Probabilistic language modeling:

An introduction to the Rational Speech Act framework

Dasha Popova

07.03.2024

Grice, 1975 “Logic and Conversation”

“what is said” vs. “what is implicated”

A: How is C getting on in his job?

B: Oh quite well, I think; he likes his colleagues, and he hasn’t been to prison yet

CONVERSATIONAL IMPLICATURES

Grice, 1975 “Logic and Conversation”

Cooperative Principle

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.

Can you think of the situations where the principle is violated?

Grice, 1975 “Logic and Conversation”

Maxims of Quality

Try to make your contribution one that is true.

a. Do not say what you believe to be false.

b. Do not say that for which you lack adequate evidence.

Grice, 1975 “Logic and Conversation”

Maxims of Quantity

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.

A: *Where are you from?*

B: *From Germany / From Berlin / From Potsdam.*

Grice, 1975 “Logic and Conversation”

Maxim of Relation (Relevance)

Be relevant

A: *Where’s Bill?*

B: *There’s a yellow VW outside Sally’s house*. (Levinson, 1983)

Grice, 1975 “Logic and Conversation”

Maxims of Manner

Be perspicuous:

a. Avoid obscurity of expression.

b. Avoid ambiguity.

c. Be brief (avoid unnecessary prolixity).

d. Be orderly.

Grice, 1975 “Logic and Conversation”

These are general principles of rational cooperative behavior not specific to the language use

Can you think of some examples outside of the linguistic domain?

Grice, 1975 “Logic and Conversation”

Alan: *Are you going to Paul’s party?*

Barb: *I have to work*.

Barb *meant* that she is not going to Paul’s party by *saying* that she has to work. She did not say that she is not going to Paul’s party, and the sentence she uttered does not mean that. Grice introduced the technical terms ***implicate*** and ***implicature*** for the case in which what the speaker said is distinct from what the speaker thereby meant or implied. Thus Barb implicated that she is not going; that she is not going was her implicature.

Horn, 1972 Scalar implicatures

A scalar implicature is a quantity implicature based on the use of an informationally weak term in an implicational scale.

Scale: SOME (less informative: some, all) -- ALL (more informative: all)

The pianist played some Mozart sonatas

More informative alternative utterance: *The pianist played all Mozart sonatas*

Scalar implicature: *The pianist did not play all Mozart sonatas*

The choice by the speaker of a less informative utterance implies the negation of the more informative alternative (some = not all).

Scalar implicatures

An implicational scale is a set of lexical items that are

- of the same constituent category, and
- ordered in terms of their informativeness.

Any sentence including one item of the set entails all of the propositions expressed by similar sentences containing less informative items. In addition, in keeping with the quantity maxim, the use of an item in the scale creates scalar implicatures that disallow the propositions expressed by similar sentences that contain more informative items.

Examples: (English)

- {all, most, many, some}
- {always, often, sometimes}
- {succeed in, try to, want to}
- {certain, probable, possible}

Can you think of more examples for English? Can you think of examples for other languages? Do you predict the scales to be universal or language-specific? What is your hunch?

The Rational Speech Act framework

- **pragmatics**: speakers and listeners reason about each other's reasoning about the literal interpretation of utterances
- **modeling**
- **probabilistic approach**
- **general theory of communication** that can be used for modeling **complex phenomena**, like metaphor (Kao et al., 2014), hyperbole (Kao et al., 2014), degree semantics (Lassiter and Goodman, 2013)

The Rational Speech Act framework

- communication as **recursive reasoning** between a speaker and a listener

- the listener interprets the speaker’s utterance by reasoning about a cooperative speaker trying to inform a naive listener about some state of affairs

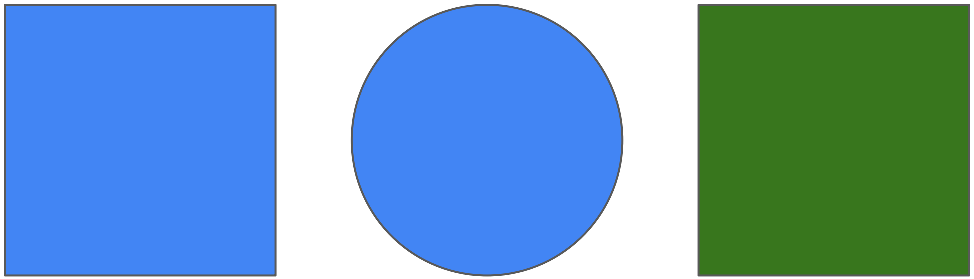
- using Bayesian inference, the listener reasons about what the state of the world is likely to be given that a speaker produced some utterance, knowing that the speaker is reasoning about how a listener is most likely to interpret that utterance

- thus, we have (at least) three levels of inference: at the top, the sophisticated, **pragmatic listener** *L1* reasons about the **pragmatic speaker** *S1* and infers the state of the world *s* given that the speaker chose to produce the utterance *u*; the speaker chooses *u* by maximizing the probability that a naive, **literal listener**, *L0* would correctly infer the state of the world *s* given the literal meaning of *u*

A vanilla RSA model

[Frank and Goodman (2012)]

In a **reference game**, a speaker wants to refer to one of the given objects



A vanilla RSA model

Set of world states:

S = {blue-square, blue-circle, green-square}

Set of utterances:

U = {“square", “circle", “green", “blue"}

$$\begin{array}{lll} L_1 & \textit{pragmatic listener} & P_{L_1}(s|u) \propto P_{S_1}(u|s) \cdot P(s) \\ \downarrow & & \\ S_1 & \textit{pragmatic speaker} & P_{S_1}(u|s) \propto \exp(\alpha U_{S_1}(u;s)) \\ \downarrow & & \\ L_0 & \textit{literal listener} & P_{L_0}(s|u) \propto [\![\mathrm{u}]\!](s) \cdot P(s) \end{array}$$

Literal Listener *L0*

The naive, literal listener *L0* interprets an utterance according to its meaning: she computes the probability of s (state) given u (utterance) according to the semantics of *u* and the prior probability of *s*

For example, the utterance “blue” is true of states “blue-square”, “blue-circle” and false of state “green-square”, $[[u]]:S \mapsto \{0,1\}$

$P_{L0}(s \mid u) \propto [[u]](s) \cdot P(s)$, where P(s) is an a priori belief regarding which state or object the speaker is likely to refer to in general

https://www.problang.org/chapters/01-introduction.html

Pragmatic Speaker *S1*

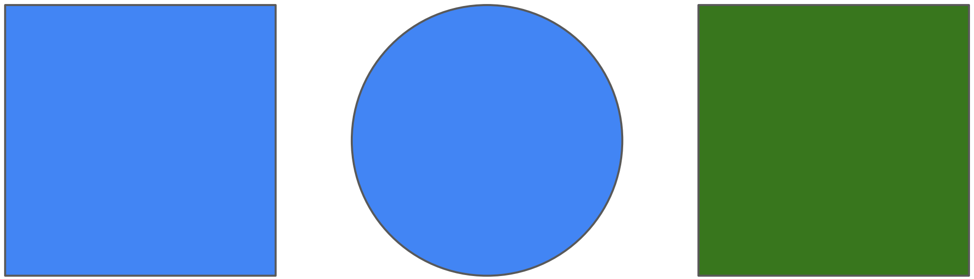
$P_{S1}(u|s) \propto \exp(\alpha(\log L_0(s|u) - C(u)))$, α -- rationality, optimality of the choice of utterance, C(u) -- cost of utterance

https://www.problang.org/chapters/01-introduction.html

Pragmatic Listener *L1*

$P_{L1}(s \mid u) \propto P_{S1}(u \mid s) \cdot P(s)$

https://www.problang.org/chapters/01-introduction.html



Parallels with Grice

Grice	**RSA**
Quality	
Quantity	
Manner	
Relevance	

Parallels with Grice

recursive nature of RSA mimics the calculable nature of the conversational implicature (the listener thinks that the speaker thinks that the listener thinks…)

Grice	**RSA**
Quality	zero probabilities for false utterances
Quantity	the speaker favours informative utterances
Manner	the function of the cost of the utterance C(u)
Relevance	conditional probabilities: the choice of the utterance depends on the choice of the state, the choice of the state depends on the choice of the utterance

Simple scalar implicature

P(r1) = P(r2) = 0.5

C(m) = 0

α = 1

r_1

r_2

P_{Lit}	r_1	r_2
'hat'	0	1
'glasses'	0.5	0.5

	'hat'	'glasses'
r_1	0	1
r_2	0.67	0.33

	r_1	r_2
'hat'	0	1
'glasses'	0.75	0.25

$$P_{\text{Lit}}(r \mid m) = \frac{[\![m]\!](r)}{\sum_{r' \in R} [\![m]\!](r')}$$

$$P_S(m \mid r) = \frac{P_{\text{Lit}}(r \mid m)}{\sum_{m' \in M} P_{\text{Lit}}(r \mid m')}$$

$$P_L(r \mid m) = \frac{P_S(m \mid r)}{\sum_{r' \in R} P_S(m \mid r')}$$

Simple scalar implicature

1) Start with the Lexicon:

	r_1	r_2
'hat'	0	1
'glasses'	1	1

2) Normalize by rows:

P_{Lit}	r_1	r_2
'hat'	0	1
'glasses'	0.5	0.5

3) Transpose:

	'hat'	'glasses'
r_1	0	0.5
r_2	1	0.5

4) Normalize by rows:

P_S	'hat'	'glasses'
r_1	0	1
r_2	0.67	0.33

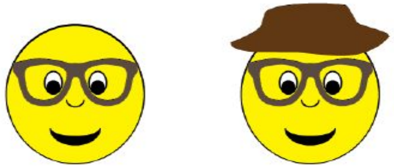
5) Transpose:

	r_1	r_2
'hat'	0	0.67
'glasses'	1	0.33

6) Normalize by rows:

P_L	r_1	r_2
'hat'	0	1
'glasses'	0.75	0.25

The role of the cost of the utterance



P(r1) = P(r2) = 0.5

C('hat') = 6

C('glasses') = 0

α = 1

P_{Lit}	r_1	r_2
'hat'	0	1
'glasses'	0.5	0.5

$$P_{Lit}(r \mid m) = \frac{[\![m]\!](r) \cdot P(r)}{\sum_{r' \in R} [\![m]\!](r') \cdot P(r')}$$

P_S	'hat'	'glasses'
r_1	0	1
r_2	0.0049	0.9951

$$P_S(m \mid r) = \frac{\exp(\alpha \cdot (\log P_{Lit}(r \mid m) + C(m)))}{\sum_{m' \in M} \exp(\alpha \cdot (\log P_{Lit}(r \mid m') + C(m')))}$$

P_L	r_1	r_2
'hat'	0	1
'glasses'	0.5012	0.4988

$$P_L(r \mid m) = \frac{P_S(m \mid r) \cdot P(r)}{\sum_{r' \in R} P_S(m \mid r') \cdot P(r')}$$

The role of the cost of the utterance

1) Lexicon

	r_1	r_2
'hat'	0	1
'glasses'	1	1

2) Normalizing by row:

P_{Lit}	r_1	r_2
'hat'	0	1
'glasses'	0.5	0.5

3) Transposing:

	'hat'	'glasses'
r_1	0	0.5
r_2	1	0.5

4) Speaker (cost):

	'hat'	'glasses'
r_1	$\exp(\log(0)-6)$	$\exp(\log(0.5)-0)$
r_2	$\exp(\log(1)-6)$	$\exp(\log(0.5)-0)$

$\Rightarrow$

	'hat'	'glasses'
r_1	0	0.5
r_2	0.0025	0.5

5) Normalizing by row:

P_S	'hat'	'glasses'
r_1	0	1
r_2	0.0049	0.9951

6) Transposing:

	r_1	r_2
'hat'	0	0.0049
'glasses'	1	0.9951

7) Normalizing by row:

P_L	r_1	r_2
'hat'	0	1
'glasses'	0.5012	0.4988

The role of alpha (rationality, optimality)

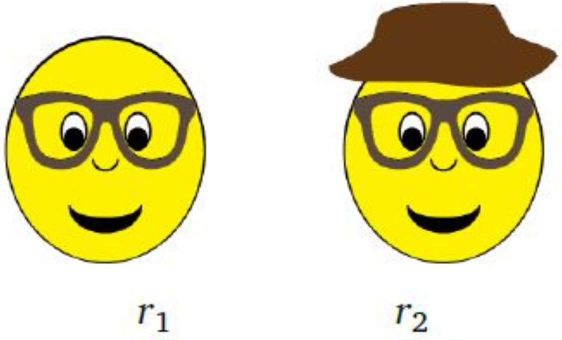
$$P_S(m \mid r) = \frac{\exp(\alpha \cdot (\log P_{\text{Lit}}(r \mid m)))}{\sum_{m' \in M} \exp(\alpha \cdot (\log P_{\text{Lit}}(r \mid m')))}$$

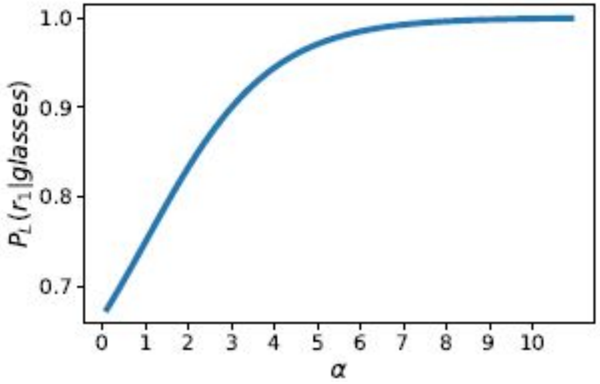
P_S	'hat'	'glasses'
r_1	0	1
r_2	0.67	0.33

$\alpha = 1$

P_S	'hat'	'glasses'
r_1	0	1
r_2	0.94	0.06

$\alpha = 4$

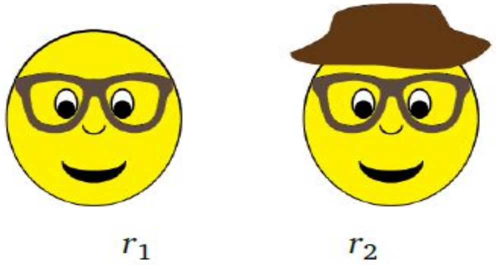




The role of the prior probability P(r)

P_{Lit}	r_1	r_2
'hat'	0	1
'glasses'	0.3	0.7

$$P_{\text{Lit}}(r \mid m) = \frac{[\![m]\!](r) \cdot P(r)}{\sum_{r' \in R} [\![m]\!](r') \cdot P(r')}$$

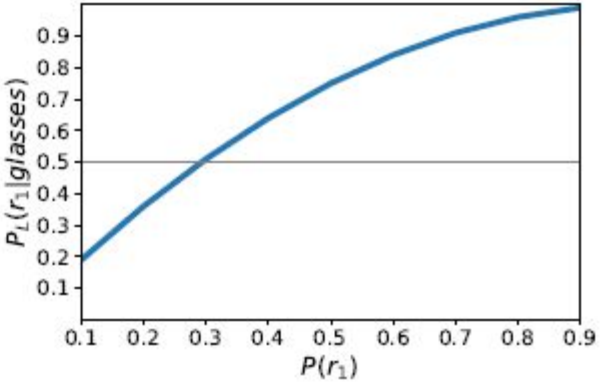


P_S	'hat'	'glasses'
r_1	0	1
r_2	0.59	0.41

$$P_S(m \mid r) = \frac{P_{\text{Lit}}(r \mid m)}{\sum_{m' \in M} P_{\text{Lit}}(r \mid m')}$$

P_L	r_1	r_2
'hat'	0	1
'glasses'	0.51	0.49

$$P_L(r \mid m) = \frac{P_S(m \mid r) \cdot P(r)}{\sum_{r' \in R} P_S(m \mid r') \cdot P(r')}$$



The role of the prior probability P(r)

1) Lexicon:

	r_1	r_2
'hat'	0	1
'glasses'	1	1

2) Introducing prior probabilities:

	r_1	r_2
'hat'	$0 \cdot 0.3$	$1 \cdot 0.7$
'glasses'	$1 \cdot 0.3$	$1 \cdot 0.7$

$\Rightarrow$

	r_1	r_2
'hat'	0	0.7
'glasses'	0.3	0.7

3) Literal listener:

P_{Lit}	r_1	r_2
'hat'	0	1
'glasses'	0.3	0.7

4) Transposing:

	'hat'	'glasses'
r_1	0	0.3
r_2	1	0.7

5) Normalizing by row:

	'hat'	'glasses'
r_1	0	1
r_2	0.59	0.41

6) Transposing:

P_S	r_1	r_2
'hat'	0	0.59
'glasses'	1	0.41

7) Introducing prior probabilities:

	r_1	r_2
'hat'	$0 \cdot 0.3$	$0.59 \cdot 0.7$
'glasses'	$1 \cdot 0.3$	$0.41 \cdot 0.7$

$\Rightarrow$

	r_1	r_2
'hat'	0	0.413
'glasses'	0.3	0.287

8) Normalizing by row:

P_L	r_1	r_2
'hat'	0	1
'glasses'	0.51	0.49

Mini-experiment

r1 r2 r3

“hat”

Mini-experiment

literal listener, pragmatic listener, we: r3

r1 r2 r3

“hat”

Mini-experiment

r1 r2 r3

“glasses”

Mini-experiment

literal listener, pragmatic listener, we: r2

r1 r2 r3

“glasses”

Mini-experiment

r1 r2 r3

“hat”

Mini-experiment

literal listener: r1 -- 0.5, r2 -- 0.5

pragmatic listener: r1 -- 0.75, r2 -- 0.25

we:

r1 r2 r3

“hat”

Mini-experiment

r1 r2 r3

“mustache”

Mini-experiment

literal listener: r1 -- 0.5, r3 -- 0.5

pragmatic listener: r1 -- 0.33(3), r3 -- 0.66(6)

we:

r1 r2 r3

“mustache”

Mini-experiment

r1 r2 r3

“mustache”

Mini-experiment

literal listener: r2 -- 0.5, r3 -- 0.5

pragmatic listener expects r2

we:

r1 r2 r3

“mustache”

Mini-experiment

r1 r2 r3

“glasses”

Mini-experiment

literal listener: r1 -- 0.5, r2 -- 0.5

pragmatic listener expects r1

we:

r1 r2 r3

“glasses”

Mini-experiment

Mini-experiment

literal listener/pragmatic listener/we: r1 -- 0.5, r3 -- 0.5

r1 r2 r3

“glasses”

Mini-experiment

we:

r1 r2 r3

“mustache”