

- Two utterances: ‘some’ and ‘all’
- Two possible states/meanings: sbna and all.

## 1 Standard RSA

$L_0(w u) \propto \llbracket u \rrbracket^w P(w)$	Literal listener $L_0$
$S_n(u w) \propto \exp(\lambda \log L_{n-1}(w u)) - c(u)$	Pragmatic speaker
$L_n(w u) \propto S_n(u w)P(w)$	Pragmatic listener

We assume a completely rational speaker  $\lambda = 1$ . Since two utterances ‘some’ and ‘all’ are not so different in cost, we ignore the cost term.

### 1.1 Uniform prior

With a uniform prior,  $P(\text{all}) = P(\text{sbna}) = 0.5$ , we get the correct result.

$$\begin{aligned}
 L_0(\text{all}|\text{'all'}) &= 1 \\
 L_0(\text{all}|\text{'some'}) &= \frac{P(\text{all})}{P(\text{all}) + P(\text{sbna})} = 0.50 \\
 S_1(\text{'some'}|\text{all}) &= 0.33 \\
 L_1(\text{all}|\text{'some'}) &= 0.25 \\
 S_2(\text{'some'}|\text{all}) &= 0.2
 \end{aligned}$$

### 1.2 The problem of skewed priors

However, when the prior is skewed, we get the wrong results (Spector, Fox & Katzir, Degen et al.). For example, with  $P(\text{all}) = 0.9$ ,  $P(\text{sbna}) = 0.1$ :

$$\begin{aligned}
 L_0(\text{all}|\text{'all'}) &= 1 \\
 L_0(\text{all}|\text{'some'}) &= 0.9 \\
 S_1(\text{'some'}|\text{all}) &= 0.47 \\
 L_1(\text{all}|\text{'some'}) &= 0.81 \\
 S_2(\text{'some'}|\text{all}) &= 0.44 \\
 L_2(\text{all}|\text{'some'}) &= 0.80
 \end{aligned}$$

The values decrease here, but will not improve much. See Figure 2.

When  $P(\text{all})$  is high, the speaker is about 50% likely to use ‘some’ to mean all (because they almost randomly choose between ‘all’ and ‘some’), and the listener is very likely to interpret ‘some’ as all (because it’s more likely to be true).

## 2 Stalnakerian Listener

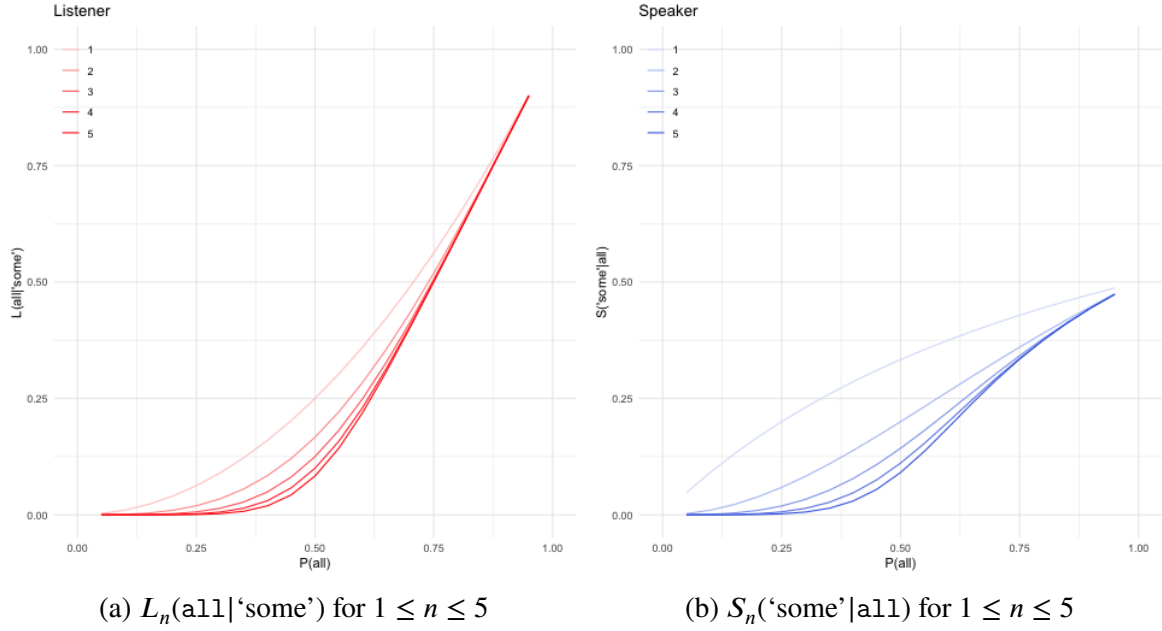
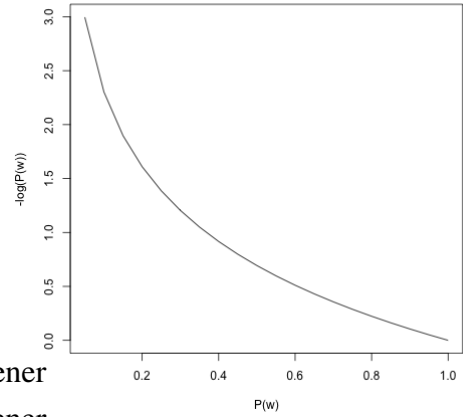


Figure 1: Probabilities of 'some' meaning all as a function of  $P(\text{all})$  in RSA.

The intuition is that when  $P(\text{all})$  is high, it does not need to be linguistically conveyed that all is the case, because the utterance would be underinformative. Knowing that the speaker would avoid redundancy, the rational listener should prefer the most informative interpretation of the utterance.

In order to implement this, let's make use of the *surprisal* (information content) of learning that  $w$  is the case,  $-\log P(w)$ .<sup>1</sup>

$$\begin{aligned}
 L_0(w|u) &\propto -\log P(w) \times \llbracket u \rrbracket^w P(w) && \text{Literal listener} \\
 L_n(w|u) &\propto -\log P(w) \times S_n(u|w)P(w) && \text{Pragmatic listener} \\
 S_n(u|w) &\propto \exp(\lambda \log L_{n-1}(w|u)) && \text{Pragmatic speaker}
 \end{aligned}$$



With a uniform prior,  $P(\text{all}) = P(\text{sbna}) = 0.5$ , the results are not so different from the standard RSA.

$$\begin{aligned}
 L_0(\text{all}|\text{'some'}) &= 0.50 \\
 S_1(\text{'some'}|\text{all}) &= 0.33 \\
 L_1(\text{all}|\text{'some'}) &= 0.25 \\
 S_2(\text{'some'}|\text{all}) &= 0.20 \\
 L_2(\text{all}|\text{'some'}) &= 0.17
 \end{aligned}$$

<sup>1</sup>Using the same Literal Listener as the standard RSA only changes the initial level.

The results are better with the skewed prior,  $P(\text{all}) = 0.9$ ,  $P(\text{sbna}) = 0.1$ :

$$L_0(\text{all}|\text{'some'}) = 0.29$$

$$S_1(\text{'some'}|\text{all}) = 0.23$$

$$L_1(\text{all}|\text{'some'}) = 0.09$$

$$S_2(\text{'some'}|\text{all}) = 0.08$$

$$L_2(\text{all}|\text{'some'}) = 0.03$$

However, we now have a problem with a different skewed prior,  $P(\text{all}) = 0.1$ ,  $P(\text{sbna}) = 0.9$ :

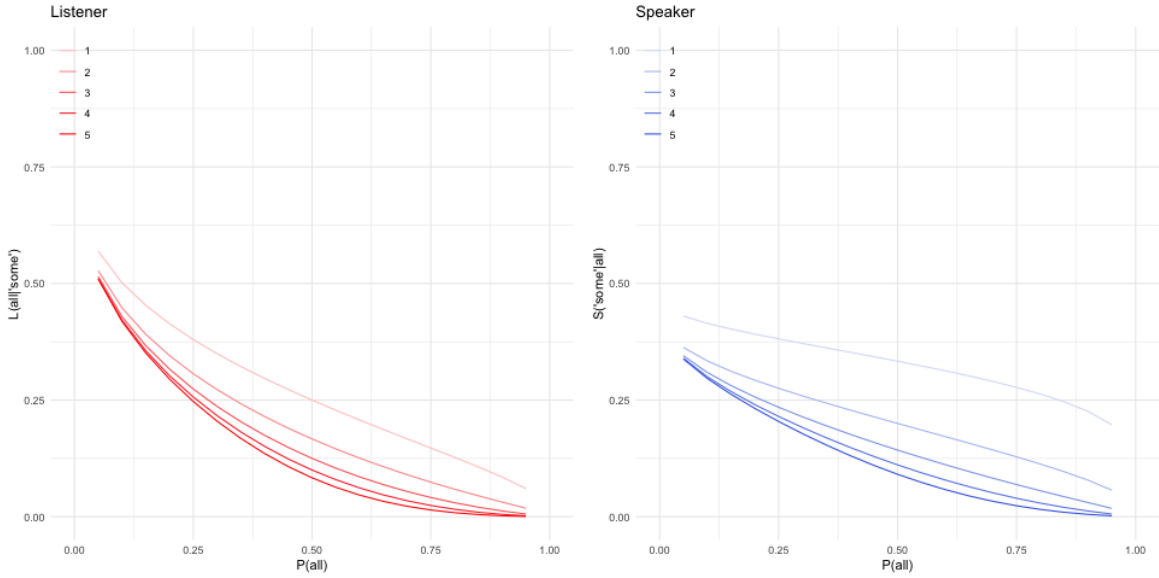
$$L_0(\text{all}|\text{'some'}) = 0.71$$

$$S_1(\text{'some'}|\text{all}) = 0.41$$

$$L_1(\text{all}|\text{'some'}) = 0.50$$

$$S_2(\text{'some'}|\text{all}) = 0.33$$

$$L_2(\text{all}|\text{'some'}) = 0.45$$



(a)  $L_n(\text{all}|\text{'some'})$  for  $1 \leq n \leq 5$

(b)  $S_n(\text{'some'}|\text{all})$  for  $1 \leq n \leq 5$

Figure 2: Probabilities of ‘some’ meaning all as a function of  $P(\text{all})$  in RSA with the Stalnakerian Listener.

In this model, when  $P(\text{all})$  is low, it makes sense to mean all with ‘some’, because all is highly informative while sbna is not. Knowing this, the speaker can choose ‘some’ to express all.

### 3 Maxent Speaker

When  $P(\text{all})$  is low, the speaker shouldn’t use ‘some’ to express all, because it could potentially receive an underinformative reading, while ‘all’ is a solid means of expressing it because it always means all.

The speaker model of the standard RSA doesn't capture this, because the speaker just chooses the most reliable message to convey  $w$  with respect to the listener model  $L$ , by maximizing the negative surprisal  $\log L(w|u)$ . So the speaker doesn't directly consider what the alternative meaning would do.

With the Stalnakerian Listener, this is indirectly captured. They try to avoid underinformative meaning, so when the Speaker maximizes the negative surprisal  $\log L(w|u)$ , the relatively more informative meaning will become more likely.

But I think it makes sense to directly encode this consideration in the speaker model. The idea is that the speaker tries to maximize the average surprisal, or *entropy*,  $H(u|L)$ , of message  $u$  relative to the listener model  $L$ .

$$S_n(u|w) \propto \exp(\underbrace{\lambda(\log L_{n-1}(w|u) + \underbrace{H(u|L_{n-1})}_{\text{entropy wrt } L_{n-1}}))_{\text{reliability}}$$

The entropy of message  $u$  with respect to the listener model  $L$  is defined as the following cross entropy:

$$H(u|L) = \sum_{w'} \underbrace{-\log P(w')}_{\text{impact on CG}} \times \underbrace{L(w'|u)}_{\text{Probability of } u \text{ meaning } w' \text{ wrt to } L}$$

This model, however, doesn't completely solve the problem. See Figure 3 (It's computationally heavy, so I could only compute up to Level 4).

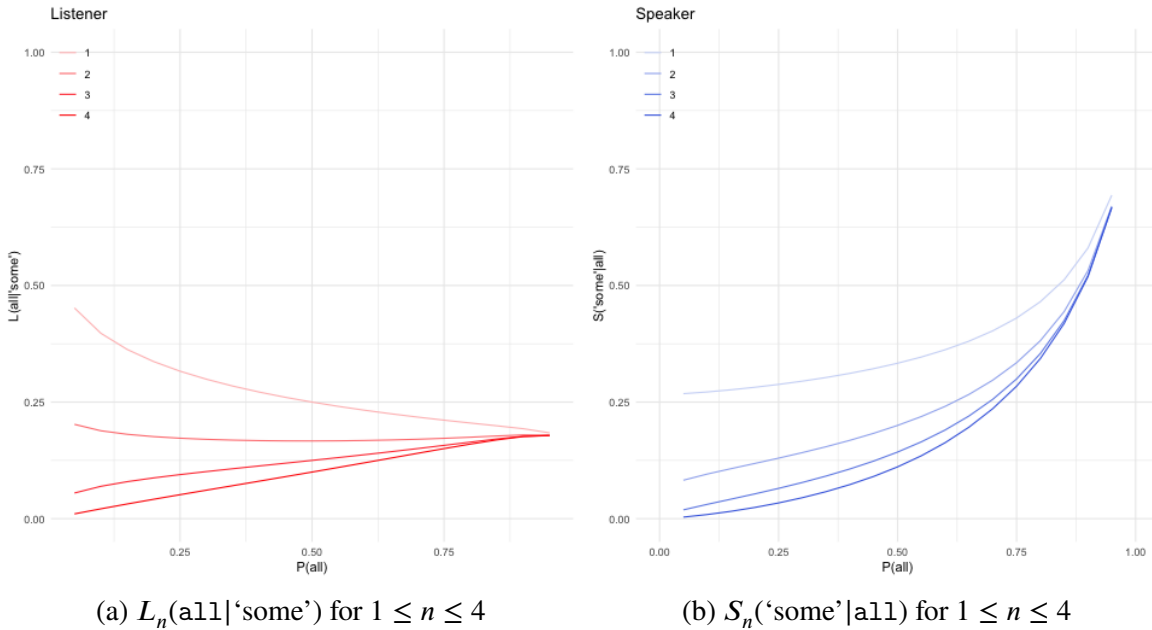


Figure 3: Probabilities of 'some' meaning all as a function of  $P(\text{all})$  in RSA with the Stalnakerian Listener and Maxent Speaker.

In particular, it recreates the initial problem for the speaker. The reason is because the utility of meaning all is so high that this consideration overrides the reliability of the listener understanding 'some' as all. To make the impact of the entropy smaller, let's regulate it via a

constant  $k$ .

$$S_n(u|w) \propto \exp(\underbrace{\lambda(\log L_{n-1}(w|u) + k}_{\text{reliability}} \cdot \underbrace{H(u|L_{n-1})}_{\text{entropy wrt } L_{n-1}}))$$

With  $k = 0.5$ , we obtain better results: But the numbers are still not very satisfactory.

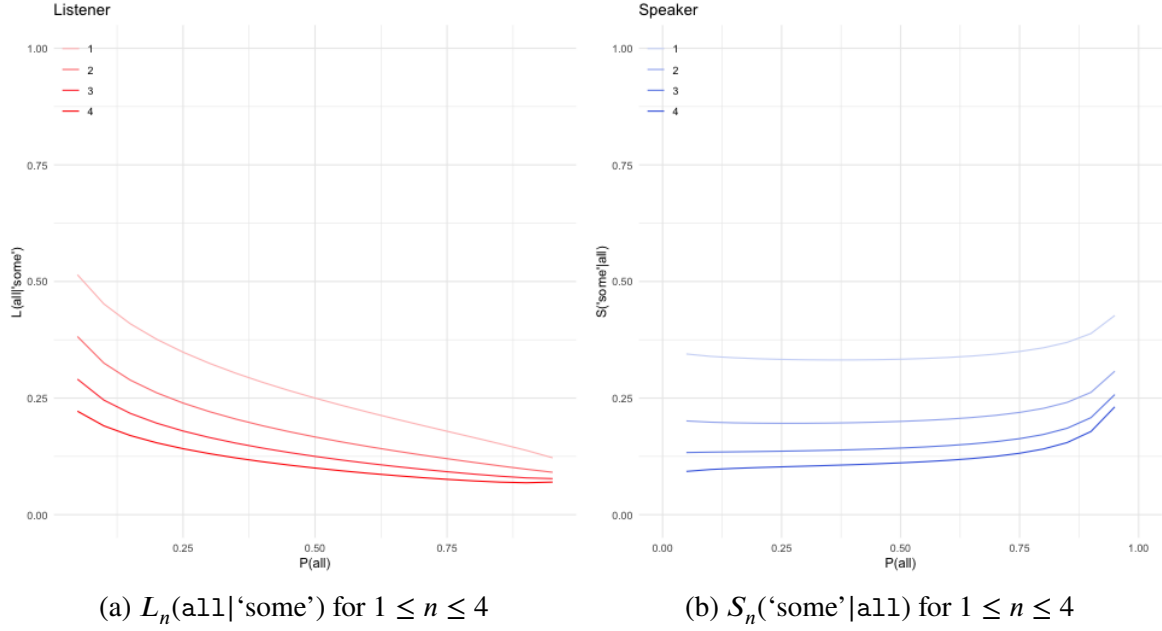
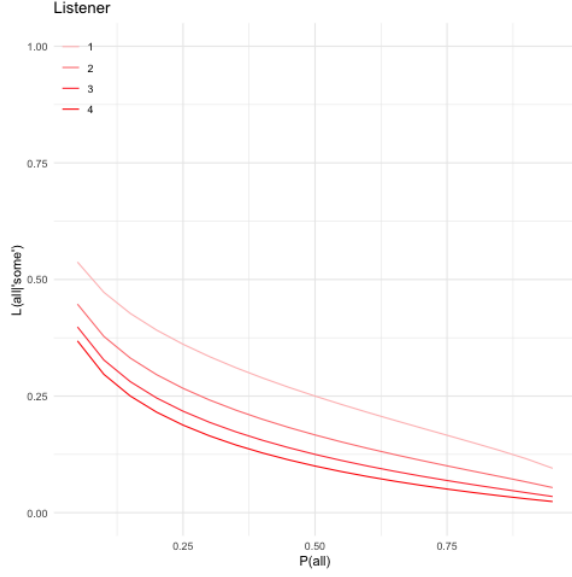
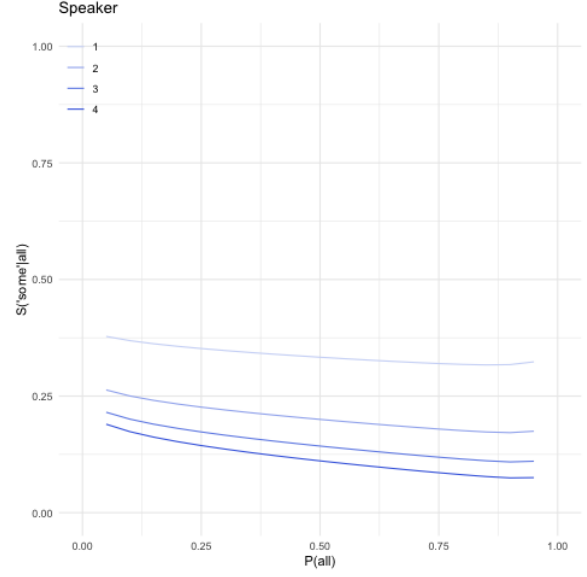


Figure 4: Probabilities of ‘some’ meaning all as a function of  $P(\text{all})$  in RSA with the Stalnakerian Listener and Maxent Speaker with  $k = 0.5$ .

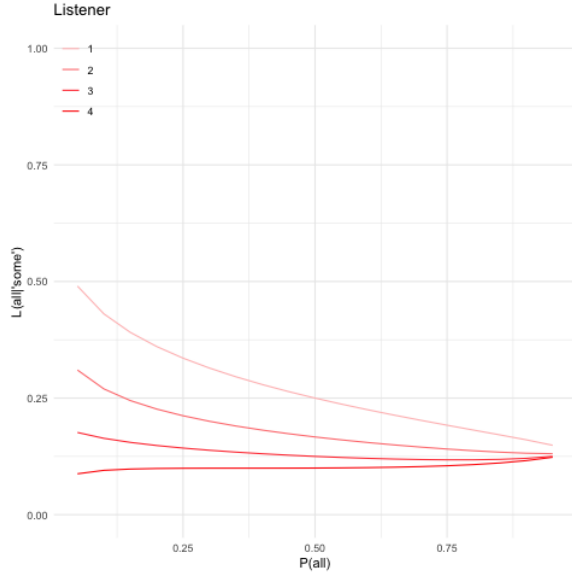
In addition, a lower or higher value of  $k$  will result in worse predictions (see the next page).



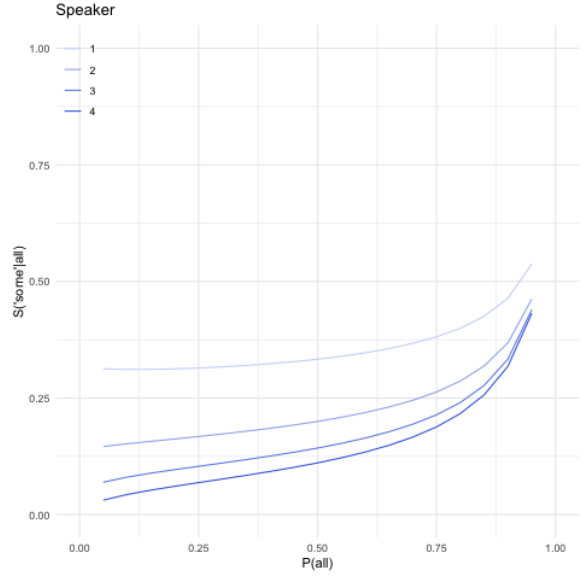
(a)  $L_n(\text{a11}|\text{'some'})$  for  $1 \leq n \leq 4$ ,  $k = 0.3$



(b)  $S_n(\text{'some'}|\text{a11})$  for  $1 \leq n \leq 4$ ,  $k = 0.3$



(c)  $L_n(\text{a11}|\text{'some'})$  for  $1 \leq n \leq 4$ ,  $k = 0.7$



(d)  $S_n(\text{'some'}|\text{a11})$  for  $1 \leq n \leq 4$ ,  $k = 0.7$