having a zero degree of familiarity). By contrast, assuming a multidimensional interpretation along the dimension-counting proposal, *not pos familiar* is consistent with being a bit familiar in some respects or others, and therefore more familiar than individuals who are in no way even a bit familiar. Such predictions are consistent with the shaky status of speakers' judgments about (13c,d).

Moreover, in the corpus investigated by Sassoon (2013), almost all the adjectives that were usually universal over their dimensions, like *typical*, admitted some existential uses and vice versa, regardless of the type of standard typically associated with them in scale-structure theory. In those cases, a complex dimensional interpretation with the standard identified by the scale structure theory tests cannot provide a sufficient account for the quantification over dimensions, and a multidimensional interpretation based on dimension-counting is needed to do the job. Dimension-counting scales seem to usually be closed on both sides ranging between 0 and the cardinality of the entire dimension set. Thus, these standard shifts are possible.

Moreover, often speaker judgments are not conclusive when applying scale structure theory's tests even with adjectives like *safe, clean* or *healthy,* which are usually maximum-standard. This happens because they may admit mid-point relative interpretations as well. Thus, speakers diverge on whether the examples in (14) are inconsistent as the literature assumes (Kennedy, 2007; Kennedy & McNally, 2005; Rotstein & Winter, 2004), or not.

(14) ?$_{\mathrm{C}}$ This is {clean, safe, healthy}, but that is (even) {cleaner, safer, healthier}.

Context relativity may pervade the interpretation of absolute multidimensional adjectives either through a shift to non-absolute standards (e.g., *safe* may convey being safe in most respects), or through a context-relative dimension (e.g., *safe with respect to robbery*). Then there are no entailment relations between the multidimensional and complex-dimension interpretations (for context effects in absolute adjectives see Bierwisch 1989; McNally 2011).

To conclude, this paper proposes an account for cases in which speakers' judgments on the tests of scale structure theory are less conclusive than expected assuming simple or complex dimensional interpretations of adjectives. The tests provide important information about the types of standards of adjectives, but they must be used with caution in order not to confound the types of standard of the different possible interpretations of each adjective, including, in particular, interpretations involving quantification over dimensions. When these exist, the test results might be affected both by the standards of the dimensions and by the force of quantification over dimensions (the standard of the multidimensional interpretation), as indicated by the exceptive phrase tests. These tests are needed to supplement judgments.

Sassoon (in progress) shows that this conclusion applies to additional tests of scale structure theory. For example, in this theory, absolute modifiers like *perfectly* select adjectives whose scale has a maximum (Rotstein & Winter 2004; Kennedy, 2007; Kennedy & McNally, 2005; Syrett, 2007). By contrast, absolute modifiers like *somewhat* or *slightly* select adjectives whose scale includes a minimum. However, Sassoon (in progress) illustrates that degree modified multidimensional adjectives have dimension-counting interpretations (see section 2). Moreover, in Sassoon's (2013) study, the universality score of each adjective (as given by the frequencies of exceptive-modification of its occurrences) strongly correlated with the frequency of its modification by *perfectly* ($r = 0.7$). Universality did not correspond with having a maximum standard by the tests of scale structure theory. Overall, this supports an account of *perfectly* as selecting multidimensional adjectives that are universal over their dimensions, even when their simple- or complex-dimensional interpretations have relative or minimum standards according to the tests of scale structure theory (e.g., *beautiful* or *similar*). Again the exceptive tests are needed to supplement scale structure theory's tests.

References

Bartsch, Renate & Theo Vennemann. 1972. The grammar of relative adjectives and comparison. *Linguistische Berichte 20*: 19-32.

Bartsch, Renate. 1984. The structure of word meanings: Polysemy, Metaphor, Metonymy. Fred Landman & Frank Veltman (eds.), *Varieties of Formal Semantics*: 25–54. Foris Publications. Dordrecht.

Bierwisch, Manfred. (1989). The semantics of gradation. In Manfred Bierwisch & Ewald Lang (eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation*, 71–261. Springer-Verlag.

Bylinina, Lisa. 2013. *The grammar of standards.* Doctoral dissertation, Utrecht University.

Gardenfors, Peter. (2004). *Conceptual Spaces—The Geometry of Thought*. MIT Press. Cambridge, Massachusetts.

Hoeksema, Jack. (1995). The semantics of exception phrases. In J. van Eijck & J. van der Does (eds.) *Quantifiers, Logic, and Language*, 145-177. CSLI. Stanford.

Kennedy, Chris. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and philosophy, 30*(1), 1-45.

Kennedy, Chris. (2013). Two sources of subjectivity: Qualitative assessment and dimensional uncertainty. *Inquiry, 56*(2-3), 258-277.

Kennedy, Chris., & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 345-381.

Lasersohn, Peter. (1999). Pragmatic halos. *Language*, 522-551.

McNally, Louise. (2011). *The relative role of property type and scale structure in explaining the behavior of gradable adjectives.* In Rick Nouwen, Robert van Rooij, Uli Sauerland & Hans-Christian Schmitz (eds.) *Vagueness in Communication 6517*: 151-168. Springer, Berlin.

McNally, Louise & Isidora Stojanovic. (2015). Aesthetic adjectives. In James Young. (ed.) Th*e semantics of aesthetic judgment*, Oxford: Oxford University Press.

Moltmann, Frederique. (1995). Exception sentences and polyadic quantification. *Linguistics and philosophy, 18*(3), 223-280.

Murphy, Gregory. (2002). *The Big Book of Concepts*. The MIT Press. Cambridge, MA.

Pothos, Emmanuel M. and Wills, Andy J. (2011). *Formal Approaches in Categorization*. Cambridge University Press, Cambridge, UK.

Rotstein, Carmen & Yoad Winter. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics, 12*(3), 259-288.

Sassoon, Galit W. (In progress). Multidimensionality in the grammar of gradability. In review process. Bar Ilan University.

Sassoon, Galit W. (2013). A typology of multidimensional adjectives. *Journal of Semantics* 30: 335-380.

Sassoon, Galit W. & Julie Fadlon. (2017). The role of dimensions in classification under predicates predicts their status in degree constructions. *Glossa: a journal of general linguistics*, 2(1), 42. DOI: http://doi.org/10.5334/gjgl.155

Solt, Stephanie. (2018). Multidimensionality, subjectivity and scales: experimental evidence. In Elena Castroviejo, Louise McNally & Galit Weidman Sassoon (Eds.) *The Semantics of Gradability, Vagueness, and Scale Structure -Experimental Perspectives.* Language, Cognition, and Mind, Springer: Switzerland.

Syrett, Kristen, Chris Kennedy & Jef Lidz. (2009). Meaning and context in children's understanding of gradable adjectives. *Journal of semantics*, 27: 1-35.

Syrett, Kristen. (2007). *Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives*. Doctoral dissertation. Northwestern University, Department of Linguistics.

Umbach, Carla. (2016). Evaluative propositions and subjective judgments. In Meier, Cecil & Jannet van Wijnbergen-Huitink. (Eds.) S*ubjective meaning: Alternatives to Relativism*. Walter de Gruyter GmbH and Co KG.

van Rooij, Robert. (2010). Measurement and interadjective comparisons. *Journal of semantics* 28(3): 335-358.

von Fintel, Kai. 1994. *Restrictions on quantifier domains.* Doctoral dissertation, University of Massachusetts, Amherst.

von Stechow, Arnim. (2007). The temporal adjectices früh (er)/spät (er) and the semantics of the positve. In Anastasia Giannakidou and Monika Rathert. (Eds.) *Quantification, definiteness, and nominalization*, 214-233. Oxford: Oxford University Press.

Yoon, Youngeun. (1996). Total and partial predicates and the weak and strong interpretations. *Natural Language Semantics, 4*(3), 217-236.