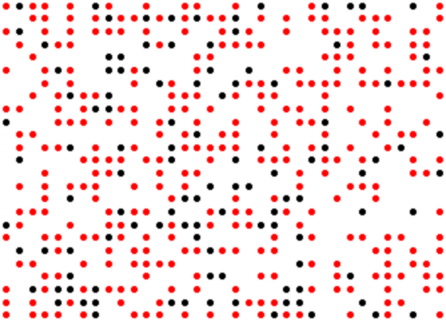
Speaking of quantifiers

Bob van Tiel

Michael Franke

Uli Sauerland

XPRAG.de annual meeting



— of the circles are red.

Generalised quantifiers

Many quantity words place thresholds on the absolute or relative intersection set size.

- 'some' means that $|S \cap P| > 0$
- 'fewer than half' means that $|S \cap P|/|S| < 0.5$
- 'most' means that $|S \cap P|/|S| > 0.5$
- 'exactly three' means that $|S \cap P| = 3$
- 'many' means that $|S \cap P| > t$

Monotonicity

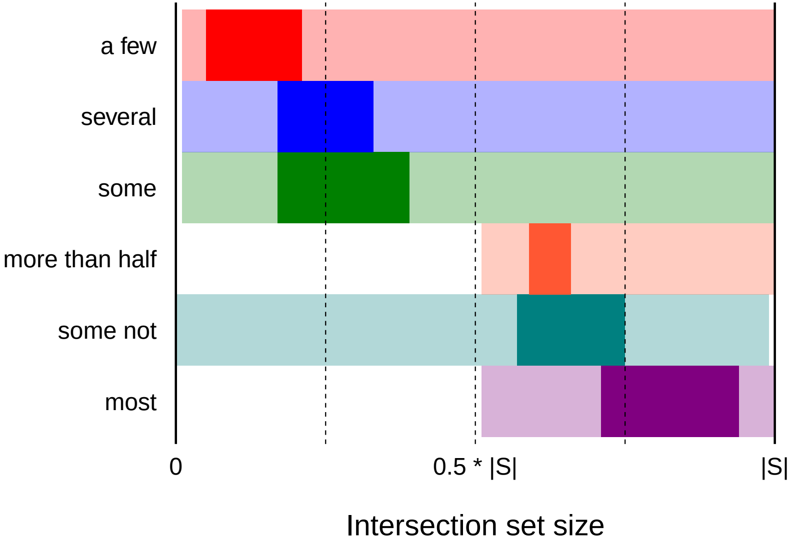
Quantity words can be…

- $>$ monotone: licensing inferences from sets to supersets.
- $>$ antitone: licensing inferences from sets to subsets.
- $>$ neither.

Monotone quantity words are learned earlier (Katsos et al., 2016) and processed faster (Geurts et al., 2010). Antitone quantity words allow for negative polarity items (Zwarts, 1981).

From meaning to use

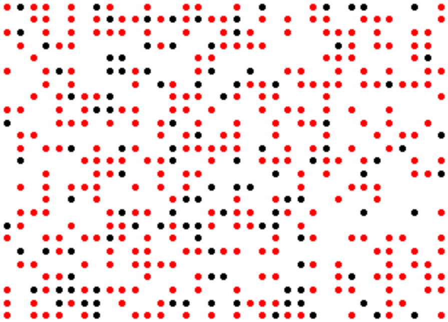
As it stands, GQ theory offers a poor account of the way speakers and hearers actually use quantity words (e.g, Hammerton, 1976; Newstead et al., 1987).



a few
several
some
more than half
some not
most
0
0.5 * |S|
|S|
Intersection set size

Production study

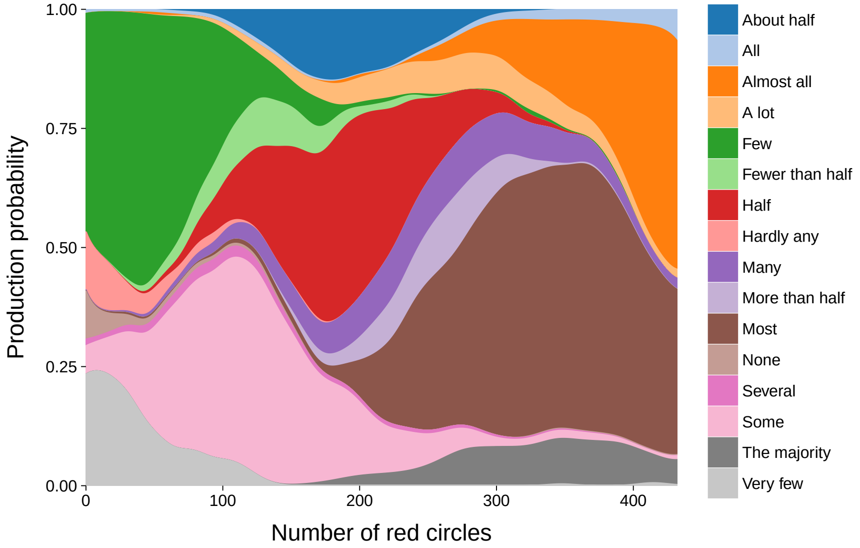
- > 600 participants on Mechanical Turk.
- > Displays consisting of 432 black or red circles.
- > Participants saw 10 displays.
- > They had to describe the number of red circles by completing the sentence '— of the circles are red'.
- > Number words were not allowed.



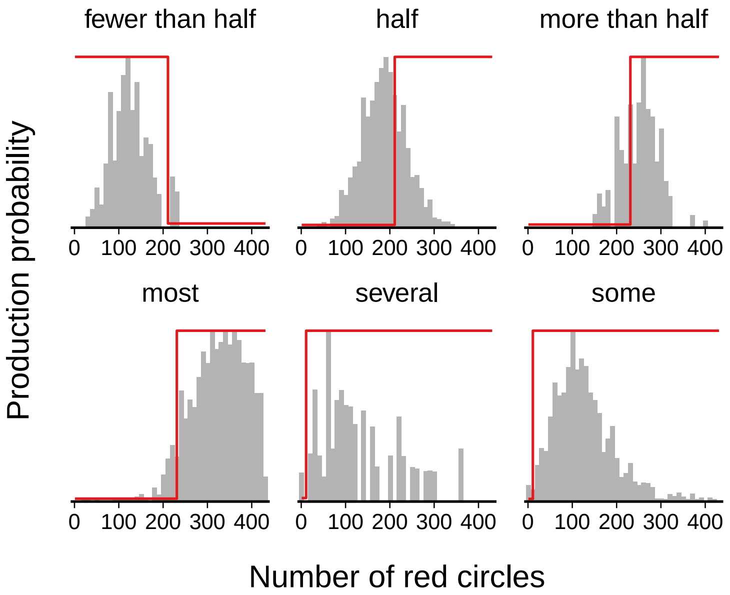
— of the circles are red.

Results

- 291 unique responses out of 5630 observations.
  - ‘all’
  - ‘less than a fraction of a third’
- Reduced to 35 types of quantity words.
- 16 types of quantity words produced more than 50 times.
- ‘Most’, ‘few’, ‘some’, ‘half’, and ‘almost all’ together make up more than half of the responses.



Production probability
1.00
0.75
0.50
0.25
0.00
0
100
200
300
400
Number of red circles
About half
All
Almost all
A lot
Few
Fewer than half
Half
Hardly any
Many
More than half
Most
None
Several
Some
The majority
Very few



fewer than half
half
more than half
most
several
some
Production probability
Number of red circles
0
100
200
300
400

Focal ranges

The focal ranges of quantity words often diverge from their semantic ranges. Can we explain the presence and location of these focal ranges within GQ theory?

Prolegomena

$>$ 433 possible intersection set sizes: $T = \{t_0, \ldots, t_{432}\}$.

$>$ 16 quantity words: $M = \{m_{\text{'a lot'}}, \ldots, m_{\text{'very few'}}\}$.

$>$ The probability that S produces m to refer to t depends on...

- the salience of m.
- the perceived intersection set size t'.
- the meaning of m.

Salience

The salience *Sal* of m is proportional to the frequency of m in the production study.

$$P_{Sal}(m) \propto n_m$$

Confusability

The perceived intersection set size t' depends on…

> the distance between t and the limits.

> the accuracy of the estimates.

We formulate confusability by analogy to confusion probabilities of the approximate number system (Dehaene, 1997).

$$P_{Conf}(t' \mid t\,;\,w) \propto P_{ANS}(t' \mid t\,;\,w)\; P_{ANS}(432 - t' \mid 432 - t\,;\,w)$$

$$P_{ANS}(t' \mid t\,;\,w) = \int_{t'-0.5}^{t+0.5} \mathcal{N}(x, \mu = t, \sigma = w\,t)\mathrm{d}x\,.$$

Meaning

B_θ maps pairs of m and t to the truth value of m in t. B_θ is parameterized, and we will use data-driven inference to determine best-fitting parameter values below.

The monotonicity of quantity words is gauged experimentally.

Monotonicity study

- 60 participants on Mechanical Turk.
- Participants were presented with arguments containing inferences from sets to subsets and supersets.
- They had to determine the validity of these arguments

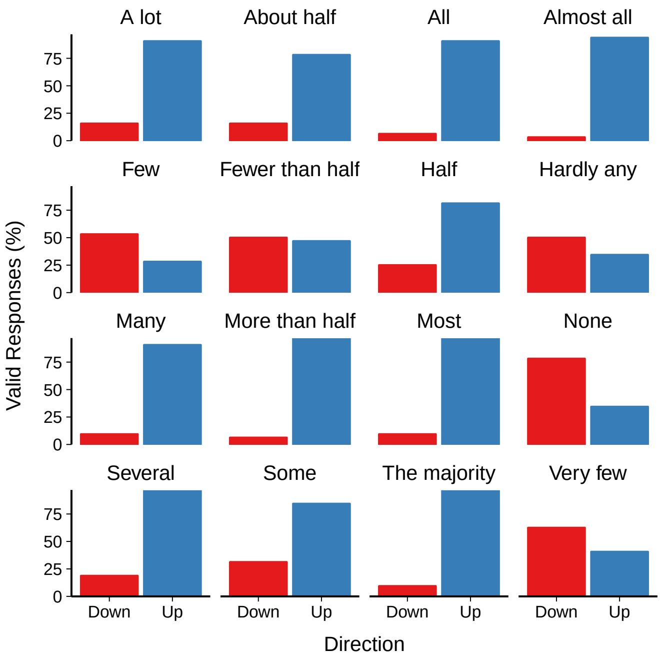
Upward argument

Premise:	Q of the people P1.
Conclusion:	Q of the people P2.

Downward argument

Premise:	Q of the people P2.
Conclusion:	Q of the people P1.

…where $P1 \subseteq P2$

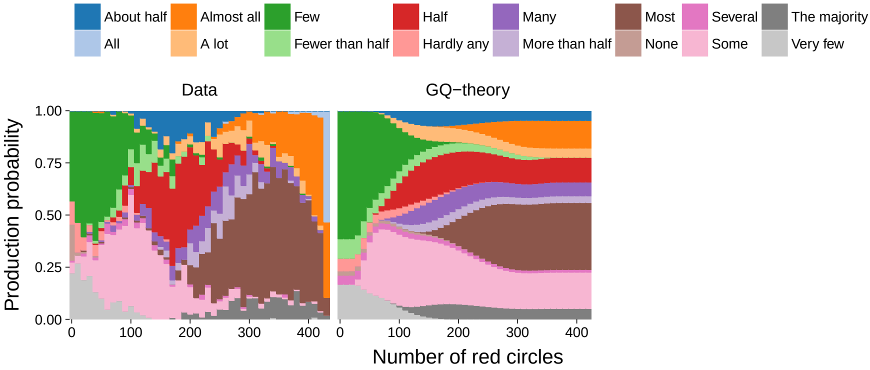


A lot
About half
All
Almost all
Few
Fewer than half
Half
Hardly any
Many
More than half
Most
None
Several
Some
The majority
Very few
0
25
50
75
Valid Responses (%)
Down
Up
Direction

A literal speaker

A literal speaker produces quantity words with a probability that is proportional to its semantic value.

$$P_{S_{Lit}}(m \mid t; \theta, w) \propto \sum_{t' \in T} P_{Conf}(t' \mid t; w)\, P_{Sal}(m)\, B_{\theta}(m, t')$$



About half
All
Almost all
A lot
Few
Fewer than half
Half
Hardly any
Many
More than half
Most
None
Several
Some
The majority
Very few
Data
GQ−theory
Production probability
1.00
0.75
0.50
0.25
0.00
0
100
200
300
400
Number of red circles

Model fit

The semantic model offers a reasonably good fit to the data (r = .809). But it fails to account for the focal ranges.

– ‘about half’ (r = .04)

– ‘fewer than half’ (r = .38)

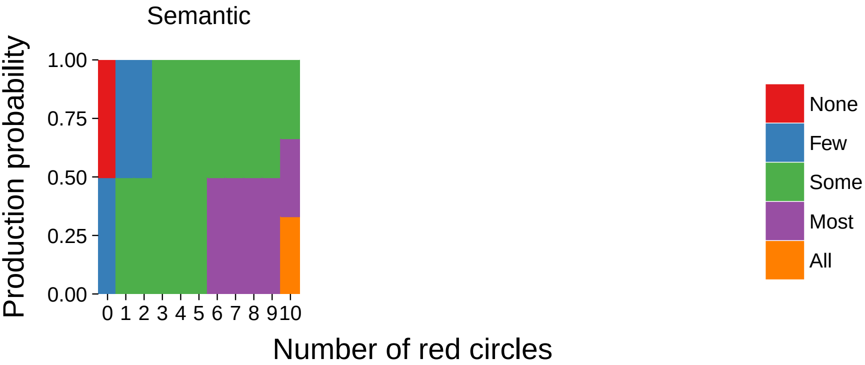
– ‘none’ (r = .36)

Can we improve the model by taking into consideration the pragmatics of communication?

Semantics

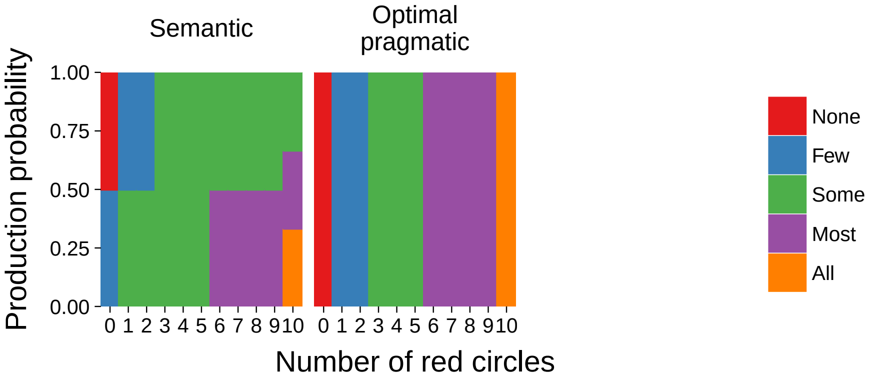
Suppose the speaker can produce 5 quantity words to describe displays with 10 circles.

- ‘no’ means that $|S \cap P| = 0$
- ‘few’ means that $|S \cap P| < 2$
- ‘some’ means that $|S \cap P| > 0$
- ‘most’ means that $|S \cap P| > 5$
- ‘all’ means that $|S \cap P| = 10$


Semantic
Production probability
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10
Number of red circles
None
Few
Some
Most
All

Pragmatics

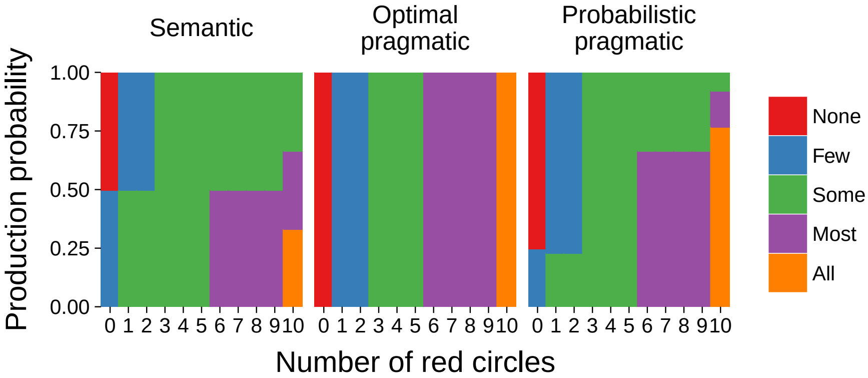
A completely rational pragmatic speaker chooses messages that are most likely to be interpreted correctly.



Semantic
Optimal pragmatic
Production probability
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10
0 1 2 3 4 5 6 7 8 9 10
Number of red circles
None
Few
Some
Most
All

Probabilistic pragmatics

A probabilistically pragmatic speaker would choose the optimal message with a probability proportional to her rationality.



Semantic
Optimal pragmatic
Probabilistic pragmatic
Production probability
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10
Number of red circles
None
Few
Some
Most
All

A pragmatic speaker

A pragmatic speaker optimises her behaviour based on how a naive hearer would behave.

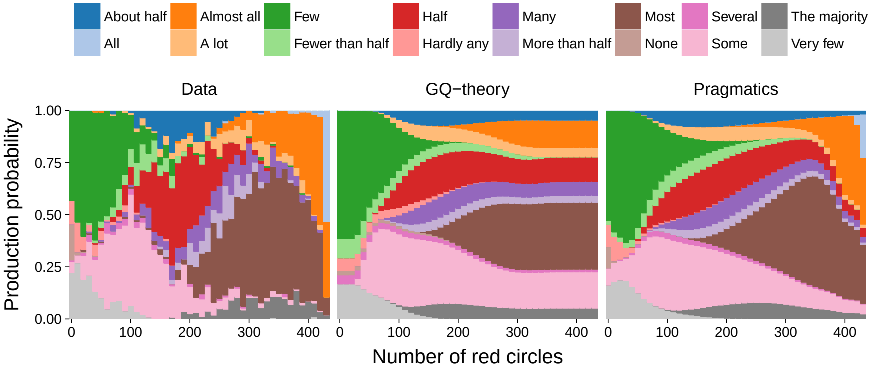
$$P_H(t \mid m) \propto B_\theta(m, t)$$

A pragmatic speaker optimises the probability of coordination with H. λ regulates how likely it is that the pragmatic speaker produces the optimal quantity word (Frank & Goodman, 2012).

$$P_S(m \mid t, \lambda) \propto P_H(t \mid m)^\lambda$$

Taken together, a pragmatic speaker is defined as follows.

$$P_{S_{Prg}}(m \mid t; \theta, w, \lambda) \propto \sum_{t' \in T} P_{Conf}(t' \mid t; w)\, P_{Sal}(m)\, P_S(m \mid t'; \theta, \lambda)$$



About half
All
Almost all
A lot
Few
Fewer than half
Half
Hardly any
Many
More than half
Most
None
Several
Some
The majority
Very few
Data
GQ−theory
Pragmatics
Production probability
1.00
0.75
0.50
0.25
0.00
0
100
200
300
400
Number of red circles

Results

- S_{Prg} offers a substantially better explanation of the production data than S_{Lit} (r = .809 vs. .924).
- The difference is most pronounced for…
  - 'about half' (r = .04 vs. .98).
  - 'fewer than half' (r = .38 vs. .82).
  - 'none' (r = .36 vs. 1.00).
- GQ-theory offers a compelling account of the use of quantity words if it is supplemented with a pragmatic module.

Conclusions

> Gradient structure may emerge as a consequence of pragmatic reasoning and constraints on perception.
>
> – Vagueness.
>
> – Prototypes.

> Further improvements to the model: individual differences in the location of the thresholds for vague quantity words.

Thank you!