

If you hear “Jones is happy” you understand that Jones’ happiness —a *gradable property*— exceeds a certain threshold (Kennedy and McNally, 2005; Kennedy, 2007). Where exactly this threshold lies is a matter of context, in particular prior expectations about usual degrees of happiness of a person like Jones, but can be derived formally by considerations of goal-oriented cooperative language use (Lassiter and Goodman, online first; Qing and Franke, 2014). The model of Lassiter and Goodman (online first), for example, captures a listener’s probabilistic reasoning about which thresholds would likely explain a speaker’s utterance if the speaker strives to maximize information flow about the true degree of Jones’ happiness. Here, we try to extend this line of inquiry to also incorporate compositional expressions involving various forms of negation: how happy or sad is an agent who is reported to be *not happy*, *not sad*, *unhappy* or *not unhappy*? We propose a probabilistic speaker-listener pragmatic reasoning model in the Rational Speech Act tradition (Frank and Goodman, 2012; Franke and Jäger, 2016; Goodman and Frank, 2016). Our model introduces elements of lexical uncertainty (Bergen, Levy, and Goodman, 2012, to appear) in order to model a listener’s uncertainty about how to interpret overt negation markers.

Negation is the semantic operation of forming an opposite, but there are several kinds of semantic opposition (Horn, 1989; Horn and Wansing, 2017). A *contrary opposition*, such as between *happy* and *sad*, is one where both predicates cannot be true at the same time, but can be false at the same time. A *contradictory opposition*, such as between *pure* and *impure*, is one where truth of one predicate entails falsity of the other. In other words, gradable terms that express contrary opposites allow for a neutral middle ground, unlike contradictory opposites. Contrary opposition is thus logically stronger than contradictory opposition, but frequently natural language expressions that appear to express contradictory opposition are pragmatically strengthened to convey contrary opposition instead. For example, today’s English adverb *never* derives from Old English *nǣfre*, a combination of *ne* (not) and *ǣfre* (*ever*) literally meaning *not ever*, but strengthened to mean *never*.

*Lexical antonyms* are pairs like *happy* and *sad* or *polite* and *rude*. Intuitively, lexical antonyms normally express contrary opposition: Jones can be neither happy nor sad without logical contradiction.

We distinguish two markers of negation in English, *adverbial negation* (e.g., *not happy*, *not polite*) and *affixal negation* (e.g., *unhappy*, *impolite*). In principle, both types of negation markers could map onto either type of semantic opposition relation, contrary or contradictory opposition. If  $Hx$  expresses that  $x$  is happy, we denote a contradictory opposition using standard bivalent negation  $\neg Hx$ . Contrary opposition is formed by a different kind of negation, which we denote as  $\tilde{H}x$ . The latter is a predicate-forming operation that is not iterable (Horn, 1989; Horn and Wansing, 2017). So while it makes sense to iterate  $\neg\neg Hx$ , it is impossible to iterated contrary negation. It is, however possible, to have  $\neg\tilde{H}x$ . But it is not possible to have  $\neg$  in the scope of  $\tilde{H}x$ .

This means that listeners may in principle be uncertain about which of several logical readings a speaker has in mind when she uses a negated gradable predicate. In particular, *not happy* and *unhappy* can both be construed as either  $\neg Hx$  or  $\tilde{H}x$ . Given the constraints mentioned above, the phrase *not unhappy* can receive two interpretations, namely  $\neg Hx$  and  $\neg\tilde{H}x$ .

In the probabilistic model that combines pragmatic reasoning about the meaning of gradable adjectives with reasoning about the meaning of negation markers, we will assume that contrary negation  $\tilde{H}x$  leads to a new independent threshold (with the right directionality). In other words, we leave it to rational pragmatic reasoning to decide optimal meaning assignments to predicates, including those formed from predicate-forming contrary negation.

This leads to the *uncertain parse model* ...

## References

- Bergen, Leon, Roger Levy, and Noah D. Goodman (2012). “That’s what she (could have) said: How alternative utterances affect language use”. In: *Proceedings of the 34<sup>th</sup> Annual Meeting of the Cognitive Science Society*.  
— (to appear). “Pragmatic Reasoning through Semantic Inference”. to appear in *Semantics & Pragmatics*.  
Frank, Michael C. and Noah D. Goodman (2012). “Predicting Pragmatic Reasoning in Language Games”. In: *Science* 336.6084, p. 998.  
Franke, Michael and Gerhard Jäger (2016). “Probabilistic pragmatics, or why Bayes’ rule is probably important for pragmatics”. In: *Zeitschrift für Sprachwissenschaft* 35.1, pp. 3–44.  
Goodman, Noah D. and Michael C. Frank (2016). “Pragmatic Language Interpretation as Probabilistic Inference”. In: *Trends in Cognitive Sciences* 20.11, pp. 818–829.  
Horn, Laurence R. (1989). *A Natural History of Negation*. Chicago: Chicago University Press.  
Horn, Laurence R. and Heinrich Wansing (2017). “Negation”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2017. Metaphysics Research Lab, Stanford University.  
Kennedy, Christopher (2007). “Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives”. In: *Linguistics and Philosophy* 30, pp. 1–45.  
Kennedy, Christopher and Louise McNally (2005). “Scale Structure, Degree Modification, and the Semantics of Gradable Predicates”. In: *Language* 81.2, pp. 345–381.

- Lassiter, Daniel and Noah D. Goodman (online first). “Adjectival vagueness in a Bayesian model of interpretation”. In: *Synthese*.
- Qing, Ciyang and Michael Franke (2014). “Gradable Adjectives, Vagueness, and Optimal Language Use: A Speaker-Oriented Model”. In: *Proceedings of SALT 44*. Ed. by Jessi Grieser et al. elanguage.net, pp. 23–41.