CS 784: Computational Linguistics
Lecture 17: Pragmatics

Freda Shi

School of Computer Science, University of Waterloo
fhs@uwaterloo.ca

March 18, 2025

Announcements

Assignment 2: For GPT-2 conditional entropy/probability calculation, since there is no official start-of-sentence token, we cannot directly calculate a well-defined $P(w_0)$. Instead, you may just start from the second word.

Final Exam: Apr 11, 2025, 4:00pm-6:30pm, MC 2034. It will be open-book, open-notes, and open-internet.

Pragmatics: Meanings Beyond Semantics

> *A diplomat who says yes means maybe, a diplomat who says maybe means no, and a diplomat who says no is no diplomat.*
> —*Talleyrand*

What one communicates may differ from what one literally says.

Pragmatics: the study of the underlying communication intentions and goals of speakers.

Near-Side vs. Far-Side Pragmatics

- **Near-side pragmatics** focuses on facts relevant to what is said.
- **Far-side pragmatics** focuses on what happens beyond.

Stop at the car.
Near-side: the speaker is talking about where to park the car (among multiple options).

I'm running late.
Far-side: the speaker is implying that the listener (the driver) should hurry up.

Grice's Maxims

- **Quantity**: Make your contribution as informative as is required.
  1. Make your contribution as informative as is required (for the current purposes of the exchange).
  2. Do not make your contribution more informative than is required.
- **Quality**: Try to make your contribution one that is true.
- **Relation**: Be relevant.
- **Manner**: Be perspicuous.
  - Unclear statements usually carry more information.

A and B are talking about a mutual friend, C, who is now working in a bank. A asks B how C is getting on in his job, and B replies: Oh quite well, I think; he likes his colleagues, and he hasn't been to prison yet.

Q-Based and R-Based Implicatures

Q(uantity-1)-based implicature: make your contribution as informative as is required.

Some students passed the exam.

- At least one (or two) student passed the exam.
- Not all students passed the exam (otherwise they could have said *all* students passed the exam).

R(elation; quantity-2)-based implicature: say no more than is required.

I broke a finger yesterday.

- I broke my own finger. There is no reason to say *my* finger as it can be inferred.

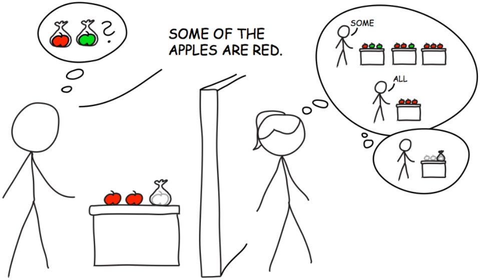
Implicatures

- The speaker had said that p (utterance).
- There is no reason to suppose that they are not observing the maxims.
- They could not be doing this unless they thought that q (communication intent).
- They know (**and know that I know that they know**) that I can see that the supposition that they think that q is required.
- They have done nothing to stop me thinking that q.
- They intended me to think, or at lease willing to allow me to think, that q; and so they have implicated that q.

Scalar Implicature

You know that there are three apples on the table, but only I can see them. You hear me say, *some of the apples are red.*

How many apples do you think are red?



[Source: Scontras et al.; https://www.problang.org/]

Indirect Scalar Implicature

Direct implicature: *some* means *some but not all*.

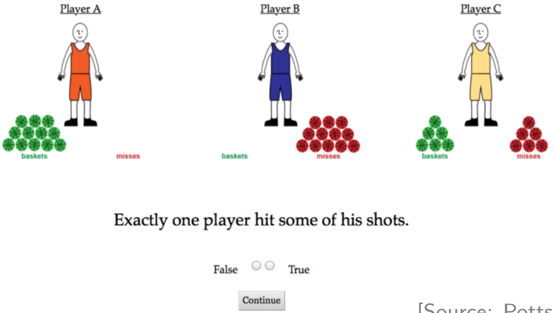
You know that there are three apples on the table, but only I can see them. You hear me say, *not all apples are red.*

How many apples do you think are red?

Indirect implicature: *not all* means *some* (i.e., at least one).

Embedded Implicature

$$\left\{\begin{array}{l}\text{Every}\\ \text{Exactly one}\\ \text{No}\end{array}\right\} \text{ player hit } \left\{\begin{array}{l}\text{all}\\ \text{none}\\ \text{some}\end{array}\right\} \text{ of his shots.}$$



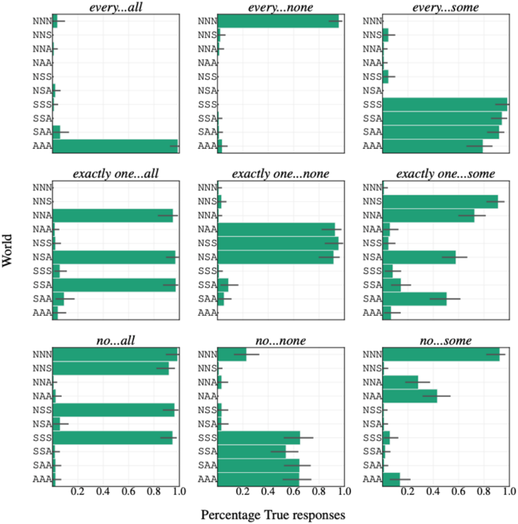
[Source: Potts et al., 2015]

Embedded Implicature

Humans show clear preferences on world states.

$$\left\{\begin{matrix}\text{Every}\\ \text{Exactly one}\\ \text{No}\end{matrix}\right\}\ \text{player}$$

$$\text{hit}\left\{\begin{matrix}\text{all}\\ \text{none}\\ \text{some}\end{matrix}\right\}\ \text{of his shots.}$$



Question Under Discussion

I've been waiting for a million years.
I paid one thousand dollars for the coffee.

Communication intents are relevant to what question under discussion (QUD) a speaker is likely addressing with their utterances.

Comparison Class

*Look at the **big** tree near the **small** lake.*

The *small lake* is larger than the *big tree* anyway.

This is because we infer the different comparison classes for them.

Overinformativeness and Politeness

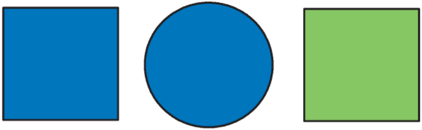
(In a coffee shop)
Q: Do you have iced tea?
A: No, but we do have ***iced coffee****.*

People often violate the informativeness rule.

In such a scenario, pragmatic inference happens when choosing the substitution:

- * *No, but we do have* ***muffins****.*
- * *No, but we do have* ***laptops****.*

Referential Communication Game

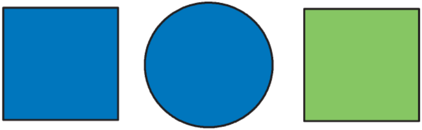


[Source: Frank & Goodman, 2012]

You and your friend are playing a referential communication game.

- The **speaker** knows the correct object and tells the listener.
- The **listener** picks up **one** object following the speaker's instruction.
- Both players will be rewarded if the listener picks up the correct object.
- The speaker can only refer to the intrinsic properties of objects (no "middle" or "left" etc.).

Referential Communication Game: Example



[Source: Frank & Goodman, 2012]

For simplicity, we assume the speakers will only use **color** and/or **shape** to refer to objects.

Suppose you are the **listener** and hear "**square**."

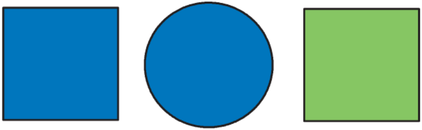
Which object will you pick? Most participants say the **first**.

Suppose you are the **speaker** and want to refer to the third object.

Which word will you use, **green** or **square**?

A: **green** as it's less ambiguous.

The Rational Speech Act (RSA) Framework



[Source: Frank & Goodman, 2012]

Computationally models the process that "square" refers to the first object.

A Bayesian approach: derive posterior (pragmatics) from the prior (semantics/literal meanings).

We need a set of pre-defined **world states** (i.e., possible choices) S and possible **utterances** U:

$$S = \{\text{blue-square}, \text{blue-circle}, \text{green-square}\}$$
$$U = \{\text{blue}, \text{green}, \text{square}, \text{circle}\}$$

RSA: Literal Listener L_0

[Source: Frank & Goodman, 2012]

$S = \{\text{blue-square, blue-circle, green-square}\}$

$U = \{\text{blue, green, square, circle}\}$

$$P_{L_0}(s \mid u) \propto \llbracket \text{u} \rrbracket (s) P(s)$$

$$\propto \begin{cases} P(s) & \text{if } u \text{ is true for } s \\ 0 & \text{otherwise} \end{cases}$$

$$\llbracket \text{blue} \rrbracket [\text{blue-square}] = 1$$

$$\llbracket \text{blue} \rrbracket [\text{green-square}] = 0$$

$$\llbracket \text{green} \rrbracket [\text{green-square}] = 1$$

RSA: Literal Listener with an Example

[Source: Frank & Goodman, 2012]

$$P_{L_0}(s \mid u) \propto [\![\mathrm{u}]\!](s)P(s)$$

$$\propto \begin{cases} P(s) & \text{if } u \text{ is true for } s \\ 0 & \text{otherwise} \end{cases}$$

Suppose $P(s)$ is uniform over all s, here is $[\![\mathrm{u}]\!](s)P(s)$:

$U \backslash S$	blue-square	blue-circle	green-square
blue	$1/3$	$1/3$	0
green	0	0	$1/3$
square	$1/3$	0	$1/3$
circle	0	$1/3$	0

RSA: Literal Listener with an Example

[Source: Frank & Goodman, 2012]

$$P_{L_0}(s \mid u) \propto [\![\mathrm{u}]\!](s)P(s)$$

$$\propto \begin{cases} P(s) & \text{if } u \text{ is true for } s \\ 0 & \text{otherwise} \end{cases}$$

P_{L_0} is a **probability distribution**, so we normalize the table:

$U \backslash S$	blue-square	blue-circle	green-square
blue	1/2	1/2	0
green	0	0	1
square	1/2	0	1/2
circle	0	1	0

RSA: Pragmatic Speaker S_1

[Source: Frank & Goodman, 2012]

$$S = \{\text{blue-square}, \text{blue-circle}, \text{green-square}\}$$
$$U = \{\text{blue}, \text{green}, \text{square}, \text{circle}\}$$

Suppose the pragmatic speaker S_1 has the **theory of mind** of literal listener L_0, and therefore knows the distribution $P_{L_0}(s \mid u)$.

S_1 will choose the utterance u that maximizes the probability of the listener picking the correct object:

$$P_{S_1}(u \mid s) \propto P_{L_0}(s \mid u)$$

In real-practice, S_1 also considers the cost of utterances.

$$U(u; s) = \log P_{S_1}(u \mid s) - \text{cost}(u) \qquad \text{(utility function)}$$
$$P_{S_1}(u \mid s) \propto \alpha \exp(U(u; s)) \qquad (\alpha : \text{ hyperparameter})$$

RSA: Pragmatic Speaker with an Example

[Source: Frank & Goodman, 2012]

$$U(u; s) = \log P_{S_1}(u \mid s) - \text{cost}(u)$$

$$P_{S_1}(u \mid s) \propto \alpha \exp(U(u; s))$$

Assume 0 cost function and $\alpha = 1$, $P_{S_1}(u \mid s) \propto P_{L_0}(s \mid u)$.

$P_{L_0}(s \mid u)$

$U \backslash S$	blue-square	blue-circle	green-square
blue	1/2	1/2	0
green	0	0	1
square	1/2	0	1/2
circle	0	1	0

RSA: Pragmatic Speaker with an Example

[Source: Frank & Goodman, 2012]

$$U(u; s) = \log P_{S_1}(u \mid s) - \mathrm{cost}(u)$$
$$P_{S_1}(u \mid s) \propto \alpha \exp(U(u; s))$$

Normalize the column to form a well-defined probability distribution: $P_{S_1}(u \mid s) \propto P_{L_0}(s \mid u)$.

$P_{S_1}(u \mid s)$

$U \backslash S$	blue-square	blue-circle	green-square
blue	$1/2$	$1/3$	0
green	0	0	$2/3$
square	$1/2$	0	$1/3$
circle	0	$2/3$	0

RSA: Pragmatic Listener L_1

[Source: Frank & Goodman, 2012]

$S = \{\text{blue-square, blue-circle, green-square}\}$
$U = \{\text{blue, green, square, circle}\}$

The pragmatic listener L_1 has the **theory of mind** of the pragmatic speaker S_1, and therefore knows the distribution $P_{S_1}(u \mid s)$.

L_1 will infer the world state s from the utterance u:

$$P_{L_1}(s \mid u) \propto P_{S_1}(u \mid s) P(s)$$

RSA: Pragmatic Listener with an Example

[Source: Frank & Goodman, 2012]

$$P_{L_1}(s \mid u) \propto P_{S_1}(u \mid s) P(s)$$

Assume uniform prior over world states. Normalize the rows to form a well-defined probability distribution: $P_{L_1}(s \mid u) \propto P_{S_1}(u \mid s)$.

$P_{S_1}(u \mid s)$

$U \backslash S$	blue-square	blue-circle	green-square
blue	$1/2$	$1/3$	0
green	0	0	$2/3$
square	$1/2$	0	$1/3$
circle	0	$2/3$	0

RSA: Pragmatic Listener with an Example

[Source: Frank & Goodman, 2012]

$$P_{L_1}(s \mid u) \propto P_{S_1}(u \mid s) P(s)$$

Assume uniform prior over world states. Normalize the rows to form a well-defined probability distribution: $P_{L_1}(s \mid u) \propto P_{S_1}(u \mid s)$.

$P_{L_1}(s \mid u)$

$U \backslash S$	blue-square	blue-circle	green-square
blue	3/5	2/5	0
green	0	0	1
square	3/5	0	2/5
circle	0	1	0

Summary: RSA

$$
\begin{aligned}
P_{L_0}(s \mid u) &\propto [\![\mathrm{u}]\!](s)P(s) \\
P_{S_1}(u \mid s) &\propto P_{L_0}(s \mid u) \\
U(u; s) &= \log P_{S_1}(u \mid s) - \mathrm{cost}(u) \\
P_{S_1}(u \mid s) &\propto \alpha \exp(U(u; s)) \\
P_{L_1}(s \mid u) &\propto P_{S_1}(u \mid s)P(s)
\end{aligned}
$$

Caveat: While being supported by human behavioral data, RSA is just one (among arguably infinite) approach to modeling human pragmatic reasoning behaviors

More at `https://www.problang.org/`

Next

Linguistic typology and computational multilingualism.