Experimenting with Degree

Summary. It is generally agreed that the interpretation of gradable adjectives (GAs) such as *tall* and *dark* in some way makes reference to **degrees**. But what exactly degrees are, and which adjectival forms invoke them, remain open to debate. Focusing on what notion of degree – if any – underlies the semantics of GAs in their positive (unmodified) form, we argue that experimental research can help to resolve these questions. Our results provide evidence that the interpretation of the positive form involves degrees organized into a scale with a distance metric, and in particular are inconsistent with proposals that scales are derived from an ordering on a comparison class (e.g. Bale 2008).

Theories of GAs. According the **delineation** approach of Klein (1980), GAs denote partial one-place predicates that induce a three-way partition on a comparison class C. A notion of degree can be added to this to account for measure phrases (e.g. *6 feet*), but plays no role in the semantics of the positive form. In contrast, **degree**-based theories (Cresswell 1976, von Stechow 1984, a.o.) take GAs to express relationships between individuals and degrees on a scale; the semantics of all adjectival forms, including the positive, are stated in terms of degrees. Here the specifics differ. Some authors (e.g. von Stechow) consider degrees and scales to be something abstract (**abstract degree** theory), while others (e.g. Cresswell) adopt the more concrete view that scales are constructed from comparison classes, as follows: an ordering is established on a comparison class relative to the dimension in question (e.g. height), with the equivalence classes under this ordering constituting the degrees of the scale (**derived degree** theory). Bale (2008) extends the derived degree approach with the proposal that for adjectives associated with a numerical system of measurement (e.g. *tall*), measurements themselves (e.g. *6 feet*) enter into the underlying ordering as individuals, with the result that the derived scale is isomorphic to that associated with the measurement system (e.g. height in feet).

Different Predictions. What has not been fully recognized is that these theories differ in how they allow the truth conditions of the positive form to be stated, and thus in the predictions they make as to how speakers' application of the adjective will vary across contexts. The delineation approach is most easily reconciled with truth conditions such as (1a), which is not based on degrees; such a definition leads us to expect that speakers will consistently call a fixed proportion of a comparison class (say, the top third) *tall*, etc. The derived degree approach also supports (1a) (which can be restated in degree terms; Bale 2011) as well as (1b) (Bale 2008). But since this approach derives only an ordinal level scale lacking a distance metric (Kranz et al. 1971), it does not allow truth conditions of the form in (1c). For (1c), we require a scale at the interval or ratio level. This is possible in the abstract degree theory, where we can assume a more informative scale whose structure is independent of that of the comparison class. More generally, the derived degree theory predicts that whether an entity is classified as *tall*, etc. relative to a comparison class C must be determined on the basis of its **ordinal degree** – the rank order of its equivalence class in the ordering on C. This limitation does not hold for the abstract degree theory, which allows a notion of **absolute degree**. Finally, Bale's mixed theory predicts a difference in behavior between adjectives that are associated with a numerical measurement system (e.g. *tall*) and those that are not (e.g. *dark*); only the former should have access to absolute (rather than ordinal) degrees.

(1) $[\![$ John is tall $]\!]^C$=1 iff...

a. ...John is among the tallest n% of the Cs

b. ...$HEIGHT(j)$ is among the top n% of heights of Cs

c. ...$HEIGHT(j) \succ mean_{x \in C}(HEIGHT(x))$

We test these predictions experimentally, following a method developed by Barner & Snedeker (2008) and Schmidt et al. (2009), in which subjects are presented with arrays of pictures representing comparison classes with varying distributions, and asked to indicate which pictures could be described by a given adjective.

Experiment 1. The first experiment involved 4 adjectives - *large*, *tall*, *dark* and *pointy* - each paired with an array of 36 pictures spanning 11 'degrees' of size/height/etc. (respectively: eggs varying in size; cartoon characters varying in height; gray squares varying in shade; triangular shapes varying in angle). Four distributions of pictures over degrees were tested: Gaussian (largest # of eggs in medium sizes; fewer very small or very large); left skewed, right skewed, moved (Gaussian distribution shifted to overall greater sizes/etc.). The study was conducted online via Amazon MTurk in 4 versions (n=194).

A linear mixed model revealed that the average # of items classified as *large/tall/ dark/pointy* was significantly different across conditions ($p<0.001$). This indicates that judgments of gradable adjectives such as *large* cannot be based simply on picking the top n% of a ranking of comparison class members on the dimension in question (cf. 1a); rather, degrees are necessary. There was also a significant difference ($p<0.001$) across conditions in the average 'cut-off points' for *large* etc. (the degree of the smallest item called *large*, etc.). Thus *large* etc. also cannot be identified with a fixed segment of the range of degrees corresponding to the comparison class, per (1b).

Experiment 2. While Experiment 1 rules out the truth conditions in (1a,b), it leaves open the possibility that the interpretation of gradable adjectives is nonetheless based in some way on ordinal rather than absolute degrees – consistent with the derived degree theory. We address this possibility in Experiment 2. For each of the 3 adjective/picture pairs *large* (eggs), *tall* (cartoon characters) and *dark* (gray squares), a baseline distribution was constructed in which a target set of items represented the 4th of 6 ordinal degrees of the relevant dimension. This was compared with a rank-equivalent distribution, featuring a target set of items identical in ordinal degree (4 out of 6) but lower in absolute degree (i.e. smaller/shorter/less dark). The # of items in the target set, and the # of items greater in degree, was held constant across distributions. The study was executed online via MTurk (n=170).

The derived degree theory predicts no difference between baseline and rank-equivalent distributions in the proportion of target set items checked. However, we found a significant difference between these two conditions (baseline 59%; rank equiv. 7%; $p<0.001$), indicating that a scale constructed on the basis of equivalence classes under an ordering on a comparison class is not sufficient to account for speakers' judgments. Rather, we require a notion of absolute degree on a scale with a distance metric (e.g. absolute size), which supports truth conditions such as (1c). Importantly, the adjective *dark* (which lacks a common numerical unit of measure) behaved the same in this respect as *large* and *tall* (for which there are such measures), suggesting that the existence of a measurement system is not responsible for the availability of the necessary scale structure.

Conclusions. Whether an entity is considered *large*, *dark* or *tall* is not based simply on its position in a ranking of members of a comparison class it belongs to, but rather must reference its degree of size, etc. This is most consistent with a degree- rather than delineation-based theory of gradability. Furthermore, the relevant notion of degree is one in which the underlying scale includes a distance metric. Nor do we find evidence that the existence of a distance metric depends on the availability of a numerical system of measurement (per Bale 2008). Overall these findings support a view of degrees and scales as something abstract, and not one in which scales are derived from comparison classes.