A Rational Speech-Act model for the pragmatic use of vague terms in natural language

Alexandre Cremers (alexandre.cremers@gmail.com)
Vilniaus Universitetas, Filologijos Fakultetas
Universiteto g. 5, Vilnius 01122, Lithuania

Abstract

The question of *why* human language relies so heavily on vague terms has received a great deal of attention from philosophers, linguists, and more recently cognitive scientists, yet much less is known about their effect on other aspects of language use. In this paper, we propose a model for the interaction between vagueness and implicatures, an important pragmatic phenomenon, incorporating recent work in the RSA framework and insights from the philosophical literature on vagueness. We show that the model offers a good fit of data from earlier studies, and discuss the scope of the model more broadly.

Keywords: Language; Vagueness; Implicatures; Bayesian pragmatics; Rational Speech-Act; Supervaluationism

Vagueness in natural languages

Vagueness is pervasive in human languages (Russell, 1923), and a multitude of theories have been put forward to explain why language is vague and how vagueness can be modelled. A defining property of vague predicates is their admittance of *borderline* cases, for which neither the predicate nor its negation is clearly applicable.

(1) Context: Anna is slightly above average height
   a. #Anna is tall
   b. #Anna is not tall

Various strategies have been adopted to offer a formal account of vagueness, including fuzzy logic (Zadeh, 1965), trivalent logics (Fine, 1975, among many others), and recently probabilistic accounts (Qing & Franke, 2014; Lassiter & Goodman, 2015). Much less has been said about how vagueness affects other aspects of language use, but the experimental literature offers a handful of observations about interactions between vague gradable adjectives and implicatures (Gotzner, Solt, & Benz, 2018; Leffel, Cremers, Gotzner, & Romoli, 2019; Mazzarella & Gotzner, 2021). Yet we still lack explicit models of such interactions. In this paper, we leave aside the question of why and how vagueness arises, and focus on its concrete effects on message choice and interpretation. We propose a quantitative model which builds on recent developments in the Rational Speech-Act framework of Frank and Goodman (2012), but incorporates important insights from the philosophical and formal semantic literature on vagueness as well. We show that this model offers a very good fit of publicly available data from Leffel et al. (2019), and we lay out a few predictions to test in future research.

Motivating puzzle

Leffel et al. (2019) observed a surprising contrast between two categories of gradable adjectives, which they attribute to a difference in how vague they are:

(2) a. John is not very tall $\not\rightsquigarrow$ John is tall
   b. The antenna is not very bent
   $\rightsquigarrow$ The antenna is (somewhat) bent

Both adjectives are gradable, but 'tall' is *relative*, while 'bent' is a *minimum-standard absolute* adjective. Relative adjectives are vague (one cannot pinpoint exactly which heights count as tall) and context-dependent ('tall' conveys very different heights whether we consider persons or buildings), while absolute adjectives can easily receive a clear threshold and thus do not depend on context much. Among absolute adjectives, it is usual to distinguish minimum-standard and maximum-standard adjectives. The former convey that the predicated object possesses at least some degree of the property (e.g., 'bent', 'open', 'wet'); the latter that it possesses the property to a maximal degree (e.g., 'full', 'straight', 'dry').

(2b) behaves as linguists would expect: the inference that the antenna is at least somewhat bent can be explained as an implicature, by competition with the simpler and more informative alternative 'not bent' obtained by deletion of the adverb 'very'.[1] The puzzle is that the exact same reasoning could apply to (2a), but unless 'very' is stressed, the implicature is absent. Leffel et al. show that the contrast can be replicated with other vague and non-vague constructions (e.g., 'not very hot' vs. 'not much hotter than average').

They remark that no height can both clearly satisfy 'tall' and clearly falsify 'very tall', making the candidate strengthened meaning of (2a) akin to *borderline contradictions* such as "tall and not tall" (Ripley, 2011). In practice, the range of heights which are somewhat compatible with "tall but not very tall" is very narrow, and with small differences between the thresholds the speaker and listener assign to 'tall', the ranges of heights each of them consider "tall but not very tall" may not overlap at all. By contrast, in (2b) one can choose a degree arbitrarily close to 0 in order to satisfy both 'not very

[1] Such implicatures are often called *structural implicatures* because the alternative is a substring of the utterance (Simons, 2001), by contrast with *scalar implicatures*, which require retrieving a lexical alternative from a scale of related terms.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

bent' and a strict interpretation of 'bent'. Leffel et al. propose to model this reasoning by generalizing a notion of *innocent exclusion* initially developed by Fox (2007) to prevent contradictory implicatures (in particular from disjunctions, such as "A or B, and not-A, and not-B"). Implicatures would be derived by an operator EXH which would avoid not only classical contradictions, but also borderline contradictions such as "tall and not very tall".

While this explanation captures the initial observation, encoding implicatures' sensitivity to vagueness in the semantics raises some concerns. Besides being inherently *ad hoc* and raising important questions about modularity and the semantics/pragmatics interface, it does not explain why stressing 'very' in (2a) would make the implicature available when no such thing happens in the classical contradiction cases which motivated Fox (2007)'s innocent exclusion. We propose an alternative model with well-defined roles for pragmatics and semantics, which explains the contrast without revising the standard definition of EXH. We show that this model goes further than Leffel et al.'s informal explanation by making accurate quantitative predictions about their data.

In the next section, we present the RSA-SvI model (for Supervaluationist Intentions). We then explain how we evaluated the model against data from Leffel et al. (2019)'s experiments, and conclude by discussing other possible applications of the model and current limitations.

The RSA-SvI model

Informal description

Our model of vagueness and implicatures fits in the Rational Speech-Act framework (Frank & Goodman, 2012), but unlike Lassiter and Goodman (2015) and successors, we do not attempt to explain how vagueness arises. Rather, we measure vagueness and take it as a starting point to offer an account of how it affects message choice for a pragmatic speaker and interpretation by a pragmatic listener. The model incorporates ideas from the philosophical logic and formal semantics literatures on gradable adjectives and vagueness. In particular, we focus on a key property of vague predicates which received limited attention in the modeling literature: higher-order vagueness (Dummett, 1959). If we were to define a predicate 'borderline-tall' to characterize individuals like Anna in (1), this predicate would be vague itself. In probabilistic terms, not only the threshold θ from which someone counts as 'tall' is uncertain, but the distribution of θ (or the set of parameters of this distribution in a parametric setting) should itself be treated as a random variable.

The second key ingredient of our model is *supervaluationism*, a concept from philosophical logic proposed by van Fraassen, Bas C (1966) and first applied to vagueness by Fine (1975). The idea is that sentences with vague terms are underdetermined and could be *precisified* in many possible ways. A sentence is *supertrue* only if it is true under any possible precisification, and conversely *superfalse* if it is false under any precisification. Borderline cases correspond to sentences which are true under some precisifications and false under others. Supervaluationism has been adapted to the RSA framework by Spector (2017), albeit for a different phenomena (*homogeneity effects*) with only two possible precisifications. His idea is to consider the utility of an underdetermined message as its average utility across all possible precisifications. In the RSA framework, utility diverges to $-\infty$ as the probability of the message being true approaches 0, so a message must be true under all possible precisifications to receive a finite utility and be used.[2]

The last ingredient to our model is the mechanism by which it derives implicatures. Following the grammatical view of implicatures (Fox, 2007; Chierchia, Fox, & Spector, 2012) and recent work in the RSA framework (Champollion, Alsop, & Grosu, 2019; Franke & Bergen, 2020), we assume that implicatures are computed in the semantics by a specialized operator EXH similar to a silent 'only'. Pragmatic reasoning is reduced to a simple disambiguation problem between parses with and without EXH. In the case of (2a), a literal parse would simply convey that John's height is less than what qualifies as 'very tall', while an exhaustive parse would convey that John is 'tall but not very tall'. We adapt Franke and Bergen (2020)'s Global Intentions RSA model for disambiguation, which differs radically from the supervaluationist treatment of underspecification: the speaker chooses the pair (message, parse) which best conveys their intention. In particular, this decision rule does not prevent the speaker from using a message u when one of its parses is false or likely false (e.g., the exhaustive parse of 'not very tall').

Piecing everything together, the model captures the observation in (2a) as follows: upon hearing 'not very tall', the pragmatic listener knows that—in principle—the speaker could have either an exhaustive or a literal interpretation in mind. However, no matter which height the speaker wanted to convey, the exhaustive interpretation has a very low expected utility (across all possible vague denotations for 'tall' and 'very tall'): in supervaluationist terms, no height makes 'EXH[not very tall]' supertrue. By contrast, the literal interpretation is compatible with low heights under any reasonable threshold for 'very tall'. The listener therefore draws the inference that the speaker almost certainly meant the literal interpretation, and that John is somewhat short.

Detailed implementation

Semantic assumptions: We assume that gradable adjectives denote measuring functions and require a silent operator POS to combine with entities in sentences like "Anna is tall" (Kennedy & McNally, 2005). POS introduces a threshold variable θ, so that "Anna is tall" is true if and only if Anna's degree of height, obtained by applying the measure function denoted by 'tall' to Anna, exceeds the threshold θ:

(3) $\quad [\![\text{Anna is POS tall}]\!] = \lambda w.\mu_{\text{tall}}(a) > \theta$

[2]Uncertainty regarding the question under discussion also plays an important role in Spector's model, but we leave it aside for now.

Following Qing (2021), we assume that minimum standard adjectives can combine with either the POS morpheme of Kennedy and McNally, yielding a loose interpretation, or MIN, resulting in a strict interpretation $\theta = 0$. Intensifiers such as 'very' are treated as overt realizations of POS which additionally shift the threshold by a positive quantity δ. By treating θ and δ as random variables, we assign the following graded truth-conditions to the different messages and parses,[3] where d is the degree to convey, and Θ a set of hyperparameters describing the distribution of θ and δ:

(4)
$$\begin{aligned}
[\![\text{adj}]\!]^{d,\text{MIN},\Theta} &= \mathbb{1}_{0<d} \\
[\![\text{adj}]\!]^{d,\text{POS},\Theta} &= P(\theta < d|\Theta) \\
[\![\text{not adj}]\!]^{d,\text{MIN},\Theta} &= \mathbb{1}_{d\leq 0} \\
[\![\text{not adj}]\!]^{d,\text{POS},\Theta} &= P(d \leq \theta|\Theta) \\
[\![\text{very adj}]\!]^{d,\Theta} &= P(\theta+\delta < d|\Theta) \\
[\![\text{not very adj}]\!]^{d,\text{LIT},\Theta} &= P(d \leq \theta+\delta|\Theta) \\
[\![\text{not very adj}]\!]^{d,\text{EXH}_{\text{MIN}},\Theta} &= P(0 < d \leq \theta+\delta|\Theta) \\
[\![\text{not very adj}]\!]^{d,\text{EXH}_{\text{POS}},\Theta} &= P(\theta < d \leq \theta+\delta|\Theta)
\end{aligned}$$

A noteworthy aspect of our semantics is that negation flips truth-values but does not affect thresholds (akin to fuzzy negation, and in line with empirical observations, e.g., Hersh & Caramazza, 1976). We also point out that the ambiguity between MIN and POS is only relevant in the case of minimum-standard adjectives, as the MIN interpretation is trivial for relative adjectives like 'tall', and that it leads to two possible exhaustive parses, depending on which alternative EXH negates.

Pragmatic model: Our L_0 listener is parametrized by Θ and a parse i. In other words, L_0 only takes into account first-order vagueness (she has uncertainty regarding θ and δ, but is certain about their distribution).

(5) $$L_0(d|u,i,\Theta) \propto P(d)[\![u]\!]^{d,i,\Theta}$$

The speaker S_1 selects the pair (u,i) such that u under parse i maximizes expected utility across all Θ values, where utility is a trade-off between informativity and a term reflecting the cost of uttering u, to which we will come back below.

(6) $$U_1(u,i|d) = \int \log L_0(d|u,i,\Theta)P(\Theta)d\Theta - c(u)$$

(7) $$S1(u,i|d) \propto \exp(\alpha U_1(u,i|d))$$

A crucial feature of this model is that supervaluationism only comes into play at the level of Θ (higher-order vagueness), and not θ (first-order vagueness). Directly averaging utility over θ would turn the model into a categorical one, where 'tall' can only apply to extremely tall individuals, removing all vagueness along the way. The current implementation is more flexible, in that it allows 'tall' to be used as long as no value of Θ makes the sentence strictly false.

The pragmatic listener L_1 jointly infers d and i by applying Bayes' rule, with uniform prior on $i|u$ (each message can receive different parses, so the set of parses over which we define the uniform prior depends on the message). For 'late' we consider all and only the parses listed in (4), for 'tall' we further exclude parses with MIN or EXH_{MIN}, which are trivial.

(8) $$L_1(d,i|u) \propto P(d)S_1(u,i|d)$$

Model evaluation

We tested the model by fitting data from Leffel et al. (2019, Exp. 1), which measured the acceptability of sentences with relative 'tall' and minimum-standard 'late' (in the sense of being late to work) at 13 scale points each. For instance, participants would have to specify how much they agree or disagree with Mary's statement in the following situation (the initial instructions specified that work started at 9am):

(9) **Fact:** Donna showed up to work at 8:48am.
Mary said: "Donna was not very late."

Both adjectives appeared in various constructions, including '*adj*', 'not *adj*', 'very *adj*' and 'not very *adj*'. Participants expressed their agreement using a continuous slider from 'disagree' to 'agree'. Crucially, 'not very tall' was interpreted as meaning roughly the same as 'not tall', while 'not very late' was heavily degraded when the subject arrived early, as in (9), suggesting a 'late but not very late' interpretation.

Methods

Before we can derive model predictions, we need to specify the distribution of Θ (recall that our model describes the effects of vagueness but not how vagueness arises in the first place). Our solution was to estimate this distribution empirically from the acceptability of the constructions "*X* is *adj*" and "*X* is very *adj*", and feed it to the RSA-SvI model to predict the acceptability of the constructions "*X* is not *adj*" and "*X* is not very *adj*", given model parameters (rationality and costs) which we fitted. Figure 1 gives an overview of the whole analysis process.

We take acceptability judgments to reflect a relatively low-level literal interpretation, namely the probability that a sentence be true, albeit averaged across parses, with the weight of each parse determined by pragmatic reasoning. This is in essence similar to Lassiter and Goodman (2015)'s assumption that the acceptability of "*X* is *adj*" reflects the cdf of θ. For 'tall' in affirmative sentences, there is no parse ambiguity. We assumed that θ_{tall} follows a normal distribution of parameters (μ,σ) and δ an exponential distribution of parameter λ. We assumed that $\Theta = (\mu,\sigma,\lambda)$ follows a hybrid multi-

[3] It is possible to keep the truth-conditions binary and have a truly probabilistic interpretation by adding a hypothetical "literal speaker" S_0 parametrized by (θ,δ) in the RSA model described below. For instance, for the positive form, S_0 would be:

$$S_0(adj,\text{POS}|d,\theta) = 1 \text{ iff } \theta < d$$

Letting L_0 average across all such S_0 yields a model formally equivalent to the one defined below. For instance:

$$\begin{aligned}
L_0(d|adj,\text{POS},\Theta) &\propto P(d)\int S_0(adj,\text{POS}|d,\theta)P(\theta|\Theta)d\theta \\
&\propto P(d)P(\theta < d|\Theta)
\end{aligned}$$

u	i	cost
$\emptyset$	—	0
adj	POS MIN	c_{adj}
not *adj*	POS MIN	$c_{adj} + c_{\text{neg}}$
very *adj*	—	$c_{adj} + c_{\text{very}}$
not very *adj*	LIT EXH_{MIN} EXH_{POS}	$c_{adj} + c_{\text{neg}} + c_{\text{very}}$

Table 1: List of messages and parses tested in the model for Leffel et al. (2019)'s data, with associated costs. $\text{MIN/EXH}_{\text{MIN}}$ parses only appeared with 'late' since they are trivial for 'tall'.

variate normal/log-normal distribution (Fletcher & Zupanski, 2006), with hyperparameters $\Omega = (m_\mu, m_\sigma, m_\lambda, \Sigma)$, where μ is marginally normal and σ, λ are marginally log-normal. For 'late', we assumed both θ and δ to follow exponential distributions, with parameters λ_θ and λ_δ respectively. The origin for θ was 9am, and times earlier than this were encoded as negative degrees (it was therefore encoded in the model that arriving early falsifies 'POS late'). The two λ parameters were assumed to follow a multivariate log-normal distribution; we also included a parameter ζ which described the probability of a 'MIN late'interpretation. Acceptability was fitted as normally distributed around this prediction:[4]

$$(10) \qquad Acc(\text{“}X \text{ is late”}; d) \sim \mathcal{N}\Big((1-\zeta)P(\theta < d) + \zeta \mathbb{1}_{d>0}, \varepsilon\Big)$$

We fitted the data on affirmative constructions with hierarchical models implemented in Stan (Carpenter et al., 2017). Each participant in Leffel et al.'s study was assigned a Θ sampled from a distribution parametrized by Ω. From these models, we saved the fitted hyperparameters $\hat{\Omega}$, as well as the vector of random effects $(\hat{\Theta}_s)$. Using $\hat{\Omega}$ to parametrize the distribution of Θ, the first term of the utility function in equation (6) was precomputed for 346 d values for each message-parse pair in Table 1. We used normal priors on d (tall: 69.2 ± 2.66in, from real-world data; late: 9:00am$\pm$10min, arbitrary).

We then fitted a second hierarchical model on the negative constructions "not *adj*" and "not very *adj*" with parameters α and the three costs parameters used in Table 1, again with a multivariate hybrid distribution where all parameters were marginally log-normal except the cost c_{adj}, which admitted negative values.[5] Given these parameters and the precomputed utility terms, we computed the L_1 posterior on parse i

[4]Note that slider data is bounded to $[0,1]$, so it would be better modeled with a censored Gaussian (or logit-transformed as suggested by Franke et al., 2016, although it's not clear how one would deal with boundary values in this case). However, the RSA models described below were particularly difficult to fit, so we adopted this simpler Gaussian model as a compromise.

[5]The set of costs is defined up to an additive constant, so this is equivalent to allowing a positive cost for the null message.

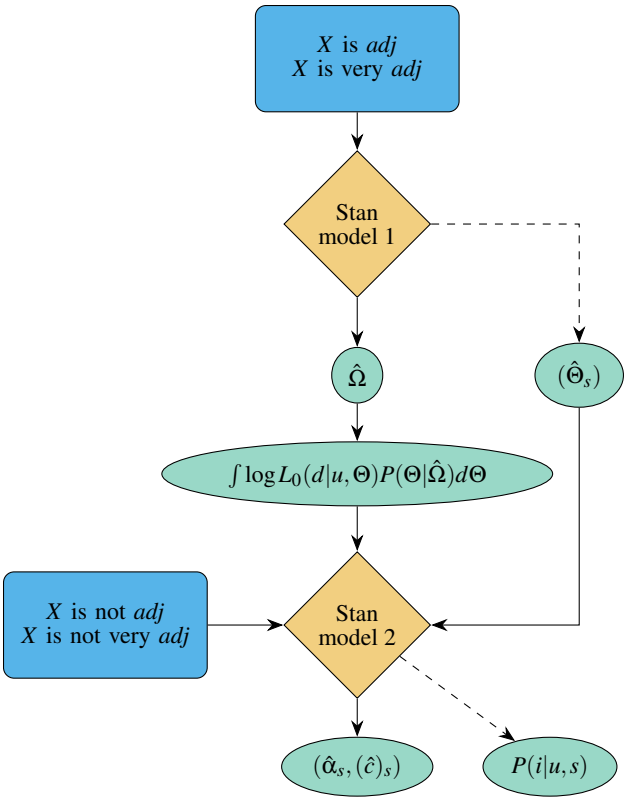

Figure 1: Flowchart for the analysis of Leffel et al's data. The variable s indexes participants. Blue squares represent data, orange diamonds – models, and green ellipses – fitted parameters. Dashed arrows indicate tranformed parameters extracted from the models in addition to the free parameters.

for the two negative constructions (marginalizing over d). Within the Stan model, integrals were approximated using Simpson's 3/8 rule. The acceptability of message u for a participant s was assumed to follow:

$$Acc(u;s) \sim \mathcal{N}\left(\sum_i P([\![\text{u}]\!]^i = 1|\hat{\Theta}_s) L_1\left(i|u, \hat{\Omega}, \alpha_s, (cost)_s\right), \varepsilon\right)$$

$\hat{\Omega}$ and the $\hat{\Theta}_s$ are taken directly from the fit on affirmative constructions, so the only free parameters in this second model are the hyperparameters governing the distribution of $(\alpha, c_{adj}, c_{\text{neg}}, c_{\text{very}})$ across participants and the noise parameter ε. Unlike the first model which was fitted separately for 'tall' and 'late', the second model was fitted on both adjectives simultaneously. In other words, we assume that the α and costs for participants in both datasets come from the same distribution (except the cost for the adjective which was allowed to differ for 'tall' and 'late').

Results

The posterior estimates for the hyperparameters are given in Table 2 and the model fit for individual participants is presented in Figure 2. The resulting distribution on posterior probability of exhaustive interpretations for "not very

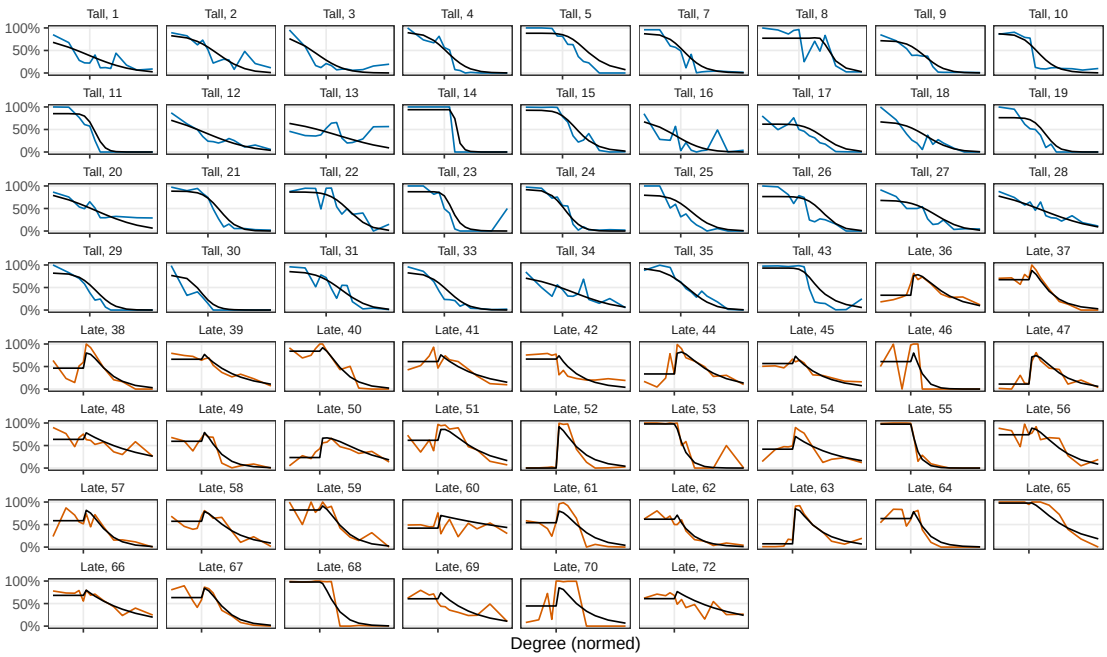

Figure 2: Data (colored line) and model predictions (black line) for each participant in Leffel et al. (2019)'s Experiment 1.

parameter	mean	95% CI	
m_{α}	−0.54	−1.00	0.05
s_{α}	1.41	1.05	1.82
$m_{c_{\text{tall}}}$	6.18	0.74	12.78
$m_{c_{\text{late}}}$	0.33	−1.24	2.28
$s_{c_{adj}}$	1.86	0.85	3.13
$m_{c_{\text{neg}}}$	2.05	0.14	4.10
$s_{c_{\text{neg}}}$	0.39	1.6e−4	1.28
$m_{c_{\text{very}}}$	−0.82	−5.02	3.44
$s_{c_{\text{very}}}$	0.33	8.2e−5	0.93

Table 2: Hyperparameters from the second model, with estimated posterior mean and highest-density credible intervals. All parameters followed a marginal log-normal distribution except c_{tall} and c_{late}. We skip the correlation matrix as the strongest correlation (between c_{adj} and c_{neg}) only reached a mean of −.07.

adj" was .15 for 'tall' (95% HDI-CI on the mean, $[.15, .20]$, range among participants: $[.06, .38]$), and .42 for 'late' (95% HDI-CI on the mean, $[.40, .44]$, range among participants: $[.02, 1]$), split roughly equally between EXH_{MIN} and EXH_{POS}. The model therefore predicts 'not very tall' to be exhaustified significantly less than 'not very late'. Besides, the highest exhaustification probabilities predicted for 'tall' appear to be artifacts from participants who gave noisy responses or didn't use the top of the slider (e.g., subject 17). By contrast, we observe genuine variation in the propensity to exhaustify with 'late' (from nearly 0 for subjects 53 and 55 to nearly 1 for subject 63). To confirm that high exhaustivity with 'tall' does not reflect an optimal strategy given certain combinations of costs and Θ, we looked at the effect of the rationality parameter, displayed in Figure 3. Crucially, as α increases, $P(\text{EXH}|\text{not very late})$ converges to either 0 or (usually) 1, but $P(\text{EXH}|\text{not very tall})$ always falls to 0.

It is not unreasonable to imagine that the good fit offered by the RSA-SvI model is not due to the model itself but to its use of parameters fitted to the affirmative sentences, which contain a lot of information on how individual participants interpret such sentences. To evaluate the specific added value of our model, we compared it with a "literal model", which simply treats the negative sentences as the literal negation of their affirmative counterparts. PSIS leave-one-out cross-validation (Vehtari, Gelman, & Gabry, 2017) with participants as unit data points indicated that the RSA-SvI performed much better than the literal model ($\Delta ELPD = 1041 \pm 151$, effective number of parameters: $p_{loo} = 65.3$ vs. 14.7).

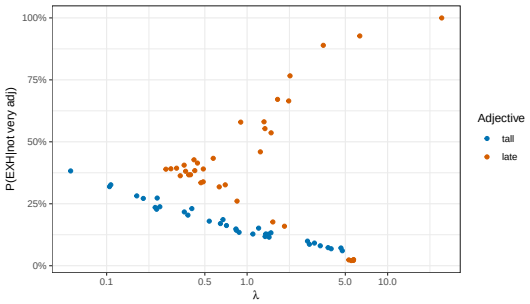

Figure 3: Estimated probability to exhaustify 'not very *adj*', function of the fitted rationality parameter $\hat{\alpha}$ (by participant).

Discussion

We proposed a theoretically-motivated model incorporating recent advances from the RSA framework as well as important insights from formal semantics and philosophical logic. We showed how this model allows us to formalize an explanation for a puzzle regarding the interaction between vagueness and implicatures. Quantitatively, the fit of the data is very good, and qualitatively, the model correctly predicts that a rational speaker would not use 'not very tall' to convey "tall but not very tall". Note however that some fitted parameter are rather extreme—in particular the cost of negation which is estimated at 8.4 on average—, raising concerns that our model overfitted the data (it may be selecting implausible sets of costs in order to attain the best-fitting posterior exhaustivity for each participant). To address this concern, we fitted a simplified version of our model where both adjectives shared the same cost c_{adj}, and without the EXH_{MIN} parse for 'not very late' (which may bias the model towards more exhaustivity for this adjective). Formally, the simpler model only had one less degree of freedom (one hyperparameter less for costs). Removing the EXH_{MIN} parse does not affect the number of parameters, but makes the model inherently less flexible. The resulting fit was significantly worse ($\Delta ELPD = 381 \pm 78$, effective number of parameters 65.3 vs. 62.9), but still much better than the literal model, and the fitted costs were less extreme (in particular the mean c_{neg} across participants went down from 17.6 to 3.9). Qualitatively the main difference was on participants who had a very peaked interpretation of 'not very late' (which EXH_{MIN} captures best).

Let us now come back to the observation that stressing 'very' in (2a) allows the implicature to resurface. To the best of our knowledge, the only work addressing stress in the RSA framework is Bergen (2016). Without going into too much detail, Bergen proposes that prosodic stress is a way for the speaker to selectively reduce noise (and therefore potential misperceptions on the listener's side) on part of an utterance, at a small positive cost. In an RSA model, the presence and position of stress thus leads to pragmatic effects without any semantic contribution. By stressing 'very', the speaker indicates their intention to draw the listener's attention to the contrast with the bare adjective, and signals that the question they are addressing makes this difference relevant (both Bergen, 2016 and Spector, 2017 include a model of the *question under discussion*, which we haven't touched upon, as mentioned in fn. 2). Independently of this, Bennett and Goodman (2018) demonstrate that costlier intensifiers have a larger effect. Stressing 'very' (a costly move according to Bergen) could therefore increase its intensifying effect (our variable δ). Our model predicts "EXH not very tall" to have low but finite utility because the range of heights that count as both 'tall' and 'not very tall' is too narrow and unstable. Increasing this range would mechanically increase the utility of the EXH parse, and stress would further signal that the contrast between 'not very tall' and 'not tall' is relevant to the speaker, making the implicature much more attractive. Without offering explicit account of the effect of stress yet, the fact that one seems at least possible is a direct improvement on the proposal of Leffel et al. (2019).

In this paper, we focused on a very specific puzzle regarding negated intensified adjectives, but we would like to underline how general our model really is. It can in principle be applied to any sentence involving vague terms, and a particularly important avenue for future research is the interaction with quantification. Take (11) for instance, where 'tall' appears in the restrictor of 'every':

(11) Every tall student laughed

A supervaluationist account like ours predicts (11) to convey that even students who are borderline tall laughed. Indeed, if a student is borderline tall, some amount of Θ's make her count as tall. If such a student laughed, the sentence would be false under these Θ, rendering it unusable. In addition, by negation of its simpler alternative "Every student laughed", (11) is expected to implicate that not every student laughed, and therefore that at least one student is clearly not tall. Such predictions will be interesting to test in future research.

At this point, the main limitation of the model is that it requires empirical measurements for each vague term appearing in a given sentence. A few quantitative models have been proposed to predict the interpretation of gradable adjectives (Qing & Franke, 2014; Lassiter & Goodman, 2015), but they only consider first-order vagueness, their empirical adequacy is still low (Zhao, 2018), and they require knowledge of a prior on the degree distribution (so not much is gained in practical terms). Progress in this area would immediately widen the empirical scope of our model.

Acknowledgement

I would like to thank Paul Egré, Nicole Gotzner and Benjamin Spector for discussion of this work, as well as four anonymous reviewers, whose precious feedback significantly improved this paper. This research was funded by the European Social Fund and the Lithuanian Research Council (LMT) under the measure 09.3.3-LMT-K-712 "Development of Competences of Scientists, other Researchers and Students through Practical Research Activities".

References

Bennett, E. D., & Goodman, N. D. (2018). Extremely costly intensifiers are stronger than quite costly ones. *Cognition*, *178*, 147-161.

Bergen, L. (2016). *Joint inference in pragmatic reasoning*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, *9*.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1), 1–32.

Champollion, L., Alsop, A., & Grosu, I. (2019). Free choice disjunction as a rational speech act. In K. Blake, F. Davis, K. Lamp, & J. Rhyne (Eds.), *Proceedings of SALT 29* (Vol. 29, pp. 238–257).

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In *Semantics: An International Handbook of Natural Language Meaning* (Vol. 3, pp. 2297–2331). Berlin: Mouton de Gruyter.

Dummett, M. (1959). Wittgenstein's philosophy of mathematics. *The Philosophical Review*, *68*(3), 324–348.

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, *30*(3-4), 265–300.

Fletcher, S. J., & Zupanski, M. (2006). A hybrid multivariate normal and lognormal distribution for data assimilation. *Atmospheric Science Letters*, *7*(2), 43–46.

Fox, D. (2007). Free choice disjunction and the theory of scalar implicature. In U. Sauerland & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (p. 71-120). New York, NY: Palgrave Macmillan.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Franke, M., & Bergen, L. (2020). Theory-driven statistical modeling for semantics and pragmatics: A case study on grammatically generated implicature readings. *Language*, *96*(2), 77–96.

Franke, M., Dablander, F., Schöller, A., Bennett, E., Degen, J., Tessler, M. H., ... Goodman, N. D. (2016). What does the crowd believe? a hierarchical approach to estimating subjective beliefs from empirical data. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society.* Austin, TX: Cognitive Science Society.

Gotzner, N., Solt, S., & Benz, A. (2018). Scalar diversity, negative strengthening, and adjectival semantics. *Frontiers in Psychology*, *9*. doi: 10.3389/fpsyg.2018.01659

Hersh, H. M., & Caramazza, A. (1976). A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General*, *105*(3), 254.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*.

Lassiter, D., & Goodman, N. D. (2015). Adjectival vagueness in a bayesian model of interpretation. *Synthese*.

Leffel, T., Cremers, A., Gotzner, N., & Romoli, J. (2019). Vagueness in Implicature: The Case of Modified Adjectives. *Journal of Semantics*, *36*(2), 317-348.

Mazzarella, D., & Gotzner, N. (2021). The polarity asymmetry of negative strengthening: dissociating adjectival polarity from face-threatening potential. *Glossa*, *6*(1), 1–17.

Qing, C. (2021). Zero or minimum degree? Rethinking minimum gradable adjectives. *Proceedings of Sinn und Bedeutung 25*. doi: 10.18148/sub/2021.v25i0.964

Qing, C., & Franke, M. (2014). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In T. Snider, S. D'Antonio, & M. Weigand (Eds.), *Proceedings of SALT 24* (pp. 23–41).

Ripley, D. (2011). Contradictions at the borders. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *Vagueness in communication* (pp. 169–188). Springer.

Russell, B. (1923). Vagueness. *The Australasian Journal of Psychology and Philosophy*, *1*(2), 84–92.

Simons, M. (2001). On the conversational basis of some presuppositions. In R. Hastings, B. Jackson, & Z. Zvolensky (Eds.), *Proceedings of SALT 11* (pp. 431–448).

van Fraassen, Bas C. (1966). Singular terms, truth-value gaps, and free logic. *The journal of Philosophy*, *63*(17), 481–495.

Spector, B. (2017). The pragmatics of plural predication: Homogeneity and non-maximality within the rational speech act model. In A. Cremers, T. van Gessel, & F. Roelofsen (Eds.), *Proceedings of the 21st Amsterdam Colloquium* (p. 435).

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, *27*(5), 1413–1432. Retrieved from `https://doi.org/10.1007/s11222-016-9696-4` doi: 10.1007/s11222-016-9696-4

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338–353.

Zhao, Z. (2018). Interpreting intensifiers for relative adjectives: Comparing models and theories. In *Proceedings of the ESSLLI 2018 student session* (pp. 142–151).