Class 10: Game Theory 3

1 Gricean Pragmatics

Grice (1957, 1975)'s research program:

- Analyse linguistic communication as a kind of rational behaviour.

  How do we calculate speaker meaning (according to Grice)?

1. We assume that speakers are rational.
2. We assume that speakers are cooperative.

(1) **The Cooperative Principle** (Grice 1975)
Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Broken down into maxims (among others):

(2) **Quality:** 'Don't lie'
  a. Try to make your contribution one that is true.
  b. Do not say what you believe to be false.
  c. Do not say that for which you lack adequate evidence.

(3) **Quantity**: 'Be informative'
  a. Make your contribution as informative as is required (for the current purposes of the exchange).
  b. Do not make your contribution more informative than is required.

(4) Mary ate some of the cookies.
Normally implies *Mary didn't eat all of the cookies.*

Listener's reasoning w.r.t. *some*:

- The speaker used the utterance with *some* instead of *all* which would be more informative (**quantity**).
- If the speaker knew that Mary ate all the cookies, they would have used it.
- The speaker is well informed and respects the maxim of quality.
- ∴ It's not the case that Mary ate all the cookies.

Very nice idea, but not very formal…

- We can give a formalization on Gricean reasoning using Game Theory.
- I'll present the *Rational Speech Act Model* (RSA)

2 Signaling games

Signaling game: A formal description of an interaction situation between two agents which is

- A situation of **coordination**.
- A game of incomplete information (all the information is not available to all the agents).

Developed for linguistic pragmatics by Lewis 1969.

(5) Definition of a signaling game

a. 2 players: N = {S, L} (S = sender, speaker and L = receiver, listener).
b. S observes a fact in the world (their type) and wants to communicate this fact to L.
c. To help them communicate their type, S chooses a **message**, a pair of a form and an interpretation, to send to S.
d. L hears the message and assigns it an interpretation taking into account the semantic meaning of the message and their prior beliefs about the world.
e. S and L win if L's interpretation corresponds to S's type. Otherwise, both lose.

More formally:

(6) A signaling game is a tuple $\langle \{S, L\}, W, M, [\![\cdot]\!], U, Pr \rangle$

- {S, L} are the players.
- W = a set of possible worlds.
- M = a set of messages
- $[\![\cdot]\!]$ is a semantic interpretation function.
- U is a utility function
- Pr is a probability distribution over W representing L's prior beliefs.

World	Description
w_0	Mary ate 0 cookies
w_1	Mary ate 1 cookie
w_2	Mary ate 2 cookies
w_3	Mary ate 3 cookies

Table 1: Possible worlds

Short name	message	$[\![\text{message}]\!]$
NONE	Mary ate none of the cookies	$\{w_0\}$
SOME	Mary ate some of the cookies	$\{w_1, w_2, w_3\}$
ALL	Mary ate all of the cookies	$\{w_3\}$

Table 2: Messages in cookie example

Suppose L has no prior expectations about how many cookies Mary ate. We can represent the listener's uncertain belief state through having their prior beliefs, Pr, be **uniform** over the set of possible worlds.

w_0	w_1	w_2	w_3
0.25	0.25	0.25	0.25

Table 3: L has uniform prior beliefs (Pr(w)).

2.1 Solution concept

Nash equilibria too simple: multiple equilibria. (ex. pooling).

- Integrate a formalization of Gricean reasoning (Quality and Quantity) into solution concept.
- Multiple steps.

When the listener hears a message m, the first thing that they do is restrict their attention to the worlds in which m is true. More technically, L conditions their beliefs on the meaning of the message, which is equivalent to intersection followed by renormalization of the measure.

- Conditionalization formalizes the maxim of Quality.

(7) $Pr(w|m) = \frac{Pr(\{w\} \cap [\![m]\!])}{Pr([\![m]\!])}$

Message	w_0	w_1	w_2	w_3
NONE	1	0	0	0
ALL	0	0	0	1
SOME	0	$0.33\dot{3}$	$0.33\dot{3}$	$0.33\dot{3}$

Table 4: L's beliefs immediately after hearing m ($\Pr(w|m)$).

Formalization of the maxim of quantity:

- Building informativity into the speaker's *utility function* (U_S). S's utility function is a measure of how useful a message would be for S to communicate their desired piece of information to L
- Informativity is formalized as in information theory (Shannon 1949): natural log of prior conditioned on the meaning of the message.

(8) Utility of a message m to communicate w ($U_S(m, w)$):
$U_S(m, w) = ln(Pr(w|m))$

In signalling games, speaker utility functions often also encode information associated with **costs** for different messages. Since, in our small example, there is no major length or other grammatical difference between the messages under consideration, we will not incorporate message costs into the speaker utility function; however, message costs are an important way in which linguistic conditioning factors can be captured in game-theoretic pragmatics.

Message	w_0	w_1	w_2	w_3
NONE	0	$-\infty$	$-\infty$	$-\infty$
ALL	$-\infty$	$-\infty$	$-\infty$	0
SOME	$-\infty$	-1.099	-1.099	-1.099

Table 5: S's utility for m for communicating w ($U_S(m, w)$).

Bayesian game-theoretic models can be used to make gradient quantitative predictions concerning linguistic production and interpretation. This arises under the hypothesis that speakers are **approximately rational**: they are trying to make the choice that will have the best chance of accomplishing their goals (whatever they may be), but may not always be perfect.

(9) For a world w, a message m and a real number α, called the *temperature*,

$$P_S(m|w) = \frac{exp(\alpha \times U_S(w, m))}{\sum_{m' \in M} exp(\alpha \times U_S(w, m'))}$$

When we are modelling actual quantitative studies, the value for α that best fits the observed data can be estimated, but for an example: $\alpha = 4$.

Listeners interpret messages using their hypotheses that speakers are (approximately) rational and motivated by informativity, combined with their prior beliefs.

- We treat linguistic interpretation as Bayesian inference, and, from the **probability that S would use m given** w) and L's **prior beliefs concerning** w), we derive a probability distribution over interpretations of messages.

(10) $P(w|m) = \frac{Pr(w) \times P(m|w)}{\sum_{j=1}^{|W|} Pr(w_j) \times P(m|w_j)}$

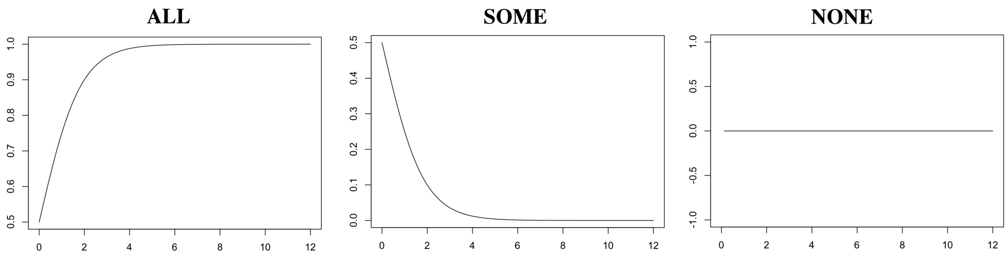

Figure 1: Predicted probabilities of using ALL vs SOME vs NONE to communicate w_3, by α

Message	w_0	w_1	w_2	w_3
NONE	1	0	0	0
ALL	0	0	0	0.99
SOME	0	1	1	0.01
Prediction	Cat. NONE	Cat. SOME	Cat. SOME	Favored ALL

Table 6: S's predicted use of m, given w with $\alpha = 4$ ($P_S(m|w)$).

Message	w_0	w_1	w_2	w_3	PREDICTION
NONE	1	0	0	0	Categorical w_0
ALL	0	0	0	1	Categorical w_3
SOME	0	0.498	0.498	0.005	Favoured w_1, w_2

Table 7: L's predicted interpretation of w, given m ($P_L(w|m)$).

Listener prior beliefs affect interpretation.

- Suppose that L knows that Mary usually likes to have two cookies for her dessert. So, before hearing what S has to say, they are expecting Mary to have eaten two cookies.

w_0	w_1	w_2	w_3
0.1	0.1	0.7	0.1

Table 8: L's priors heavily weighted on w_2.

Then, L's probability of interpreting w_2 after hearing SOME jumps up to **0.87**.