# How to be Kind with Prosody

*Michael Wagner*

McGill University

chael@mcgill.ca

*"How interesting!"*
*Stefanie Shattuck-Hufnagel in just about any conversation on prosody.*

## Abstract

What was said is often interpreted relative to what was left unsaid. Evaluate statements such as *That's good* can sound negative, because the speaker could have said *great* instead. *That's great*, on the other hand, might be interpreted as 'not so great', if we believe the speaker was just being nice. How, then, can we ever credibly convey our true intentions when making evaluate statements? We present evidence showing that prosody can be used to modulate the interpretation of evaluative statements, and can specifically be used to preempt inferences about positive evaluations toward a more negative interpretation. It is less able to modulate negative evaluations. The observed asymmetry makes sense if we tend to be kind to each other, and inflate our evaluative statements toward the nicer end of the spectrum.
**Index Terms**: prosody, emotion, implicature, sarcasm

## 1. Evaluations and their Enrichment

Suppose A comments to B about one of B's drawings by saying:

(1) A to B: It's good.

With this assertion, A expresses his evaluation of the drawing along some contextually relevant dimension, maybe artistic value or true-to-lifeness. We will refer to assertions that convey a speaker's opinion or attitude about something as an 'evaluative statement'. A lot can be at stake in making such statements, an overly harsh assessment by A could be taken by B as discouraging or unkind, an overly positive one as unctuous. In order to avoid such adverse effects, A might overstate or understate her appreciation of the drawing. Intuitively, there is an asymmetry though: It seems likely that B will assume that A's actual opinion is at most as positive as what A literally said, but might conclude it to be less positive. In other words, people seem more likely to overstate rather to understate their evaluations.

(2) Overstatement Inferences
   a. Good-means-bad
      A to B: It's is good. → B concludes that A actually thinks it's bad.
   b. Great-means-good
      A to B: It's great. → B concludes that A actually thinks it's merely good.

One could call overstatement inferences *just-being-nice*-inferences, but overstatements could have different motivations. A might act out of kindness, or might just be following a conventional social norm, and avoid being *damning with faint praise*. Being polite is a complex skill, and issues like saving face play an important role designing polite discourse moves [1]. The likelihood of overstatement will be modulated, for example, by whether A thinks B cares, and how A thinks B evaluates the drawing herself. What about inferences in the opposite direction? Would B ever conclude that A actually thinks it's great when she says 'good', based on the fact that A didn't say *terrible*? Or suppose that A does say *it's terrible*. When would B conclude that A thinks it's actually good?

(3) Understatement Inferences
   a. Good-means-great
      A to B: It's is good. → B concludes that A actually thinks it's great.
   b. Terrible-means-good
      A to B: It's terrible. → B concludes that A actually thinks it's pretty good.

Understatement inferences might arise if A is generally known to be prone to understatement, or is B's public nemesis, and there scenarios are imaginable, but they seem intuitively less plausible. Now consider reversals:

(4) Reversal Inferences
   a. Great-means-terrible
      A to B: It's is great. → B concludes that A actually thinks it's terrible.
   b. Terrible-means-great
      A to B: It's terrible. → B concludes that A actually thinks it's great.

Such reversals are what we prototypically think of when thinking about sarcasm. Mark Liberman notes in a 2003 blogpost that the reversal in (4-b) seems harder than (4-a)—this more evidence for the asymmetry we are after here. Note, however, that understatements and overstatements can also be ironic/sarcastic, and the terms 'sarcasm' and 'irony' should not be reserved only for reversals alone (cf. [2]). In fact, even the choice of tone of voice can be ironic/sarcastic (e.g., fake anger) in utterances whose content is intended literally.

The logic behind these inferences is reminiscent of the logic of scalar implicatures. Consider:

(5) A to B: I did some of the work. → B concludes that A didn't do all of the work.

A's assertion would lead B to the conclusion that A didn't do all of the homework. The reasoning given in the Gricean literature is that B draws the inference because A is presumably competent about what she did or didn't do, and if A had done all of the home work she would have said so since it's relevant. One crucial ingredient of scalar implicatures is that there are lexical alternatives that could have been said, that are at least as relevant, and that are stronger than what was actually said. In the

*Draft. To appear in the Proceedings of the Satellite Session on Framing speech: Celebrating 40 years of inquiry with Stefanie Shattuck-Hufnagel, at Speech Prosody 2016, Boston.*

case of (5), the alternative utterances involve elements of the scale <some, most, all> (cf. [3]). A second ingredient is the assumption that the interlocutor is cooperative, often assumed to involve following the Gricean maxims of conversation [4].

The scalar implicature in (5) is very similar to *good*-means-*not great* inferences. Note, however, that for the <*some, all*>-scale an inference is parallel to the Great-means-good inference in (2-b) seems intuitively much less likely:

(6) A to B: I did all of the work. → B concludes that A actually did some but not all of the work.

But why? Clearly there could be many reason why A might want B to think that A did all of the work. The problem is that if A actually overstates her accomplishments, the claim could easily be refuted. Compare the following variation:

(7) A to B: I spent all afternoon working on the homework. → B concludes that A actually spent a good portion of time on the homework but not all afternoon.

The inference in (7) seems more plausible than the one in (6). They differ in how hard it would be to find evidence that refutes the stronger statement if only the weaker statement is true, and *spending all afternoon* seems like a more subjective predicate and leave room for more interpretation than *did all the homework*. Evaluative statements have a special property in this regard: The truth of the matter is *entirely* subjective, and only directly accessible to the speaker, and this subjectivity seems to affect which the inferences we are liable to draw based on them.

The main hypothesis of this paper is that there is an asymmetry in the interpretation of evaluative statements favoring overstatements inferences. If true, we should see an asymmetry between positive and negative words, such that negative words are generally interpreted literally, while non-negative words should be able to convey a broader range of meanings since their interpretation is more likely to be enriched.

A second hypothesis regards how overstatement inferences can be avoided: If any evaluative statement can be misinterpreted due to a range of unintended inferences, how can we ever convey our true intentions? If overstatement is the rule, it seems that we would need to resort to constant hyperbole to avoid negative interpretations of our assessments, and words on evaluative scales should constantly shift toward the negative and need to be replaced by yet more positive lexical words. There is indeed 'grade inflation': Words like *good* and *nice* can easily be used when the actual assessment is average or even negative. In quantitive sentiment evaluations they tend to come out in the middle of the spectrum rather than on the positive end [5]. And yet, in general we are able to convey our intentions successfully.

How do we do it? The answer is, arguably, prosody. While prosody has been argued to play a role in evaluative statements (cf. [6], i.a.), the precise mechanics remain relatively unexplored. As far as we know, the interaction between lexical choices and emotive prosody has not been previously explored.

## 2. Prosody and pragmatic inferences

Intuitively, the tone of voice used in an evaluative statement seems important in the interpretation of evaluative expressions. There are at least three ways in which it might be.

### 2.1. Prosody reveals degree of emotive arousal

The first possibility is that prosody conveys level of excitement or emotional arousal, similar to exclamation marks in orthography. Especially in informal writing, such as email or text messages, exclamation marks play an important role in influencing the inferences drawn by the reader. An evaluative statement missing an exclamation mark might easily be interpreted as less positive than intended. And yet a non-literal interpretation can also come about through exuberance, e.g. by adding too many exclamation marks. Similarly, a highly positive lexical choice used without commensurate emotive excitement might lead to a overstatement inference. Emotive prosody in terms of pure excitement should make negative words more negative and positive words more positive, so we might find opposite effects of prosody on opposite ends of the lexical scale.

### 2.2. Prosody reveals emotional state or attitude

However, prosody has been argued to be capable of encoding emotional valence, not just level of arousal [7]. If true, it is more similar to emoticons and emojis [8] than to exclamation marks. Prosody has been shown to be able to reflect various basic emotions such as being happy, sad, or angry [9, 10, 11]. To some extent this might work simply by conveying information about the physiological state of the speaker, for example, we can hear when someone is smiling while speaking [12]. But emotive effects could occur at different levels: Depending on the situation, the prosodic valence might reveal something about the speaker emotional state (happy, sad, angry, ...), it might reflect the speakers attitude toward the speaker (critical, condescending, ...), or it might encode a propositional attitude [13].

Suppose emotive prosody was able to directly encodes the valence of the evaluative proposition, similar to the lexical choice: Then we might simply see a main effect of prosody, entirely separate from the effect of lexical choice. But we might also see more complex interactions: A statement that is seemingly positive based on the lexical choices pronounced with a prosody that conveys a negative evaluation might lead to an inference that the lexical choice was not genuine. A discrepancy in the opposite direction, a very negative statement said with a positive tone of voice might also trigger some inferences, although it is intuitively less clear which. However, it seems actually more likely that emotive prosodic cues will be taken to reveal something about the speaker's emotional state and attitude toward the addressee, rather than the propositional content. Any inference about the content of the message drawn by the listener might hence be much more indirect.

### 2.3. Conventionalized tunes

A third way in which prosody might affect evaluative statements is by way of conventionalized tunes, that encode a speakers' attitude or some aspects of the perlocutionary part of the message [14]. Suppose, for example, that there is something like an intonational wink, which conveys that something about a statement (its content or maybe the tone of voice) is meant ironically—just like an actual wink, and it would be a great cool to mark utterances as sarcastic. (Mark Liberman has cast doubt on the existence of such an intonational tool in English in a 2010 blogpost on the language log).

What aspects of prosodic meaning reflect conventionalized linguistic representations and which aspects are purely 'paralinguistic' is complex (cf. [15]). It seems, however, that there is some cross-linguistic consistency in how emotion is conveyed, suggesting that it is less conventionalized than other parts of its meaning [16]. Convention clearly plays an important role in emotive interpretation, however, for example the presence of periods when meessaging has taken on negative emotive import,

since it less obligatory to use it by default, as noted by Ben Crair in an article in the New Republic in 2013, and others.

### 2.4. The present study

Our first hypothesis is that there is an asymmetry in the interpretation of evaluative statements: Negative lexical choices tend to lead to negative evaluations, while positive lexical choices tend to lead to a greater range of interpretations, since we are prone to draw overstatement inferences. Testing this hypothesis is not trivial. For example, the original intent of evaluative statements is hard to retrieve from spoken corpora without a lot of additional knowledge about the individuals involved and the contexts in which their statements were made. We used scripted experiments in which we asked speakers to convey a certain intent using particular lexical choices. We then ran perception experiments to study how their utterances were interpreted. While this method has its own weaknesses, the results we obtained suggests that this is a productive method. Our second, related, hypothesis relates to prosody: If prosody is used to mitigate unintended pragmatic inferences, we should see that it can preempt overstatement inferences, and in general have a greater effect on words that are not on the negative end of the spectrum, since they are less likely to be enriched by pragmatic inferences.

## 3. The Experiments

We conducted a production experiment varying the lexical choice of an evaluative statement (highly negative words, highly positive words, and more or less neutral words). The words were normed with the help of the UMass sentiment corpus, which quantifies sentiment ratings for words based on the average Amazon rating of reviews that they occur in [5]. We tested 4 adjectives in each group, for a total of 12 test sentences. A total of 25 participants were recorded, and instructed that "You will be recorded on several short sentences. You will be asked to say the statement either in a very negative way (you express your feeling that you really don't like something and make a negative statement), or neutrally (you're neither very positive or very negative about it), or in a positive and enthusiastic way (you really like something and want to convey this emphatically)." On each trial, they were prompted with a sentence, e.g. *This movie was amazing*, and were instructed on what they were actually intending to convey. Each speaker produced 48 sentences.

A subset of about half of the 1200 total recordings were played to listeners in a perception study. They rated on a 7-point Likert scale what the believe the speaker actually thought. Each participant rated 60 sentences. The experiment took about 10 minutes. A total of 100 participants took part in the perception study, each individual soundfile was rated by 9-15 listeners.

### 3.1. Results and Discussion

The results of the perception experiment (Fig. 1) show a clear asymmetry: Utterances with very positive words were interpreted similarly to medium words unless speakers had been asked to convey a very positive evaluation, and intent also affected medium words. By contrast, intent had no effect on listeners ratings involving words with negative words.

Given the limited range of lexical choices we looked at this could be a floor effect—our negative words were maybe so negative that prosody was not able to counteract this. But very positive words, on the other hand, showed a very big effect of intent. The observed asymmetry is compatible with the hypothesis that overstatement inferences are more prevalent. We used

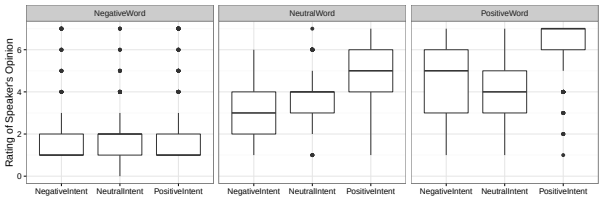


Figure 1: *1=very negative, and 7=very positive*

mixed model cumulative link regression with random effects for speaker, listener, and item, to test whether the interaction between lexical choice and intent was significant. We found that the effect of positive intent was significantly bigger in the case of very positive words compared to neutral words, which together showed a bigger effect than negative words.

Maybe speakers actually have a hard time sounding negative, but the observed pattern is also compatible with speakers simply having used level of excitement to manipulate their utterances, rather than encoding emotive valence in their utterances, and not having bothered to prosodically modulate statements with negative words, maybe knowing that prosody would not have a big effect. A detailed look at the acoustic results of the production is beyond what can be done in the confines of this short paper, but just looking at maximum pitch on the evaluative word suggests that it is not the case that speakers did not modulate their prosody on the negative end: In fact, the differences in maximum pitch are *bigger* on the negative end (Fig. 2).

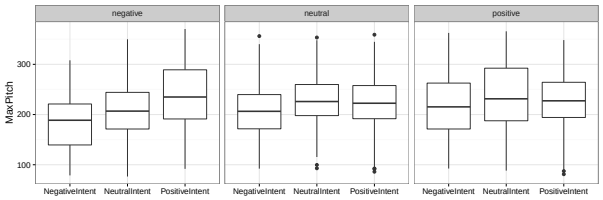


Figure 2: *Maximum pitch on the evaluative word.*

If prosody simply encoded level of excitement, we might expect greater prosodic effort on negative words to result in more *negative* evaluations—the observed use of greater maximum pitch to convey positive intent would hence be unexpected. Instead, the results suggest that speakers put on their best *positive* intonation, more so than on the positive end, maybe knowing that this would be a tough feat. The perception evidence shows that despite their efforts, they didn't succeed.

## 4. Conclusion

Our experiments go beyond prior findings in that they look at *interactions* between lexical choices and emotive content. The results suggest that listeners use prosody in deciding which inferences about a speaker's actual intent they should draw. Many other factors are likely to enter the equation (e.g., [17] show that emotive cues can be overridden by semantic cues), but our results show that prosody can play an important role.

While it seems likely that under the right circumstances, any of the inferences we outlined above might occur, our results support the idea that overstatement inferences are more likely. A reversal or an understatatement might simply require

much more prior knowledge about the situation, the likely intent, and/or the typical prosody of the speaker than our experiments provided. Prosody was mostly important when using positive and neutral words to avoid overstatement inferences, and not important at all when using negative words. The observed asymmetry makes sense if people are generally kind to each other: Listeners will correct for this, unless positive words are backed up by a prosody that supports the genuineness of the praise.

It would be interesting to test this in written communication: Exclamation marks and emojis should be more important (and hence frequent) for positive evaluations than for negative ones. A similar asymmetry could exist for facial expressions. It would also be interesting to explore cross-linguistic and cultural variation: Variation could occur in the degree of lexical inflation, in the degree of bias toward overstatement, and in the prosodic tools. The inability of our speakers to convey understatement inferences for negative words might simply show that listeners do not expect negative words to be used in non-literally. It could also be that we actually lack the prosodic tool of a negative intonation—such a gap in our repertoire would make sense if there is a general bias toward kindness. But maybe speakers find ways to express negative intent in situations that make understatement inferences more likely.

An analysis of the inferences with evaluative scales using the same tools used for other scales seems desirable. However, we saw that the basic logic of Gricean inferences needs to be enriched: The subjective nature of evaluative scales makes inferences plausible that seem implausible for other scales. An account based on more general principles of rational reasoning as in [18, 19] might be be more apt to work for both.

The limitations of the present study are numerous. Asking speakers to use positive words to convey a negative message and vice versa may create confusion, and also lead to unnatural utterances. The assumption that naive participants can act out intentions according to a script might also be overly optimistic. Our very limited set of lexical choices are also a limitation, it is very possible that individual words have a very different propensities to be interpreted non-literally due to their conventionalized meaning. For example, [5] report on words that have a 'bimodal' distribution varying between very positive and very negative, but we only controlled for the mean sentiment rating, and not their distributions. We also did not have a manipulation of the degree excitement without specifying the emotive valence, and important comparison case. Finally, the actual wording of our instructions confounded emotive valence and emphasis, which makes it harder to interpret the results.

It is likely that there are individual differences both in producing and interpreting emotive cues. While we might correct for a speaker's pitch range, it could be that some people are simply better at conveying their emotive intentions than others. We have not explored this in our data, but point to one outlier. It seems apparent that Stefanie Shattuck-Hufnagel is not just a maven when it comes to the study of prosody, but also a virtuoso of its use, in particular in the art of being kind and generous with prosody—in evaluative statements, but not just in those. We believe that this prosodic aptitude, and sometimes exuberance, as in the rendition of *Wowee!* with two intonational phrases reported in [20], is in this particular individual case a reflex of profound kindness and generosity. We base this on anecdotal non-prosodic evidence, and probably everyone in the field of prosody has had similar experiences. We will not speculate on whether the prosody of a speaker can tell us something about their personality in general.

# 5. Acknowledgments

This project started out as a class project conducted by Rachel Morasse in my course *Laboratory Linguistics* in the fall of 2009. We thank the members of prosodylab for their help in conducting this project, and acknowledge very helpful discussions about this project with Meghan Clayards, Dan Goodhue, Oriana Kilbourn-Ceron, Chris Potts, Morgan Sonderegger, Martina Wiltschko, as well as comments from the P* reading group at McGill. This project was funded by SSHRC grant 410-2011-1062 and funding through the Canada Research Chair program.

# 6. References

[1] P. Brown and S. Levinson, *Politeness: Some universals in language usage*. Cambridge University Press, 1987.

[2] D. Wilson and D. Sperber, "Relevance theory," in *Handbook of Pragmatics*, L. Horn and G. Ward, Eds. Oxford: Blackwell, 2005.

[3] C. Potts, "Presupposition and implicature," in *The handbook of contemporary semantic theory*, 2nd ed., S. Lappin and C. Fox, Eds. Oxford: Wiley–Blackwell, 2015, pp. 168–202.

[4] H. Grice, *Studies in the Way of Words*. Cambridge, Ma.: Harvard University Press, 2005.

[5] N. Constant, C. Davis, C. Potts, and F. Schwarz, "The pragmatics of expressive content: Evidence from large corpora," *Sprache und Datenverarbeitung*, vol. 33, no. 1-2, pp. 5–21, 2009.

[6] R. Ogden, "Phonetics and social action in agreements and disagreements," *Journal of Pragmatics*, 2006.

[7] R. Banse and K. Scherer, "Acoustic profiles in vocal emotion expression." *Journal of personality and social psychology*, vol. 70, no. 3, p. 614, 1996.

[8] B. Medlock and G. McCulloch, "The lingjuistic secrets found in billions of emojis," 2016, talk presented at South-by-South West. [Online]. Available: http://www.slideshare.net/SwiftKey/the-linguistic-secrets-found-in-billions-of-emoji-sxsw-2016-presentation-59956212

[9] R. Cowie and R. Cornelius, "Describing the emotional states that are expressed in speech," *Speech Communication*, vol. 40, no. 1, pp. 5–32, 2003.

[10] J. Liscombe, J. Venditti, and J. Hirschberg, "Classifying subject ratings of emotional speech using acoustic features," in *Proceedings of EUROSPEECH*. Citeseer, 2003, pp. 725–728.

[11] K. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech communication*, vol. 40, no. 1, pp. 227–256, 2003.

[12] V. Aubergé and M. Cathiard, "Can we hear the prosody of smile?" *Speech Communication*, vol. 40, no. 1, pp. 87–97, 2003.

[13] A. Wichmann, "Attitudinal intonation and the inferential process," in *Proceedings of Speech Prosody*, 2002.

[14] S. Jeong and C. Potts, "Intonational sentence-type conventions for perlocutionary effects: An experimental investigation," 2016, abstract for a paper to be presented at SALT 26.

[15] C. Gussenhoven, *The phonology of tone and intonation*. Cambridge: Cambridge University Press, 2004.

[16] M. Pell and V. Skorup, "Implicit processing of emotional prosody in a foreign versus native language," *Speech Communication*, vol. 50, no. 6, pp. 519–530, 2008.

[17] S. Paulmann, D. Titone, and M. Pell, "How emotional prosody guides your way: Evidence from eye movements," *Speech Communication*, vol. 54, no. 1, pp. 92–107, 2012.

[18] M. Franke, *Signal to act: Game theory in pragmatics*. Institute for Logic, Language and Computation, 2009.

[19] L. Bergen, R. Levy, and N. D. Goodman, "Pragmatic reasoning through semantic inference," 2014, ms. MIT/UCSD/Stanford.

[20] S. Shattuck-Hufnagel and A. E. Turk, "A prosody tutorial for investigators of auditory sentence processing," *Journal of Psycholinguistic Research*, vol. 25, no. 2, pp. 193–247, 1996.