

Gradable Adjectives, Vagueness and Optimal Language Use: A Speaker-Oriented Model

Ciyang Qing & Michael Franke, ILLC, University of Amsterdam

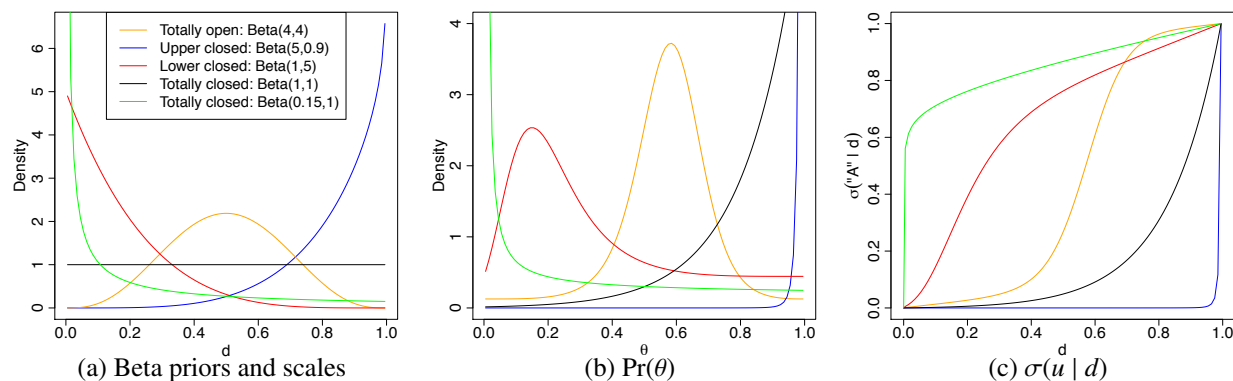
Degree-based approaches to the semantics of gradable adjectives [e.g. 1] hold that the meaning of the *positive form*, such as “tall” in the sentence “John is tall,” is obtained by composition with a silent morpheme *pos*: $\llbracket pos \text{ tall} \rrbracket = \lambda x. \mathbf{height}(x) \geq \theta$, where **height** is a function that maps individuals to *degrees* on an underlying *degree scale*, and θ is the *standard of comparison*. [1] proposed that abstract properties of degree scales influence the contextual resolution of θ . *Relative adjectives* like “tall” have totally open scales without minimal or maximal elements and allow θ to be resolved quite freely in context. *Absolute adjectives* like “wet” have closed degree scales including either their upper or lower bound (or both) and force θ to have a relatively rigid, context-independent interpretation corresponding to one of the scale’s end points. Absolute adjectives with totally closed scales can have a maximal and minimal standard reading depending on context.

Kennedy tried to explain this interaction by appeal to *interpretive economy* (IE): resolution of θ should make maximal use of semantic resources, including, if available, salient endpoints of degree scales. Subsequent contributions have tried to give a functional grounding of the resolution of θ in terms of evolutionary pressure for optimal language use [e.g. 2, 3]. A related approach is taken by the rational speech-act model (RSA) of [4], where the interpretation of gradable adjectives is given by probabilistic reasoning about hypothetical (sub-)optimal speaker behavior.

We present a model that combines basic tenets of these previous approaches, but overcomes some of their major shortcomings. In particular, the new model (1) adds a fully predictive speaker component (RSA really only covers interpretation, not production), (2) incorporates a cost parameter as a general contextual factor and makes plausible predictions for various values, and (3) predicts the context-independence of absolute adjectives’ interpretation, which few previous models have attempted. The key idea is motivated by evolutionary considerations, i.e. speakers employ a standard of comparison θ with a probability proportional to the *communicative efficiency* that results from using θ as a general convention. The relevant contextual variance is the *contextual degree distribution*, i.e., the general probability with which objects (of the given general reference class) have the property in question to a certain degree. Concretely, following [3, 4], we adopt the metaphysically austere view that degree scale types are relevantly different mainly because they are associated with different classes of probability distributions over degrees. Whether a scale is open or closed reduces to whether the probability of lower and upper bounds is negligible or not. Examples for the ensuing relation between scale types and degree distributions are given in Fig. (a) (degree ranges rescaled to fit the unit interval).

Like the RSA model, we assume a descriptive use of gradable adjective A to answer the question under discussion “what degree of A -ness does x have?” The efficiency of using θ as a conventional standard, given a contextual distribution over degrees ϕ , can then be measured in terms of the *expected success* of a speaker trying to raise the listener’s level of credence in the actual degree d_x to which x has property A under a literal interpretation given threshold θ . Expected success is defined in the usual way as “probability-weighted sum” over all potentially actual degrees d , times the utility for the case that d is actual, which in this case is the listener’s level of credence in d given that the speaker follows the convention: $ES(\theta) = \int_{-\infty}^{\infty} \phi(d) \cdot \phi(d|u, \theta) dd = \int_{-\infty}^{\theta} \phi(d)^2 dd + \int_{\theta}^{\infty} \phi(d) \frac{\phi(d)}{\int_{\theta}^{\infty} \phi(d) dd} dd$. The left summand applies when A is not true of x given θ , in which case the speaker cannot use utterance u “ x is A ” truthfully, and only listener’s prior beliefs apply. (The speaker could say different things, but we are only measuring the quality of a level of applicability for the phrase “ x is A .”) The right summand applies when A is true of x given θ , in which case the speaker can utter “ x is A ” truthfully and the listener can update his prior beliefs with the information that $d_x \geq \theta$.

Using a standard soft-max function [5], we capture actual threshold choices in production as



the probability $\Pr(\theta) \propto \exp(\lambda \cdot U(\theta))$, where $U(\theta)$ is the general utility after taking cost into account. This captures the probability with which speakers would adhere to standard θ if they tend to use language optimally, but might make mistakes of various sorts, as captured by rationality parameter λ in the usual way. Probabilities over conventional thresholds under this rule are shown in Fig. (b) for the different priors in Fig. (a) (corresponding cases are colored equally, and for space reasons we only show the basic case in which cost is negligible). The corresponding production probabilities are shown in Fig. (c), based on the rule $\sigma(u | d) = \int_{-\infty}^d \Pr(\theta) dd$, i.e., the sum probability of all thresholds no greater than d [6]. We can further derive the listener’s interpretation rule by applying Bayes’ rule: $\rho(d | u) \propto \phi(d) \cdot \sigma(u | d)$ (plots are as expected and skipped for space reasons).

The plots suggest that availability of endpoints on a scale (in the sense of sufficient probability mass) makes endpoint-conventions optimal. We can show this suggestive trend even analytically. Our formal arguments target distributions in the beta-family, but are easily seen to generalize. Concretely, we can show that (modulo cost): (i) if there is a sufficient amount of probability mass on the upper end point, we receive a maximal standard reading; (ii) otherwise if the probability mass at the lower endpoint is sufficiently larger than elsewhere, we receive a minimal standard reading; (iii) otherwise we receive a relative standard that is free to vary with the prior distribution ϕ . This has a direct bearing on Kennedy’s original observation. For relative adjectives on open scales, case (iii) is relevant, and we predict the expected relatively free contextual variation with ϕ (and cost). For upper closed scales we predict a maximum standard reading, as desired. For totally closed scales either (i) or (ii) applies, so we predict either maximal or minimal standards, depending on properties of ϕ . Finally, for lower closed scales that violate (ii), the model captures the same exception predicted by [4]. In addition, our model predicts that (non-radical) contextual variance in ϕ or cost will not affect the maximum and minimum standards of absolute adjectives.

- [1] Christopher Kennedy. “Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives”. In: *Linguistics and Philosophy* 30 (2007), pp. 1–45.
- [2] Christopher Potts. “Interpretive Economy, Schelling Points, and Evolutionary Stability”. Manuscript, UMass Amherst. 2008.
- [3] Michael Franke. “On Scales, Saliency & Referential Language Use”. In: *Amsterdam Colloquium 2011*. Ed. by Maria Aloni, Floris Roelofsen, and Katrin Schulz. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012, pp. 311–320.
- [4] Daniel Lassiter and Noah D. Goodman. “Context, Scale Structure, and Statistics in the Interpretation of Positive-Form Adjectives”. In: *Proceedings of SALT 23*. forthcoming.
- [5] Duncan R. Luce. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley, 1959.
- [6] Daniel Lassiter. “Vagueness as Probabilistic Linguistic Knowledge”. In: *Vagueness in Communication*. Ed. by Rick Nouwen et al. Springer, 2011, pp. 127–150.