What do you mean, you're uncertain?: The interpretation of cue words and rising intonation in dialogue

Catherine Lai

Department of Linguistics
University of Pennsylvania

Interspeech 2010
Sept 28, 2010

What's this about?

- Expressions of speaker attitude, like surprise, uncertainty, and agreement, help determine the structure of a dialogue.
- We see this manifest in the various attitude related strategies speakers employ to shape the discourse.
  - Overt linguistic markers: e.g. question syntax, verbs of prop. attitude ('know', 'doubt'), cue words (*really*?)...
  - Prosody: e.g. rising intonation, pitch range...
- How can we model what these actually do to a discourse? At what level do they work? What sort of meaning does prosody convey?

What's the meaning of this?

This talk is about cue words and rising intonation.

- What effect do cue words and rises have with respect to discourse structures?
- How is this related to the perception of attitudes like uncertainty?
- How does the gradability of prosody and cue word semantics relate to gradability of belief?

$\Rightarrow$ Probe these questions with a perception experiment.

Cue words in dialogue

We model speaker's public beliefs and the Question Under Discussion (QUD), c.f. Farkas and Bruce (2009); Ginzburg (2009).

(1) a. B: Do you like Lubbock better than Dallas? ($= ?p_1$)
b. A: Yeah
c. B: Why?
d. A: Uh, because people are so much nicer ($= p_2$)

(Switchboard Corpus: LDC2004T12)

	Public(A)	QUD	Public(B)
(a)		p_1?	
(b)	p_1		
(c)		Why p_1?	
(d)	p_2	p_2?	

e. B: right
B: yeah
B: okay
B: uh-huh
B: really?
B: well...
B: No!

(e) depends on the cue word semantics and prosody. Let's focus on rises...

What about Rises?

- Rises have been linked to the perception of uncertainty (Pon-Barry, 2008; Litman et al., 2009; Gravano et al., 2008).
- Formally, rises have analyzed as requesting hearer commitment or responsibility (Gunlogson, 2008), or a test on the common ground (Nilsenova, 2006) with respect to the content 'under' the rise.

  $\rightsquigarrow$ Implication of speaker uncertainty.
- However, backchannels interpretations of affirmative cue words, e.g. *okay*, are distinguished by rising pitch (Benus et al., 2007) and pitch upturn is employed to encourage the interlocutor to continue speaking (Ward and Escalante-Ruiz, 2009).

  $\rightsquigarrow$ Not really cases of speaker uncertainty

In all these cases, the rise-speaker seems to want the hearer to talk more.

This Experiment

How does cue word semantics interact with rising intonation?

- Hypotheses:
  - Rises signal that the current question under discussion is unresolved.
  - The underlying semantics of the utterance constrains how a rise is interpreted.
- Rather than ask directly about the QUD, we consider:
  - EXPECTEDNESS reflects certainty with respect to B's prior beliefs. (c.f. Lai (2009) the relationship of pitch range and surprise.)
  - CREDIBILITY reflects how willing B is to believe A, i.e. add the content of A's utterance to their public beliefs.
  - EVIDENCE reflects the status of the QUD, i.e. whether A's utterance has been resolved/accepted or whether it is still contentious.

We can then also relate uncertainty to different aspects of dialogue structure.

Stimuli

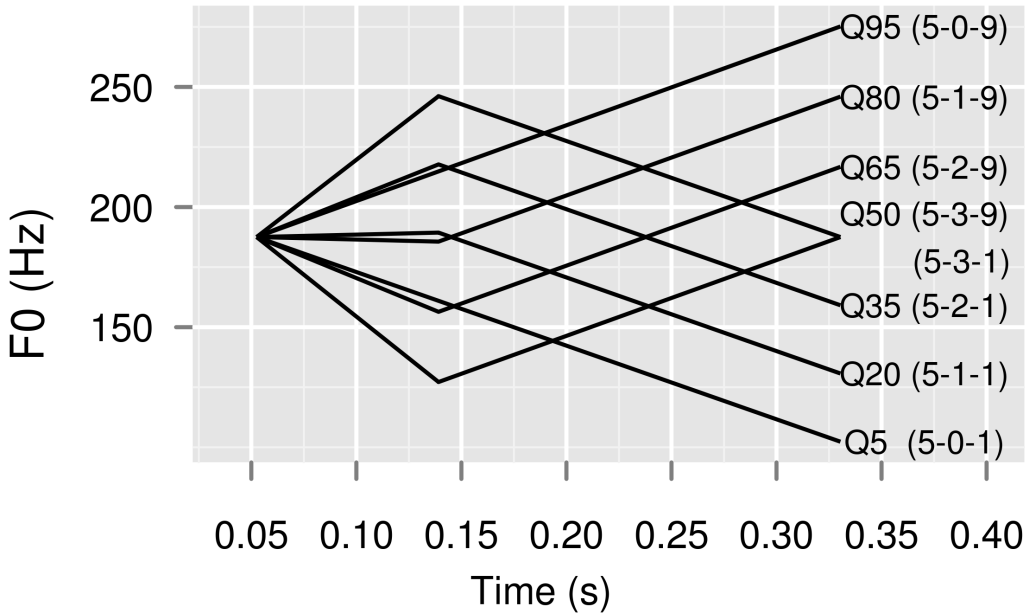
In this experiment, subjects to evaluate context + resynthesized cue word pairs with respect to EXPECTEDNESS, CREDIBILITY, EVIDENCE.

- Cue words from Switchboard II (LDC97S62):
  - $2 \times \{$*really*, *well*, *okay*, *sure*, *yeah*, and *right* $\}$
  - one word turns according to the transcripts.
  - checked for voice quality
- Contexts were drawn from turns immediately preceding one of the cue words, representing different levels of certainty (not exhaustive!)
  - factual, e.g *X is Y*,
  - evaluative, e.g. *X is good*,
  - attributed, e.g. *I heard that X*,
  - inferred e.g. *probably X*.

Resynthesis 8 ways

For each base token:

- F_0 values were based on quantiles of F_0 values of the speaker for that conversation.
- The start point was the median value and the gradient between the mid- and endpoints remained the same.
- Timing was set with respect to the start, end, and the midpoint of the stressed vowel (manually identified).



- Varies overall pitch range and peak height but not slope.
- Test whether pitch range $\propto$ unexpectedness.

The Task

14 native speakers of American English, undergraduate students, paid, were asked to:

- Read the context: e.g. *the book was just ever so much better*
- Listen to the response: e.g. really (right)
- Answer the following questions (1-7 scale):
  - How expected does what A said seem to B?
    (1=completely unexpected, 7=completely expected)
  - How credible does what A said seem to B?
    (1=not at all credible, 7=completely credible)
  - Given B's reaction, how much would you expect A to explain or provide more evidence for what they say/why they said it?
    (1=wouldn't expect a follow up, 7=definitely expect a follow up).

$\Rightarrow$ 6x2x8 = 96 cue words and 6x4x4 = 96 contexts

Experiment Design

- Written context and audio (with text) response with replay enabled.
- Contexts and responses were randomly paired.
- 4 practice slides, 64 main experiment slides (human error!)

Results: Means

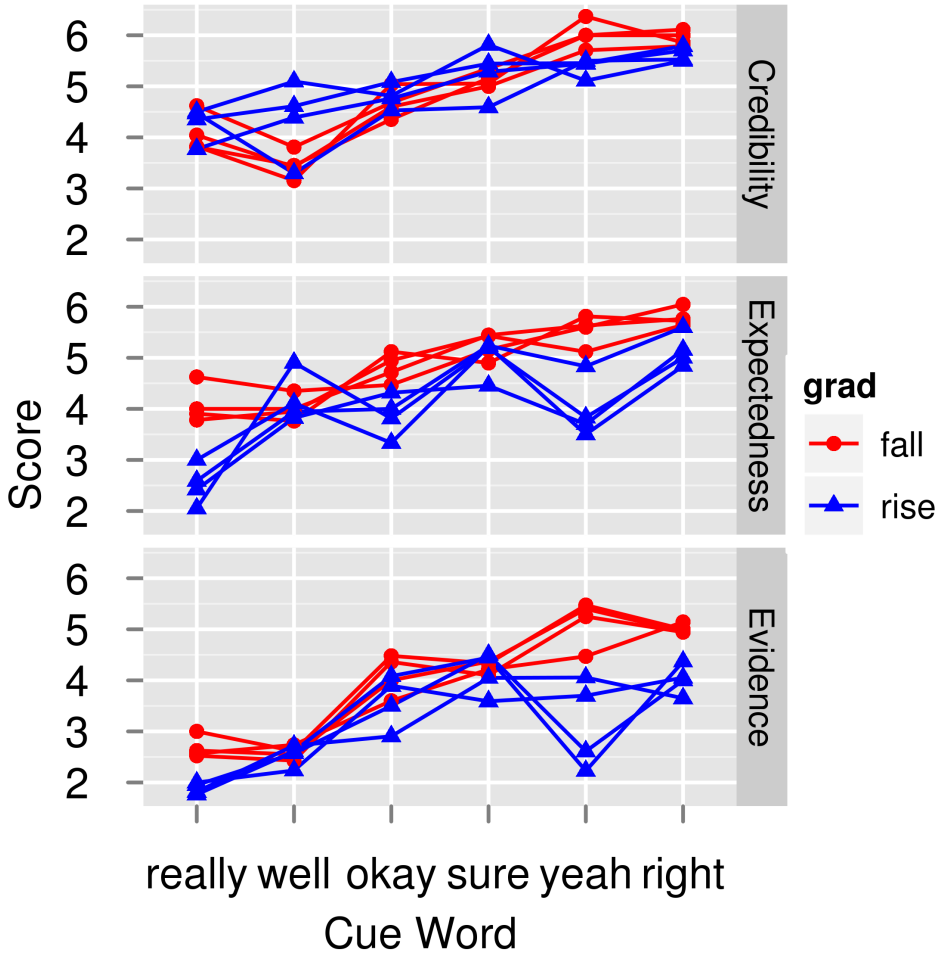


Figure: Mean scores for each cue word by question (question 3 reversed).

Multilevel Model

- Model the effects of cue words, contours, contexts, subjects and the cue word/contour interaction as arising from different normal distributions (groups).
- The model parameters, along with finite population standard deviations for each group, were estimated using the Markov Chain Monte Carlo technique (JAGS)
- This gives us distribution rather than a point estimate!

Multilevel Model

- Following Gelman and Hill (2007), for each question the observed scores, y, for each question were modelled as follows.

$$y_i \sim \mu + \alpha^{cw}_{j[i]} + \alpha^{ct}_{k[i]} + \alpha^{cx}_{l[i]} + \alpha^{s}_{m[i]} + \alpha^{cw.ct}_{j[i],k[i]} \quad (1)$$

$$\alpha^{cw}_{j} \sim N(0, \sigma^2_{cw}) \text{ for } j = 1, \ldots, 6 \quad (2)$$

$$\alpha^{ct}_{k} \sim N(0, \sigma^2_{ct}) \text{ for } k = 1, \ldots, 8 \quad (3)$$

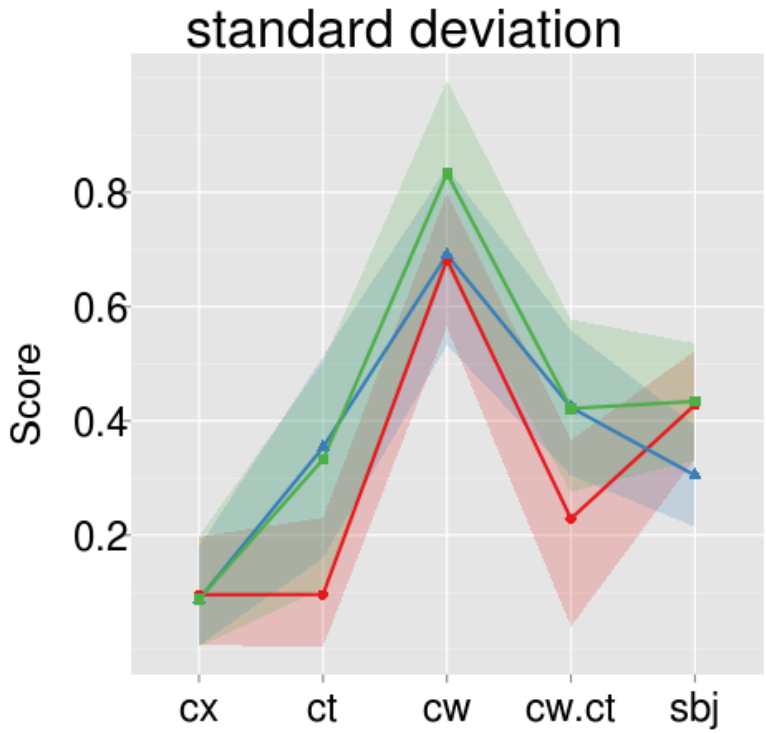
$$\alpha^{cx}_{l} \sim N(0, \sigma^2_{cx}) \text{ for } l = 1, \ldots, 4 \quad (4)$$

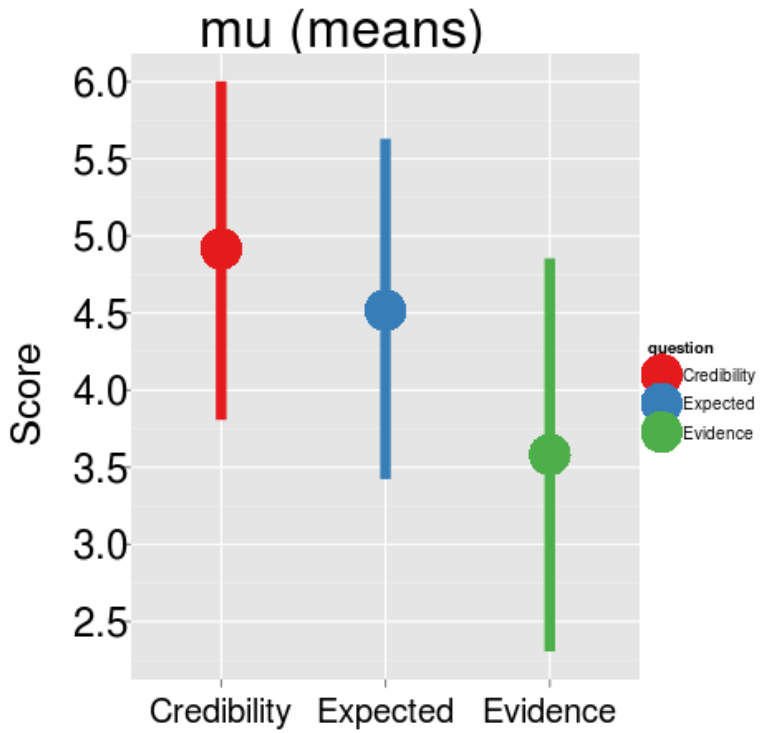
$$\alpha^{s}_{m} \sim N(0, \sigma^2_{s}) \text{ for } m = 1, \ldots, 14 \quad (5)$$

$$\alpha^{cw.ct}_{j,k} \sim N(0, \sigma^2_{cw.ct}) \text{ for } j = 1, \ldots, 6, k = 1, \ldots, 8 \quad (6)$$

- e.g. α^{cw}_{k} is a parameter representing the effect of cue word k holding the other variables constant.
- Let's look at estimated medians and 95% intervals for the different parameters for each of the scales.

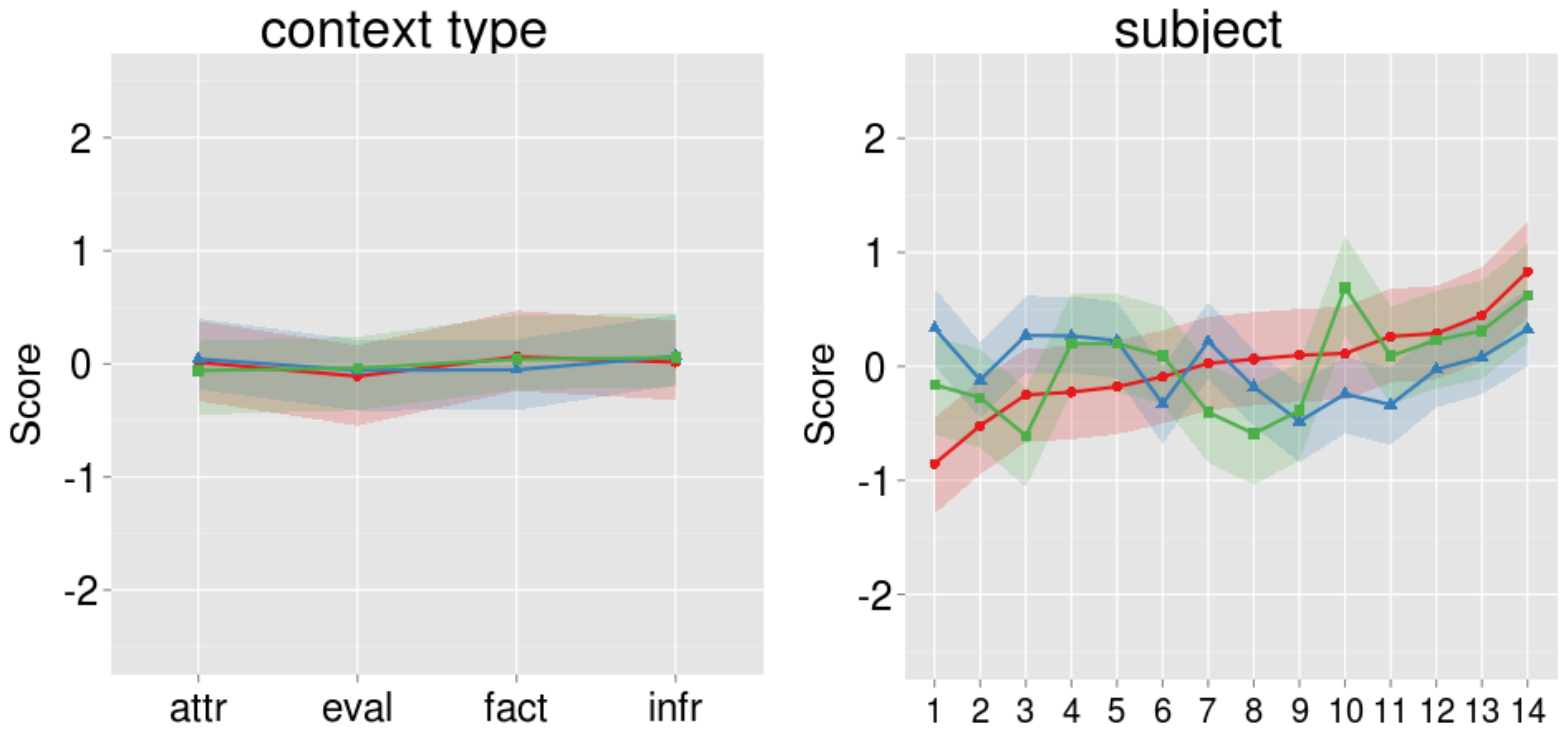
Parameter estimates





- Dot = median, shaded region = 2.5th-97.5th quantiles.
- Biggest standard deviation estimate comes from the cue word itself.
- Contour has more of an effect on EXPECTEDNESS and EVIDENCE.

Parameter Estimates

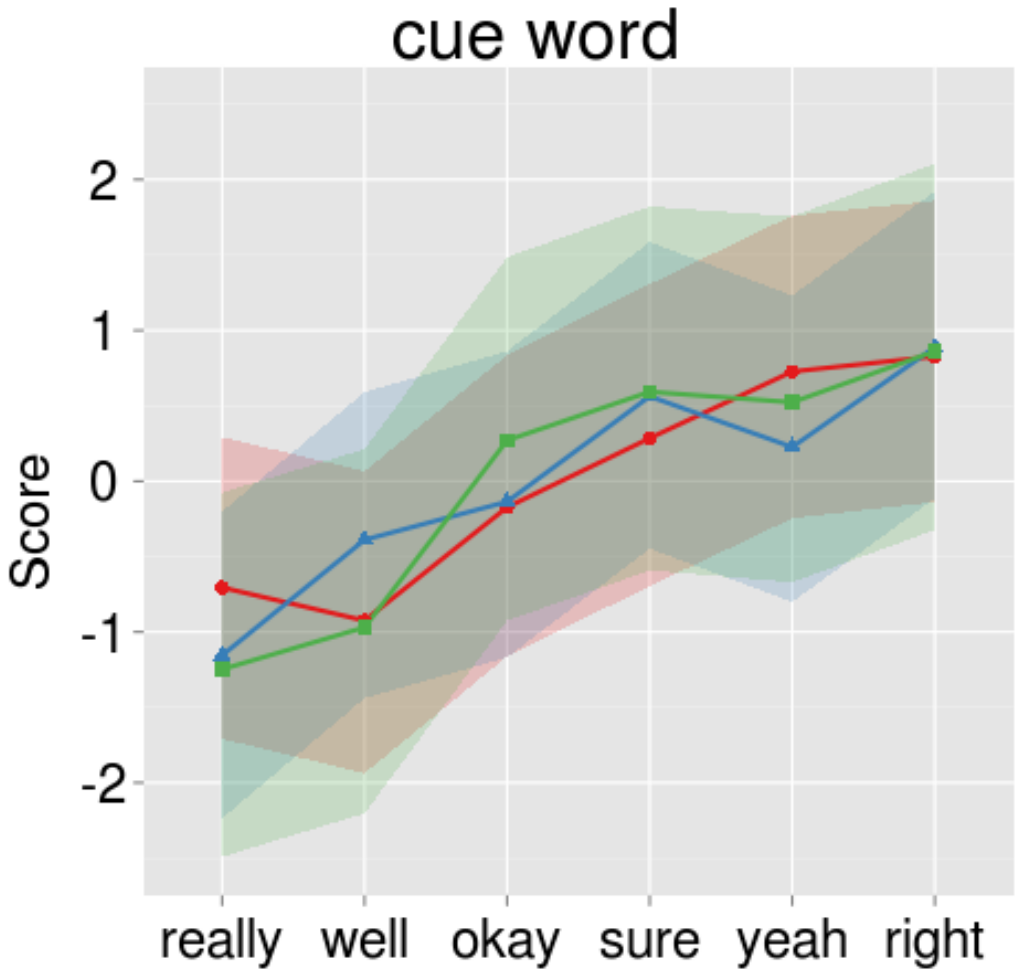


- Contexts don't have much of an effect: estimates are small and fall well inside the 95% intervals of the other type.
- Subjects have different strategies/biases.

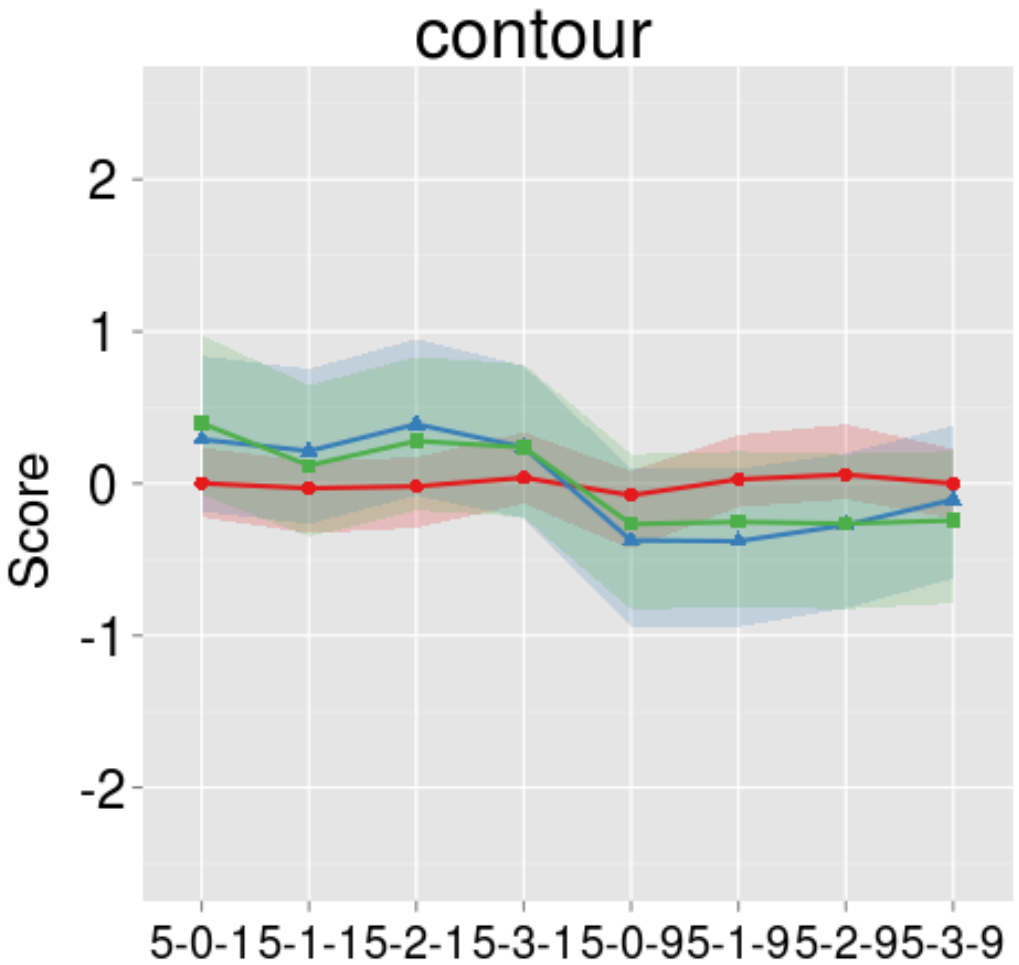
Now abstracting away from this...

Parameter Estimates

- We get a credibility ordering over cue words.
- e.g. *right* is a strong agreement word.



Parameter Estimates

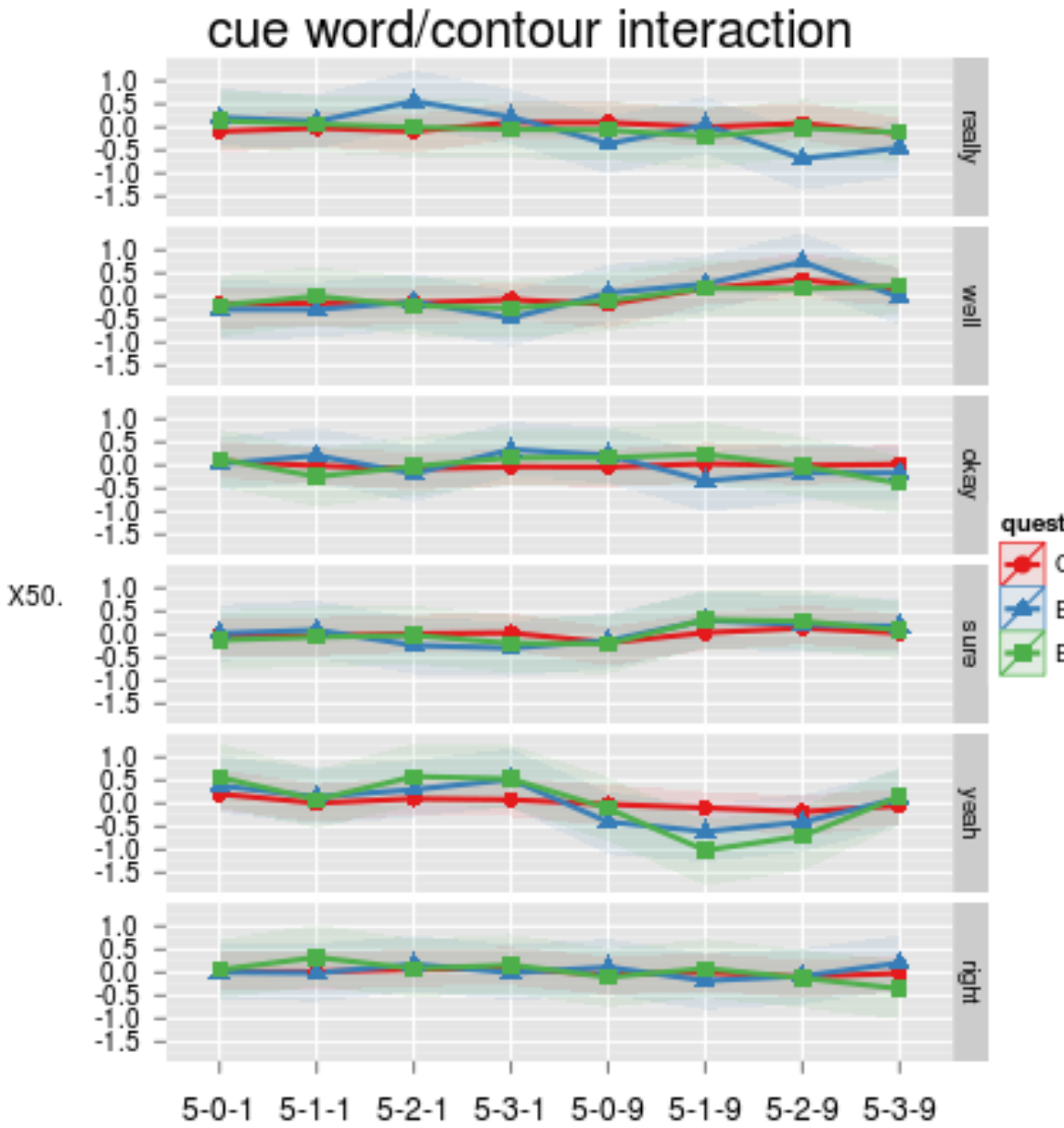


- Rising intonation lowers EXPECTEDNESS and EVIDENCE scores, but not CREDIBILITY.
- Posteriors associated with falls and rises appear quite distinct, medians for rises generally lying below the 2.5th quantile of the falls.

Parameter Estimates

Variation across cue words:

- *yeah* can express more unexpectedness then *right*.
- *yeah*'s semantics is not as strong/specific ⇝ prosody more influential.
- *really*: variation appears to be mostly on the EXPECTEDNESS scale (c.f. Lai (2009)).



The Interpretation of Rises

- Intonation did not have much of an effect on the credibility scale.
  - Rises reflect difficulty integrating the new information rather than expressing disbelief.
  - Credibility is clearly reflected in the choice of cue word
- Rises signal that question under discussion is unresolved, implicitly signalling that resolution depends on the hearer:
  - congruent with the rising intonation of affirmative backchannels (turn passing),
  - signal the expectation that more evidence will be presented
  - they do not necessarily make an utterance an interrogative!

Unexpectedness (Surprise!)

- For cue words, inability to resolve the QUD may arise due to
  - epistemically unexpected (i.e. it doesn't fit their world view)
  - unexpected from the point of view of relevance.
    e.g. *right*: the respondent may agree with the content, while still feeling that it does not resolve the current QUD.
- Greater overall pitch ranges were not really associated with the perception of more unexpectedness/surprise.

$\rightsquigarrow$ the connection between pitch range and surprise may be more to do with slope or peak position rather than a max-min measure.
But resynthesis was based on quantiles, so no strong conclusions about individual contours across cue words.

Conclusion

How can we analyze prosody? We need to know its linguistic function.

- Rising intonation works at the discourse/dialogue management level: it signals that the current QUD is unresolved.
  $\rightsquigarrow$ Co-operative interlocutors should try to resolve it!
  $\rightsquigarrow$ Conversational dialogue systems should evaluate utterances with rising intonation with respect to the QUD
- Cue words form a scale of CREDIBILITY
  $\rightsquigarrow$ Track other conversational participants public beliefs.
  $\rightsquigarrow$ Determine which type of cue word to use and when.
- To investigate the relationship between prosodic gradability and speaker attitude we need to understand the semantic/pragmatic dimensions involved. This study is another step towards this.

Further Work

- What's the contribution of pitch slope? plateaus?
- What about larger utterances? $\rightsquigarrow$ try VERUM focus...
- What about uptalk?

Thanks!

Benus, S., Gravano, A., and Hirschberg, J. (2007). The prosody of backchannels in American English. In *Proceedings of ICPhS 2007*, pages 1065–1068.

Farkas, D. and Bruce, K. (2009). On Reacting to Assertions and Polar Questions. *Journal of Semantics*.

Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press Cambridge.

Ginzburg, J. (2009). *The Interactive Stance: Meaning for Conversation (forthcoming in 2009)*. Studies in Computational Linguistics. CSLI Publications.

Gravano, A., Benus, S., Hirschberg, J., German, E. S., and Ward, G. (2008). The effect of prosody and semantic modality on the assessment of speaker certainty. In *Proceedings of 4th Speech Prosody Conference, Campinas, Brazil*.

Gunlogson, C. (2008). A question of commitment. *Belgian Journal of Linguistics*, 22(1):101–136.

Lai, C. (2009). Perceiving Surprise on Cue Words: Prosody and Semantics Interact on *Right* and *Really*. In *Proceedings of INTERSPEECH'09, Brighton, UK, September 2009*.

Litman, D., Rotaru, M., and Nicholas, G. (2009). Classifying Turn-Level Uncertainty Using Word-Level Prosody. In *Proceedings of Interspeech'09*.

Nilsenova, M. (2006). *Rises and Falls. Studies in the semantics and pragmatics of intonation*. PhD thesis, University of Amsterdam.

Pon-Barry, H. (2008). Prosodic manifestations of confidence and uncertainty in spoken language. In *Proceedings of Interspeech'08*.

Ward, N. G. and Escalante-Ruiz, R. (2009). Using Subtle Prosodic Variation to Acknowledge the User's Current State. In *Proceedings of Interspeech'09*.