Semantics and pragmatics of numerical approximation expressions

Adèle Mortier

Supervisors:
Paul Égré | Benjamin Spector

February 1, 2019

A first example

"I got it for around 10 €"

A first example

"I got it for around 10 €"

What you think of this

- She does not know for sure how much it cost

A first example

"I got it for around 10 €"

What you think of this

- She does not know for sure how much it cost
- It does not matter so much

A first example

"I got it for around 10 €"

What you think of this

- She does not know for sure how much it cost
- It does not matter so much
- **She is too "lazy" to say the exact amount and is counting on me to understand correctly**

A first example

“I got it for around 10 €”

What you think of this

- She does not know for sure how much it cost
- It does not matter so much
- **She is too “lazy” to say the exact amount and is counting on me to understand correctly**

What you understand

Between 8 and 12 €?

▷ We would like to derive this in a systematic way !

Table of contents

Setting the problem

Parameters et hypotheses

Around (10) (people) (came to the party)

$$
\begin{aligned}
[\![\text{around } n \text{ Ps did Q}]\!]^{\epsilon} &= 1 \text{ iff } |P \cap Q| \in [n - \epsilon(n); n + \epsilon(n)] \\
&\triangleq 1 \text{ iff } |P \cap Q| \in \mathcal{A}_{\epsilon}^{n}
\end{aligned}
$$

Basic semantics

- Denotes an interval centered on n, size $2\epsilon(\mathrm{n})+1$
- ϵ depends on n !!

Parameters et hypotheses

Around (10) (people) (came to the party)

$$\begin{aligned} [\![\text{around } n \text{ Ps did Q}]\!]^{\epsilon} &= 1 \text{ iff } |P \cap Q| \in [n - \epsilon(n); n + \epsilon(n)] \\ &\triangleq 1 \text{ iff } |P \cap Q| \in \mathcal{A}^n_\epsilon \end{aligned}$$

Basic semantics

- Denotes an interval centered on n, size $2\epsilon(n)+1$
- ϵ depends on n !!

Dependencies

- **Order of magnitude**: the bigger the n, the wider the interval

Parameters et hypotheses

Around (10) (people) (came to the party)

$$\begin{aligned} [\![\text{around } n \text{ Ps did Q}]\!]^{\epsilon} &= 1 \text{ iff } |P \cap Q| \in [n - \epsilon(n); n + \epsilon(n)] \\ &\triangleq 1 \text{ iff } |P \cap Q| \in \mathcal{A}_{\epsilon}^{n} \end{aligned}$$

Basic semantics

- Denotes an interval centered on n, size $2\epsilon(n)+1$
- ϵ depends on n !!

Dependencies

- **Order of magnitude**: the bigger the n, the wider the interval
- **Granularity**: the coarser the n, the wider the interval

Parameters et hypotheses

Around (10) (people) (came to the party)

$$
\begin{aligned}
[\![\text{around } n \text{ Ps did Q}]\!]^{\epsilon} &= 1 \text{ iff } |P \cap Q| \in [n - \epsilon(n); n + \epsilon(n)] \\
&\triangleq 1 \text{ iff } |P \cap Q| \in \mathcal{A}_{\epsilon}^{n}
\end{aligned}
$$

Basic semantics

- Denotes an interval centered on n, size $2\epsilon(\mathrm{n}){+}1$
- ϵ depends on n !!

Dependencies

- **Order of magnitude**: the bigger the n, the wider the interval
- **Granularity**: the coarser the n, the wider the interval
- **Salience**: in presence of a salient alternative, narrower interval (Cummins, Sauerland, and Solt 2012)

Toy dataset

(Magnitude) I got it for 600,000 € *vs* 10€

Toy dataset

(Magnitude)	I got it for	600,000 €	*vs*	10€
(Granularity)	I got it for	600,000 €	*vs*	600,050 €

Toy dataset

(Magnitude)	I got it for	600,000 €	*vs*	10€
(Granularity)	I got it for	600,000 €	*vs*	600,050 €
(Salience)	[Ideal budget: 560,000 €]			got it for 600,000 €

Toy dataset

(Magnitude)	I got it for	600,000 €	*vs*	10€
(Granularity)	I got it for	600,000 €	*vs*	600,050 €
(Salience)	[Ideal budget: 560,000 €]			got it for 600,000 €

Consequences

- Granularity and magnitude are intertwined (which one is predominant??)
- Granularity and salience are implicature-based

...	1	5	1	**2**	1	2	3	1	...	simplicity
...	219	220	221	**222**	223	224	225	226	...	divisor

Figure: Interval inferred for "around 222" using granularity implicatures

Two Bayesian models

Probabilistic intervals (Égré and Verheyen 2018)

Principle: 2 levels

- When the speaker utters "around n", he thinks of a certain interval among a **set of possible intervals** (*e.g.* intervals centered around n);

Probabilistic intervals (Égré and Verheyen 2018)

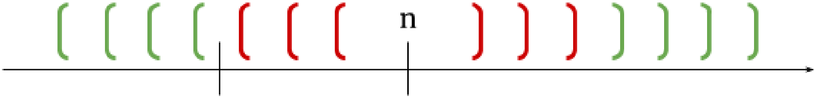
Principle: 2 levels

- When the speaker utters "around n", he thinks of a certain interval among a **set of possible intervals** (*e.g.* intervals centered around n);
- According to the listener, **each possible interval has a certain probability** (*e.g.* relatively narrow intervals might be more probable);

Probabilistic intervals (Égré and Verheyen 2018)

Principle: 2 levels

- When the speaker utters "around n", he thinks of a certain interval among a **set of possible intervals** (*e.g.* intervals centered around n);
- According to the listener, **each possible interval has a certain probability** (*e.g.* relatively narrow intervals might be more probable);
- And within a fixed interval, the **"real" number** is selected with a certain probability (*e.g.* central numbers might be more probable)



Properties of the model

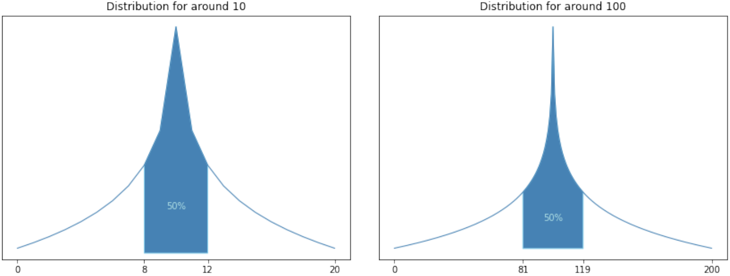


Figure: Curves generated using uniform distributions on intervals and numbers

Properties

- Symmetrical

Properties of the model

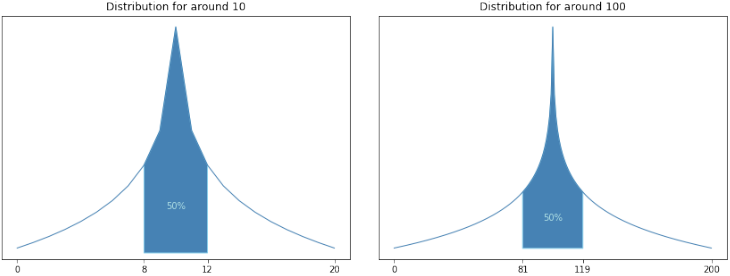


Figure: Curves generated using uniform distributions on intervals and numbers

Properties

- Symmetrical
- Scales with magnitude

Properties of the model

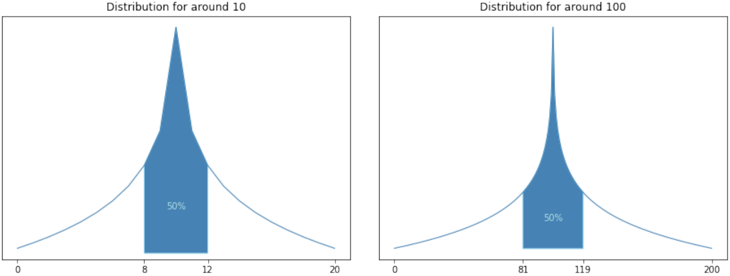


Figure: Curves generated using uniform distributions on intervals and numbers

Properties

- Symmetrical
- Scales with magnitude
- Does not account for granularity

Refinements

Behind Bayes

$$\mathbb{P}[k] = \sum_{i=|n-k|}^{n} \mathbb{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n] = -\alpha \ln(-2k+\beta) + \gamma^a$$

Hints at the Weber-Fechner law and numerical cognition?? (Dehaene 2003)

[a] $\alpha = \frac{1}{2(n+1)}$ $\beta = 2n+1$ $\gamma = \frac{\ln(2n+1)}{2(n+1)} = \alpha \ln(\beta)$

Refinements

Behind Bayes

$$\mathbb{P}[k] = \sum_{i=|n-k|}^{n} \mathbb{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n] = -\alpha \ln(-2k+\beta) + \gamma^a$$

Hints at the Weber-Fechner law and numerical cognition?? (Dehaene 2003)

[a] $\alpha = \frac{1}{2(n+1)}$ $\beta = 2n+1$ $\gamma = \frac{\ln(2n+1)}{2(n+1)} = \alpha \ln(\beta)$

Related models

- change distributions

Refinements

Behind Bayes

$$\mathbb{P}[k] = \sum_{i=|n-k|}^{n} \mathbb{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n] = -\alpha \ln(-2k+\beta) + \gamma^a$$

Hints at the Weber-Fechner law and numerical cognition?? (Dehaene 2003)

[a] $\alpha = \frac{1}{2(n+1)}$ $\beta = 2n+1$ $\gamma = \frac{\ln(2n+1)}{2(n+1)} = \alpha \ln(\beta)$

Related models

- change distributions
- relax the constraints on possible intervals

Refinements

Behind Bayes

$$\mathbb{P}[k] = \sum_{i=|n-k|}^{n} \mathbb{P}[k|\mathcal{A}_i^n]\mathbb{P}[\mathcal{A}_i^n] = -\alpha \ln(-2k+\beta) + \gamma^{a}$$

Hints at the Weber-Fechner law and numerical cognition?? (Dehaene 2003)

[a] $\alpha = \frac{1}{2(n+1)}$ $\beta = 2n+1$ $\gamma = \frac{\ln(2n+1)}{2(n+1)} = \alpha \ln(\beta)$

Related models

- change distributions
- relax the constraints on possible intervals
- higher-order uncertainty on the set of intervals

The Rational Speech Acts (RSA) model

Principle (Lassiter and N. D. Goodman 2013; Bergen, Levy, and N. Goodman 2016)

The Rational Speech Acts (RSA) model

Principle (Lassiter and N. D. Goodman 2013; Bergen, Levy, and N. Goodman 2016)

- one set of **utterances** (each with a cost and possible meanings), one set of **observations** (numbers), one set of **thresholds** (numbers);

The Rational Speech Acts (RSA) model

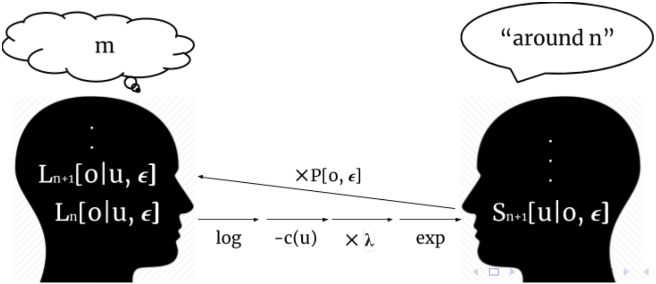
Principle (Lassiter and N. D. Goodman 2013; Bergen, Levy, and N. Goodman 2016)

- one set of **utterances** (each with a cost and possible meanings), one set of **observations** (numbers), one set of **thresholds** (numbers);
- mutually recursive Bayesian updates

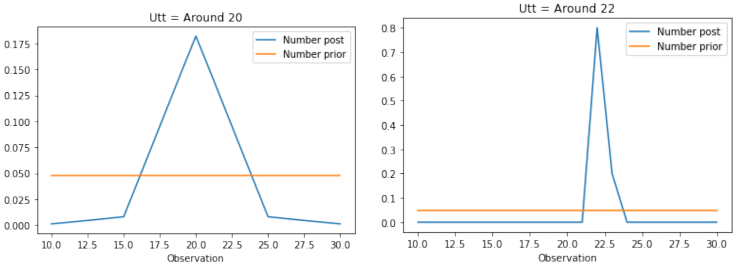
The Rational Speech Acts (RSA) model

Principle (Lassiter and N. D. Goodman 2013; Bergen, Levy, and N. Goodman 2016)

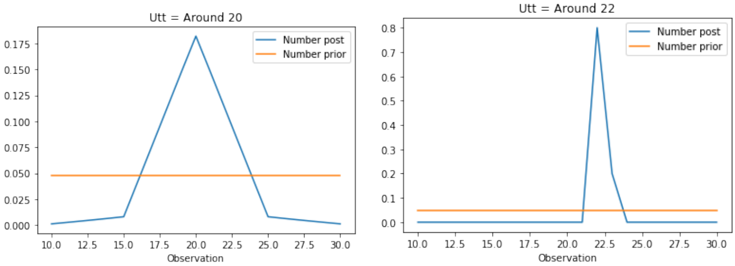
- one set of **utterances** (each with a cost and possible meanings), one set of **observations** (numbers), one set of **thresholds** (numbers);
- mutually recursive Bayesian updates
- optimality = tradeoff between **cost** and **informativity**



Results


Utt = Around 20
Number post
Number prior
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
10.0 12.5 15.0 17.5 20.0 22.5 25.0 27.5 30.0
Observation
Utt = Around 22
Number post
Number prior
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
10.0 12.5 15.0 17.5 20.0 22.5 25.0 27.5 30.0
Observation

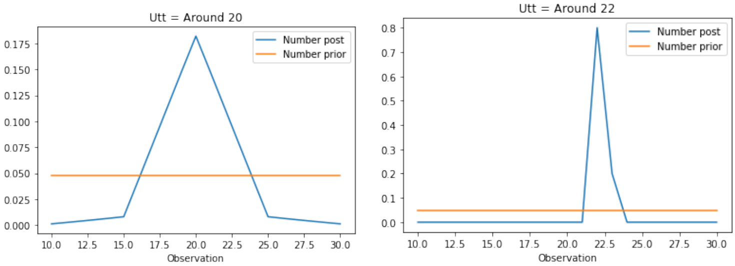
Results



Properties

- Accounts for GR (hardcoded in the cost!)

Results



Properties

- Accounts for GR (hardcoded in the cost!)
- (Almost) symmetrical

Main goals

Under quantification

#The child ate around 17 candies

All children ate around 17 candies
All children ate around 20 candies

All children ate between 16 and 18 candies
All children ate between 16 and 24 candies

Strange behavior

- with around, no "granularity violation": a number with minimum granularity can be approximated !

Under quantification

#The child ate around 17 candies

All children ate around 17 candies
All children ate around 20 candies

All children ate between 16 and 18 candies
All children ate between 16 and 24 candies

Strange behavior

- with around, no “granularity violation”: a number with minimum granularity can be approximated !
- with around, **opinionatedness** of the speaker about each number, with between, **ignorance inference**;

Under quantification

#The child ate around 17 candies

All children ate around 17 candies
All children ate around 20 candies

All children ate between 16 and 18 candies
All children ate between 16 and 24 candies

Strange behavior

- with around, no "granularity violation": a number with minimum granularity can be approximated !
- with around, **opinionatedness** of the speaker about each number, with between, **ignorance inference**;
- not a mere disjunction of numbers: it seems possible that no child ate exactly 17 candies !

Under quantification

#The child ate around 17 candies

All children ate around 17 candies
All children ate around 20 candies

All children ate between 16 and 18 candies
All children ate between 16 and 24 candies

Strange behavior

- with around, no “granularity violation”: a number with minimum granularity can be approximated !
- with around, **opinionatedness** of the speaker about each number, with between, **ignorance inference**;
- not a mere disjunction of numbers: it seems possible that no child ate exactly 17 candies !
- fine-grained, introspective data (what do people think of it?)

Conclusion

Conclusion

- probabilistic models work for **basic inferences** on numbers;

Conclusion

Conclusion

- probabilistic models work for **basic inferences** on numbers;
- but we need some additional refinements to account for **epistemic inferences**...

Conclusion

Conclusion

- probabilistic models work for **basic inferences** on numbers;
- but we need some additional refinements to account for **epistemic inferences**...
- ... and for the difference between “around” and “between”!

Conclusion

Conclusion

- probabilistic models work for **basic inferences** on numbers;
- but we need some additional refinements to account for **epistemic inferences**...
- ... and for the difference between "around" and "between"!
- a small (controlled, pre-registered) experiment might be envisaged to obtain clearer judgements about these contrasts.

Selected references I

- Leon Bergen, Roger Levy, and Noah Goodman. "Pragmatic reasoning through semantic inference". In: *Semantics and Pragmatics* 9 (May 2016). DOI: 10.3765/sp.9.20. URL: https://doi.org/10.3765/sp.9.20.

- Chris Cummins, Uli Sauerland, and Stephanie Solt. "Granularity and scalar implicature in numerical expressions". In: *Linguistics and Philosophy* 35.2 (Apr. 2012), pp. 135–169. DOI: 10.1007/s10988-012-9114-0. URL: https://doi.org/10.1007/s10988-012-9114-0.

- Stanislas Dehaene. "The neural basis of the Weber–Fechner law: a logarithmic mental number line". In: *Trends in Cognitive Sciences* 7.4 (Apr. 2003), pp. 145–147. DOI: 10.1016/s1364-6613(03)00055-x. URL: https://doi.org/10.1016/s1364-6613(03)00055-x.

- Paul Égré and Steven Verheyen. "Vagueness, approximation and the Maxim of Quality". In: Presentation. University of Groningen, Department of Theoretical Philosophy, avril 2018.

- Danny Fox. "Free Choice and the Theory of Scalar Implicatures". In: Jan. 2007, pp. 71–120. ISBN: 978-1-349-28206-7. DOI: 10.1057/9780230210752_4.

Selected references II

H. P. Grice. "Logic and Conversation". In: *Syntax and Semantics: Vol. 3: Speech Acts*. Ed. by Peter Cole and Jerry L. Morgan. New York: Academic Press, 1975, pp. 41–58. URL: http://www.ucl.ac.uk/ls/studypacks/Grice-Logic.pdf.

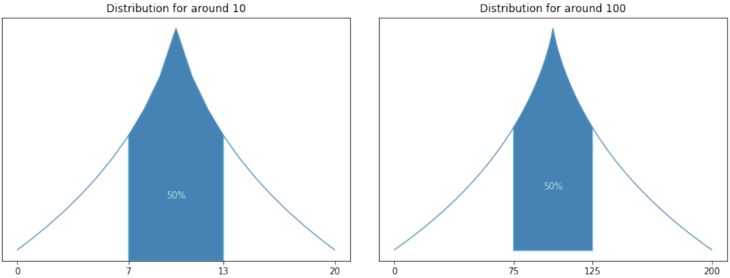
Daniel Lassiter and Noah D. Goodman. "Context, scale structure, and statistics in the interpretation of positive-form adjectives". In: *Semantics and Linguistic Theory* 23 (Aug. 2013), p. 587. DOI: 10.3765/salt.v23i0.2658. URL: https://doi.org/10.3765/salt.v23i0.2658.

Uli Sauerland. "Scalar Implicatures in Complex Sentences". In: *Linguistics and Philosophy* 27.3 (June 2004), pp. 367–391. DOI: 10.1023/b:ling.0000023378.71748.db. URL: https://doi.org/10.1023/b:ling.0000023378.71748.db.

Stephanie Solt. "Approximators as a case study of attenuating polarity items". In: *48th Annual Meeting of the North East Linguistic Society (NELS 48)* (2017).

Benjamin Spector. "Modified Numerals". In: (). to appear in the Wiley Companion to Semantics (eds. Matthewson, Meier, Rullmann, Zimmermann).

Probabilistic unconstrained intervals



Properties

- intervals containing n, but not necessarily centered around n;
- scales with magnitude;
- less peaked.

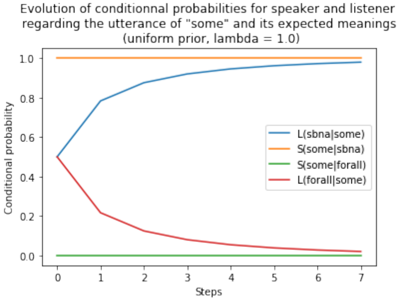
RSA formulae

$$
\begin{array}{rrcl}
 & L_0[o|u,\epsilon] & \propto & \mathbb{1}_{\{o\in[[\![u]\!]-\epsilon;[\![u]\!]+\epsilon]\}}.\mathbb{P}[o] \\
\forall n\in\mathbb{N}^+, & S_n[u|o,\epsilon] & \propto & \exp\left(\lambda(\log(L_{n-1}[o|u,\epsilon])-c(u))\right) \\
\forall n\in\mathbb{N}^+, & L_n[o|u,\epsilon] & \propto & S_n[u|o,\epsilon].\mathbb{P}[o,\epsilon]
\end{array}
$$

Explanations

- first step: for fixed u (e.g. "around x") and ϵ, just keep the observations that are in [n-ϵ; n+ϵ]; all the other have 0 probability.
- speaker step: the softmax allows to pick some non optimal possibilities with a non-zero (but very small) probability

RSA with quantifiers (replication)



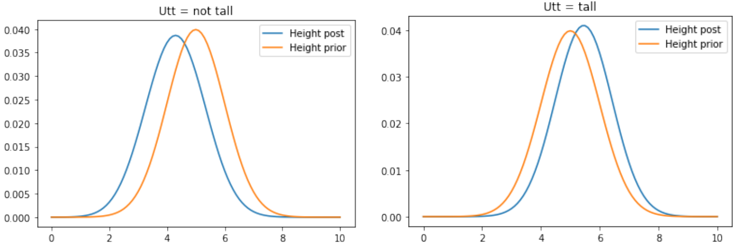
Some *vs* all

- at the beginning, “some” can mean all ($\forall$) are some but not all ($\exists_{\neg\forall}$), and “all” definitely means $\forall$.
- this asymmetry causes the meaning of “some” to converge to $\exists_{\neg\forall}$ after a few iterations.

Caveats

- Sensitive to parameter λ!
- λ is the temperature, higher λ means faster convergence but possibly to a “wrong” optimum.

RSA with gradable adjectives (replication)



Properties

- A negative utterance (“not tall”), shifts the height prior to the **left**: the listener expects the person to be **smaller**;
- A positive utterance (“tall”), shifts the height prior to the **right**: the listener expects the person to be **taller**;
- “Not tall” has a bigger effect on the prior, because it is more costly. If it has been uttered, then the person is *really* small

Quantification

$$\text{All children ate around n candies} \sim \begin{cases} \text{Child \# 1 ate around } n_1 \text{ candies} \\ \text{Child \# 2 ate around } n_2 \text{ candies} \\ \vdots \\ \text{Child \# M ate around } n_M \text{ candies} \end{cases}$$

Maybe I am saying “around n” because:

$$n \in \cap_{i=1}^{M} [[around \;\; n_i]]$$

Modality

You **can** eat around 20 candies	$=$	$\exists$ W, [16; 24]
	$\rightsquigarrow$	$\forall$ W, [0; 24]
You **must** eat around 20 candies	$=$	$\forall$ W, [16; 24]
	$\rightsquigarrow$	$\exists$ W, [24; ∞[

Strange behavior

- under existential modality, approximators rather convey an **upper bound**
- under universal modality, approximators rather convey a **lower bound**
- radically different if we use "between 16 and 24" instead !!

Additional data (Solt 2017)

Lisa has about 50 sheep.
Lisa doesn't have about 50 sheep.
Lisa has more than about 50 sheep.
Lisa doesn't have more than about 50 sheep.