

## On the role of context and prosody in the generation of scalar implicatures of adjectives

Utterances of sentences with weak, gradable adjectives, like *attractive*, may give rise to a scalar implicature denying a logically stronger alternative adjective meaning (e.g. “attractive, but not beautiful”). Our work explores the role of context and prosody in the generation of such implicatures, i.e. on the degree to which listeners infer from utterances of sentences with weak, gradable adjectives that the speaker intended to convey a scalar implicature. Exp 1 shows that utterances with weak adjectives that realize indirect answers to polar questions lead to a stronger degree of belief in the scalar implicature when the utterance was realized in the ‘rise-fall-rise’ contour than when the utterance was realized in a ‘neutral’ contour. Exp 2 replicates this finding for a different context, namely when the speaker utters the sentence with the weak adjective, but the speaker is known to be committed to the stronger adjective meaning. Exp 3 provides evidence that context affects the interpretation of utterances with weak adjectives: the degree of belief in the scalar implicature is weaker if the utterance is made in the context of Exp 2 than if it is made in the context of Exp 1 or in an out-of-the-blue context.

**Exp 1 (n=57):** Speakers of American English listened to dialogues between Mike and Julie, who they had been told were catching up over coffee (in all three experiments, participants were recruited via Amazon’s Mechanical Turk platform). The target dialogues were created from 16 adjective pairs. In ‘SW’ dialogues, the stronger adjective was realized in the question and the weaker one was in the answer (*Is your sister beautiful? – She is attractive.*). Julie’s answers were realized with a ‘neutral’ (H\* L-L%, in ToBI notation) or the ‘rise-fall-rise’ (L\*+H L-H%) contour. A separate norming study established that the adjective pairs in the dialogues are interpretable on the same scale, with one adjective stronger than the other. ‘WS’ dialogues served as a control condition: here, the weaker adjective was realized in Mike’s question and the stronger adjective in Julie’s answer, which was only realized in the neutral contour (*Is your sister attractive? – She is beautiful.*); we expected WS dialogues to lead to more positive answers by the participants than the SW ones. Participants were asked to judge, on a scale from 1 (“definitely no”) to 7 (“definitely yes”), whether Julie meant the adjective in Mike’s question (e.g. *Did Julie mean that her sister is beautiful?* for the SW dialogue).

**Results:** Mixed-effects ordinal regression models predicting response with by-participant and by-item random intercepts and random slopes for prosody shows significant differences between the SW and the WS dialogues ( $p < 0.001$ ) and between the two SW dialogues ( $p < 0.001$ ), with more negative responses (indicating a higher degree of belief in the implicature) in the WS than the SW dialogues, and, in the SW dialogues, when Julie’s answer was realized with the “rise-fall-rise” contour than with the “neutral” one.

**Exp 2 (n=33):** Target stimuli consisted of Julie’s utterances from the SW dialogues from Exp 1. In each trial, participants were told that Julie thinks that the strong adjective is true (e.g. the sister is beautiful) and then listened to an utterance with the weak adjective to Mike (e.g. *She is attractive*). Participants judged whether what Julie said is true (“yes”) or not (“no”).

**Results:** A logistic regression model predicting response with by-participant and by-item random intercepts shows significant differences between the ‘neutral’ and the “rise-fall-rise” contours ( $p < 0.01$ ), with more negative responses when Julie’s answer was realized with the rise-fall-rise contour than with the neutral one. Exps 1 and 2 thus showed that the prosodic realization of the utterance containing the weak, gradable adjective affects the degree of belief in the scalar implicature across two types of contexts in which such utterances are made.

**Exp 3 (n=80):** Target stimuli were created from eight adjective pairs that had the highest implicature rates in the SW dialogues with the neutral prosody in Exp 1. In each of the three conditions, experiment participants read an utterance by Julie with the weak adjective. The three conditions differed in the context in which this utterance was made. In the first ‘indirect answer’ condition (cf. Exp 1), participants read a short dialogue between Mike and Julie, where Mike’s question contained the strong adjective and Julie’s indirect answer the weak adjective. Participants were asked the same question as in Exp 1 and indicated their response on a 7-point Likert scale labeled at four points (1/No, 3/Possibly no, 5/Possibly yes, 7/Yes). In the second ‘verification’ condition (cf. Exp 2), participants read a statement about what Julie is committed to, namely the meaning of the strong adjective, and then Julie’s utterance to Mike containing the weak adjective. Participants judged whether Julie’s utterance was true and indicated their response on the 7-point Likert scale used in the previous condition. In the third ‘out-of-the-blue’ condition, participants were not given a context but only read Julie’s utterance to Mike with the weak adjective. Participants were asked whether they would infer that Julie denies the strong adjective meaning (e.g. *Would you infer from this that Julie thinks that her sister is not beautiful?*; see e.g. Geurts & Pouscoulous 2009, van Tiel et al 2014 for this paradigm). Responses were given on the same 7-point Likert scale. Each participant was randomly assigned to one condition, and the stimuli were presented at random to each participant.

**Results:** An ordinal mixed-effects model predicting response with by-participant and by-item random intercepts shows a significant difference between the verification and the indirect answer conditions ( $p < .05$ ), with lower responses (i.e. more implicatures) in the latter condition, but not between the verification and the out-of-the-blue condition ( $p = .47$ ). While responses in the indirect answer condition were lower than in the out-of-the-blue condition (implicature rates 19% versus 14%, respectively), this difference was not significant.

**Discussion:** Our results show that whether an utterance with a weak, gradable adjective is taken to give rise to a scalar implicature denying a stronger adjective meaning is affected by the prosody of the utterance and the context in which the utterance is made. That prosody affects scalar implicature generation was already established in prior literature: when scalar items like *some* or *or* are realized with a pitch accent, the listener’s degree of belief in the scalar implicatures (“some, but not all” or “or, but not both”) is higher than when the scalar item is not realized with a pitch accent (i.a. Chevalier et al. 2008, Thorward 2009, Zondervan 2010, Schwarz et al. to appear). Our Exps 1 and 2 establish that not just the presence/absence of a pitch accent matters (in our stimuli, the adjective was always realized with a pitch accent), but also the prosodic contour of the utterance containing the scalar item.

Exp 3 addresses the role of context on the interpretation of utterances with weak, gradable adjectives. Our results expand on those of Geurts & Pouscoulous (2009) and Degen & Goodman (2014), who find that the degree of belief in the scalar implicature with utterances of *some* and other determiners is context- (or paradigm-)dependent. Our findings indicate that the degree of belief in the scalar implicature is highest in the indirect answer condition. We hypothesize that this is because the listener recognizes that the speaker, rather than just responding “yes” or “no” to the polar question, uttered a weaker adjective, thus providing a clue to the listener that she is not committed to the stronger adjective. In the verification condition, on the other hand, the listener knows that the speaker is committed to the stronger adjective, which entails the weaker one: although the listener may reason about why the speaker uttered the weaker adjective (i.e. violated Quantity), it seems to be rare that this reasoning leads the listener to conclude that the speaker is not committed to the stronger adjective after all.