Pragmatic language interpretation as probabilistic inference

Noah Goodman and Michael Frank

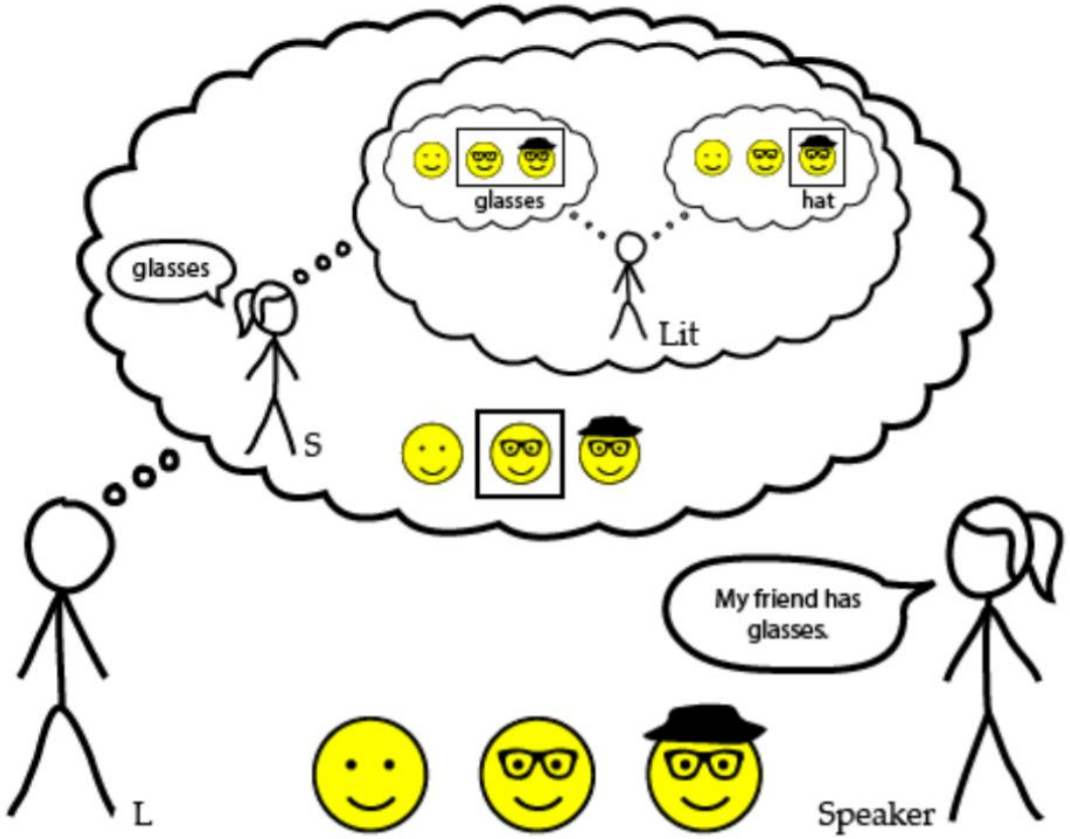
Contributions

1. RSA framework
2. “Empirical support for RSA”
3. Extension of RSA – uRSA

Grice

- Conversational maxims?
  - “Modestly useful” – Goodman and Frank
- “… one of my avowed aims is to see talking as a special case or variety of purposive, indeed *rational* behavior …” – Grice

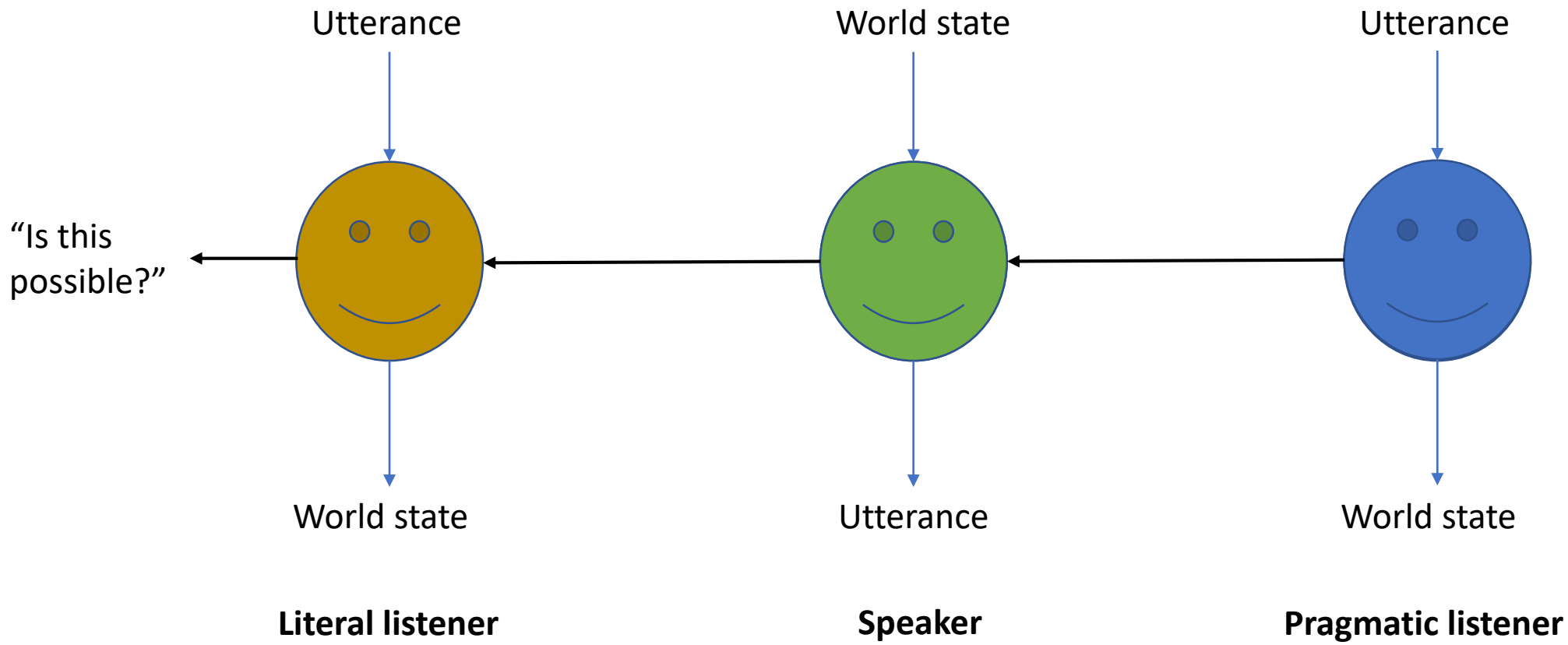
RSA



RSA

- The *pragmatic listener* reasons over the likelihood of possible world states, in terms of the *speaker's* utility (given an utterance)
- The *speaker* reasons over the utility of possible utterances, in terms of the likelihoods of the *literal listener*
- The *literal listener*, too, reasons over the likelihood of possible world states, in terms of the possibility of an utterance in the world state (given an utterance)

RSA


Utterance
World state
Utterance
"Is this possible?"
World state
Utterance
World state
Literal listener
Speaker
Pragmatic listener

RSA

$$P_L(w \mid u) \propto P_S(u \mid w)P(w)$$

$$P_S(u \mid w) \propto exp(\alpha U(u; w))$$

$$U(u; w) = \log P_{\text{Lit}}(w \mid u)$$

$$P_{\text{Lit}}(w \mid u) \propto \delta_{[\![u]\!](w)} P(w)$$

Assumptions

1. Listener and speaker are rational
    1. Listener thinks in terms of the speaker when interpreting
    2. Speaker's behavior can be modelled by a utility function
2. *Social recursion* is capped at depth two
    1. Speaker thinks in terms of a literal listener
3. Possible utterances and world states are enumerable
4. δ and P(w) are calculable

Worked example

- One man is wearing a hat and glasses
- Another man is wearing no hat and glasses
- The speaker says, “My friend is wearing glasses”
- Objective: compute the most likely world state as the listener
- We limit to two world states (equally likely)
  - The speaker’s friend is wearing a hat and glasses (HG)
  - The speaker’s friend is wearing no hat and glasses (G)
- And two utterances
  - “My friend is wearing a hat and glasses” (“HG”)
  - “My friend is wearing glasses (“G”)

Literal listener

$$P_{\text{Lit}}(w \mid u) \propto \delta_{[\![u]\!](w)} P(w)$$

u	w	$\delta_{[[u]](w)}P(w)$	$P_{lit}(w\|u)$
“HG”	HG	1/2	1
“HG”	G	0	0
“G”	HG	1/2	1/2
“G”	G	1/2	1/2

In the G world state, “HG” isn’t true.

Speaker

$$P_S(u \mid w) \propto exp(\alpha U(u; w))$$

$$U(u; w) = \log P_{\text{Lit}}(w \mid u)$$

u	w	$\delta_{[[u]](w)}P(w)$	$P_{lit}(w\|u)$	$P_s(u\|w)$
“HG”	HG	1/2	1	2/3
“HG”	G	0	0	0
“G”	HG	1/2	1/2	1/3
“G”	G	1/2	1/2	1

Assuming $\alpha=0$ gives us $P_S(u|w)$ prop. to $P_{lit}(w|u)$

In a world state of HG, the fact that “HG” would be interpreted with certainty leads the model to give preference over “G”, which is more ambiguous.

Pragmatic listener

$$P_L(w \mid u) \propto P_S(u \mid w)P(w)$$

u	w	$\delta_{[[u]](w)}P(w)$	$P_{lit}(w\|u)$	$P_s(u\|w)$	$P_L(w\|u)$
“HG”	HG	1/2	1	2/3	1
“HG”	G	0	0	0	0
“G”	HG	1/2	1/2	1/3	1/4
“G”	**G**	1/2	1/2	1	3/4

Assuming equal probability gives us $P_L(w|u)$ prop. to $P_S(u|w)$

With a world state of HG, “G” is ambiguous, but with a world state of G, “G” is the only option.

The model captures the phenomenon that the speaker would have said “HG” if the world state was “HG”!

Extension

$$P_L(w, s \mid u) \propto P_S(u \mid w, s)P(s)P(w)$$

- Joint inference of the speaker type and the world state
- Applications
  - Figurative language
  - Ambiguous adjectives
  - Ambiguous quantifiers

  “My kettle cost a thousand dollars!”

Takeaways

Positives:

- Captures phenomena like the hat and glasses example
  - Matches the way humans reason over the example
- Elegant mathematical formulation

Concerns:

- How to enumerate the world states, utterances, and speaker types?
- How to calculate δ and $P(w)$?
  - Does this model presuppose having an effective language model?
- Are we making realistic assumptions about speaker behavior?

Discussion questions

- Speaker as rational and/or not captured by utility function
  - “I don't like starting from the idea that language use is rational (and efficient) and instead would come from the position that language use "in the wild" is at least as often expressive.”
  - “…all motivation and learning is motivated by the loss function, which is never complex enough to truly guide extensible learning.”
  - “To what extent, can most conversations be characterized as rational in the sense of this paper (maximizing expected value of informativeness/relatedness after accounting for cost)?”
  - “Can we extend the RSA model to work in cases where collaboration cannot be assumed?”
  - “Isn’t it far more cognitively expensive to compute a number of alternative utterances and then compare and produce the most concise speech, as opposed to produce whichever speech comes to our head and clarify with additional sentences when asked?”
- Concerns with tractability of implementation
  - “In some real world application, we cannot have some denotation for EVERY possible speaker.”
  - “It's discrete in that it requires a fixed set of states in the world that we could maximize the probability over which is unrealistic in most computational scenarios. How do we come up with all these "things" in a natural language understanding situation?”
- Social recursion
  - What depth?