Negation, polarity and scale structure: Inferences of gradable adjectives

Stavroula Alexandropoulou & Nicole Gotzner
University of Potsdam

The type of standard invoked by gradable adjectives crucially affects entailment relations between antonymic pairs. The negation of an absolute adjective like *clean* entails the assertion of its antonym *dirty* while this is not the case for relative adjectives like *large* and *small* (Cruse, 1986, Rotstein & Winter, 2004, Kennedy, 2007, a.o., see Table 1). This is because relative adjectives allow for a middle ground ('neither small nor large'). Here, we investigate how the scale structure underlying different types of adjective affects the derivation of pragmatic inferences, given different entailment patterns. Table 1 compares different candidate inferences for relative and absolute terms involving weak and strong scale-mates. '??' indicates that the respective inference is not predicted by standard accounts (see Discussion).

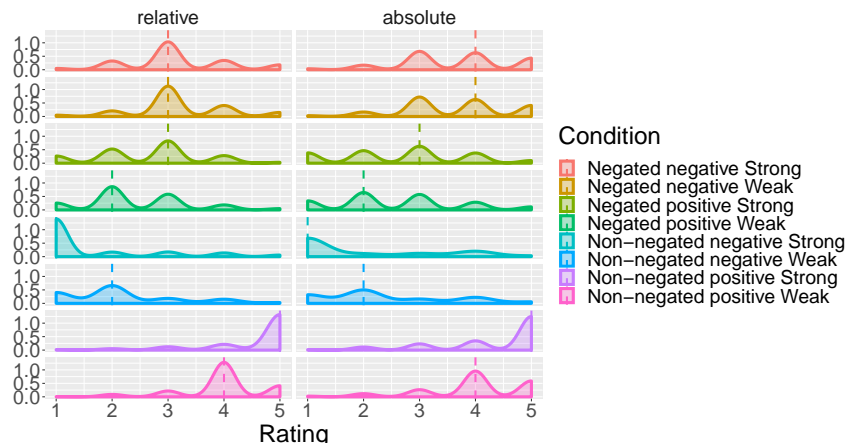Table 1: **Different interpretations of negated (positive) weak and strong terms.**

Adjective type		Relative		Absolute	
Inference	*The door is*	Strong *not gigantic*	Weak *not large*	Strong *not pristine*	Weak *not clean*
Entailment	'The door is'	'less than gigantic'	'less than large'	'less than pristine'	'dirty'
Indirect scalar implicature	'The door is'	'large'	??	'clean'	??
Middling interpretation	'The door is'	'neither large nor small'		??	??
Negative strengthening	'The door is'	??	'(rather) small'	'(rather) dirty'	??

Previous work has identified polarity to be a main factor in the derivation of pragmatic inferences: While positive terms like *large* are strengthened under negation ('rather small'; so-called *negative strengthening*, Table 1), negated negative terms like *not small* tend to receive a middling interpretation ('neither large nor small', e.g., Horn, 1989). For absolute adjectives, negative strengthening—if at all—should be available for strong positive terms (*not pristine*), as negated weak terms (*not clean*) are expected to be interpreted semantically (cf. *Entailment* in Table 1). Our **main research question** is how weak and strong terms are interpreted under negation, depending on their polarity.

Methods—We developed a rating scenario that made distinctions between different interpretations relevant via an action-based task (inspired by Tessler & Franke, 2018; Benz & Gotzner, 2018). Our scenario capitalizes on the notion of evaluative polarity and introduces a fine granularity level. We ran two experiments with a 2 scalar strength (weak/strong) x 2 polarity (negative/positive) x 2 negation (non-negated/negated) design. Each participant judged 8 simple predication statements (1 per condition) involving different adjectives in 7 (Experiment 1) or 8 (Experiment 2) contexts, each with a different adjective scale. The judgments were made on a 1-5 point Likert scale with the endpoints standing for the strong scale-mates (e.g., 1:*tiny* and 5:*gigantic*).

Results—Figure 1 presents the ratings across conditions for relative (n = 53) and absolute adjectives (n = 48). The results of both experiments were analyzed with cumulative link models involving the maximal converging random effects structure. Contrast-coded models for both experiments showed that: Overall,

Figure 1: **Ratings in Experiment 1 (relative) and Experiment 2 (absolute). Dashed line represents median.**

Condition
- Negated negative Strong
- Negated negative Weak
- Negated positive Strong
- Negated positive Weak
- Non−negated negative Strong
- Non−negated negative Weak
- Non−negated positive Strong
- Non−negated positive Weak

people distinguish between positive and negative terms differing in scalar strength (scalar strength*polarity interaction; relatives: $z=-7.51$, $p<.0001$, absolutes: $z=-5.31$, $p<.0001$). These distinctions are found to be less pronounced under negation compared to the non-negated forms (3-way scalar strength*polarity*negation interaction; relatives: $z=-7.81$, $p<.0001$, absolutes: $z=-5.07$, $p<.0001$). Under negation, there were two main differences between the two adjective types, as shown by a model where scalar strength was nested under adjective type and polarity: (i) Weak conditions were significantly different from strong ones for positive relative items ($z=-3.45$, $p<.001$), but not for positive absolute ($p=.2$), negative absolute ($p=.91$) or negative relative items ($p=.27$), (ii) negative conditions of absolute terms received significantly higher ratings than the same conditions of relative terms (Polaritynegative:Type interaction: $z=5.04$, $p<.0001$). That is, among the four negated conditions in each experiment, people only distinguish between weak and strong positive conditions of relative terms (*not large* vs. *not gigantic*), while the two negated negative conditions of absolute terms (*not dirty/filthy*) receive generally higher ratings than the corresponding conditions of relative terms (*not small/tiny*).

Discussion—In our novel rating scenario, participants used distinct portions of a scale when interpreting statements involving relative or absolute adjectives. This is an indication that they perceived the difference in strength between expressions. However, under negation, they did not distinguish between negative weak and strong terms (*not tiny/not dirty* vs. *not small/not filthy*). For **relative** terms, the middle ratings to these two conditions, as well as to the negated strong positive one (*not gigantic*), manifest the derivation of a middling interpretation ('neither large nor small', Table 1; Horn, 1989; Israel, 2004). Moreover, the data of the negated weak conditions are in line with the so-called polarity asymmetry of negative strengthening (Horn, 1989; Ruytenbeek et al., 2017; Mazzarella & Gotzner, 2021), with positive terms (*not large*; Table 1) being more likely to be strengthened than negative ones (*not small*). For **absolute** terms, no middling interpretation is expected given that their semantics does not allow for a middle ground. However, we see evidence of such interpretations ('neither clean nor dirty') in the negated negative conditions (*not dirty/filthy*), and less in the positive strong one (*not pristine*; see also Paradis & Willners, 2006). We conjecture that this is due to the fact that our contexts impose a fine granularity level. Furthermore, the higher ratings of negated negatives (peak on 4, Fig.1) reveal inferences to the antonym, which, crucially, do not arise via entailment for negative strong terms like *not filthy* (contrary to the entailment for weak ones *not dirty* ⇒'clean'), nor via negative strengthening. Granularity is relevant here too: Negative terms are minimum standard adjectives, and based on their minimum standard semantics and a fine-grained scale, one can draw more distinctions w.r.t. to the property at stake than in the case of maximum standard adjectives (*clean/pristine*) that require a scale maximum (see Sassoon & Zevakhina, 2012). Thus, the scale range communicated by negated minimum standard adjectives given a fine granularity level is largely restricted, resulting in an overlap of the two negative conditions (*not dirty/filthy*). Going back to Table 1 and **comparing relative and absolute** terms under negation, we conclude that middling interpretations favor relative adjectives. A polarity asymmetry due to negative strengthening arises for weak relative terms, and possibly for strong absolute terms (cf. peak on 2 for positive items, Fig.1), if at all. Weak absolute terms are typically interpreted semantically, while granularity interacts with minimum/maximum standard semantics triggering additional inferences: middling and inference to the antonym.

Implications—Overall, our findings are in line with degree-based analyses of gradable adjectives and they show that different properties of measurement scales—the type of

standard and granularity— as well as evaluative polarity are responsible for the derivation of different (pragmatic) inferences (see also Gotzner et al., 2018a;b). In our talk, we discuss the mechanisms underlying the different types of implicature tested here. To account for the role of scale structure in implicature, we propose the so-called *measurement mechanism* (cf. Gotzner, 2021), which assumes that implicature involves reasoning about positions on a measurement scale.

References

Cruse, A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.

Gotzner, N. (2021). The use and interpretation of gradable adjectives. Invited talk at Bochum Language Colloquium, 28.1.2021, Ruhr University Bochum.

Gotzner, N., & Benz, A. (2018). The best response paradigm: a new approach to test implicatures of complex sentences. *Frontiers in Communication*, 2(21).

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

Gotzner, N., Solt, S., & Benz, A. (2018b). Adjectival scales and three types of implicature. In *Proceedings of Semantics and Linguistic Theory*, 28, 409-432.

Horn, L. R. (1989). *A natural history of negation*. Chicago: University of Chicago Press.

Israel, M. (2004). The pragmatics of polarity. *The handbook of pragmatics*, ed. by L. R. Horn & G. Ward, 701–723. Oxford: Blackwell.

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

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

Mazzarella, D., & Gotzner, N. (2021). The polarity asymmetry of negative strengthening: dissociating adjectival polarity from face-threatening potential. *Glossa: A Journal of General Linguistics*, 6(1), 47. http://doi.org/10.5334/gjgl.1342.

Paradis, C., & Willners, C. (2006). Antonymy and negation: The boundedness hypothesis. *Journal of Pragmatics*, 38(7), 1051-1080.

Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics*, 12, 259–88.

Ruytenbeek, N., Verheyen, S., & Spector, B. (2017). Asymmetric inference towards the antonym: Experiments into the polarity and morphology of negated adjectives. *Glossa: A journal of general linguistics*, 2(1), 92, 1-27.

Sassoon, G., & Zevakhina, N. (2012). Granularity shifting: Experimental evidence from degree modifiers. In *Proceedings of Semantics and Linguistic Theory*, 22, 226–246.

Tessler, M. H., & Franke, M. (2018). Not unreasonable: Why two negatives don't make a positive. https://doi.org/10.31234/osf.io/tqjr2.