Linguistics

ERASMUS UNIVERSITEIT ROTTERDAM

ENS ÉCOLE NORMALE SUPÉRIEURE

An experimental investigation of the around / between contrast

Adèle Hénot-Mortier (MIT), Steven Verheyen (EUR), Paul Égré & Benjamin Spector (CNRS, ENS-PSL, EHESS)

December 19, 2022

Amsterdam Colloquium 2022

Approximation and vagueness

(1) a. **Around 20** people came to the party.

b. **Between 15 and 25** people came to the party.

- Both (1a) and (1b) convey **uncertainty** about an exact numerical value k. Yet (1a) is **vague** while (1b) isn't!
- [Égré et al., 2022]: for any fixed prior distribution over k, (1a) gives rise to a more "**peaked**" posterior distribution than (1b) does.

We propose to test this prediction via an online probability-elicitation experiment.

Approximation and vagueness

(1) a. **Around 20** people came to the party.

b. **Between 15 and 25** people came to the party.

- Both (1a) and (1b) convey **uncertainty** about an exact numerical value k. Yet (1a) is **vague** while (1b) isn't!
- [Égré et al., 2022]: for any fixed prior distribution over k, (1a) gives rise to a more **"peaked"** posterior distribution than (1b) does.

We propose to test this prediction via an online probability-elicitation experiment.

Approximation and vagueness

(1) a. **Around 20** people came to the party.

b. **Between 15 and 25** people came to the party.

- Both (1a) and (1b) convey **uncertainty** about an exact numerical value k. Yet (1a) is **vague** while (1b) isn't!
- [Égré et al., 2022]: for any fixed prior distribution over k, (1a) gives rise to a more **"peaked"** posterior distribution than (1b) does.

We propose to test this prediction via an online probability-elicitation experiment.

Approximation and vagueness

(1) a. **Around 20** people came to the party.

b. **Between 15 and 25** people came to the party.

- Both (1a) and (1b) convey **uncertainty** about an exact numerical value k. Yet (1a) is **vague** while (1b) isn't!
- [Égré et al., 2022]: for any fixed prior distribution over k, (1a) gives rise to a more **"peaked"** posterior distribution than (1b) does.

We propose to test this prediction via an online probability-elicitation experiment.

Approximation and vagueness

(1) a. **Around 20** people came to the party.

b. **Between 15 and 25** people came to the party.

- Both (1a) and (1b) convey **uncertainty** about an exact numerical value k. Yet (1a) is **vague** while (1b) isn't!
- [Égré et al., 2022]: for any fixed prior distribution over k, (1a) gives rise to a more **"peaked"** posterior distribution than (1b) does.

We propose to test this prediction via an online probability-elicitation experiment.

A Bayesian model of approximation expressions

- Basic semantics: **around** and **between** denote **intervals**, but in the case of **around** the half-width of the interval is parametrized by a **free variable** i.
- A Bayesian listener $\mathcal{L}$ processing **around n** draws inferences about both i and k. An **around n**-update then redistributes more weight to values of k closer to n.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{around n}] \propto \mathcal{P}_{\mathcal{L}}[k] \sum_{i=|n-k|}^{n} \mathcal{P}_{\mathcal{L}}[i]$$

- A **between x and y**-update homogeneously rescales the weights for values of $k \in [x; y]$, and assigns a 0-weight to values of $k \notin [x; y]$.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{between x and y}] \propto \begin{cases} \mathcal{P}_{\mathcal{L}}[k] & \text{if} \quad k \in [x; y] \\ 0 & \text{if} \quad k \notin [x; y] \end{cases}$$

A Bayesian model of approximation expressions

- Basic semantics: **around** and **between** denote **intervals**, but in the case of **around** the half-width of the interval is parametrized by a **free variable** i.
- A Bayesian listener $\mathcal{L}$ processing **around n** draws inferences about both i and k. An **around n**-update then redistributes more weight to values of k closer to n.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{around n}] \propto \mathcal{P}_{\mathcal{L}}[k] \sum_{i=|n-k|}^{n} \mathcal{P}_{\mathcal{L}}[i]$$

- A **between x and y**-update homogeneously rescales the weights for values of $k \in [x; y]$, and assigns a 0-weight to values of $k \notin [x; y]$.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{between x and y}] \propto \begin{cases} \mathcal{P}_{\mathcal{L}}[k] & \text{if} \quad k \in [x; y] \\ 0 & \text{if} \quad k \notin [x; y] \end{cases}$$

A Bayesian model of approximation expressions

- Basic semantics: **around** and **between** denote **intervals**, but in the case of **around** the half-width of the interval is parametrized by a **free variable** i.
- A Bayesian listener $\mathcal{L}$ processing **around n** draws inferences about both i and k. An **around n**-update then redistributes more weight to values of k closer to n.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{around n}] \propto \mathcal{P}_{\mathcal{L}}[k] \sum_{i=|n-k|}^{n} \mathcal{P}_{\mathcal{L}}[i]$$

- A **between x and y**-update homogeneously rescales the weights for values of $k \in [x; y]$, and assigns a 0-weight to values of $k \notin [x; y]$.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{between x and y}] \propto \begin{cases} \mathcal{P}_{\mathcal{L}}[k] & \text{if} \quad k \in [x; y] \\ 0 & \text{if} \quad k \notin [x; y] \end{cases}$$

A Bayesian model of approximation expressions

- Basic semantics: **around** and **between** denote **intervals**, but in the case of **around** the half-width of the interval is parametrized by a **free variable** i.
- A Bayesian listener $\mathcal{L}$ processing **around n** draws inferences about both i and k. An **around n**-update then redistributes more weight to values of k closer to n.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{around n}] \propto \mathcal{P}_{\mathcal{L}}[k] \sum_{i=|n-k|}^{n} \mathcal{P}_{\mathcal{L}}[i]$$

- A **between x and y**-update homogeneously rescales the weights for values of $k \in [x; y]$, and assigns a 0-weight to values of $k \notin [x; y]$.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{between x and y}] \propto \begin{cases} \mathcal{P}_{\mathcal{L}}[k] & \text{if} \quad k \in [x; y] \\ 0 & \text{if} \quad k \notin [x; y] \end{cases}$$

A Bayesian model of approximation expressions

- Basic semantics: **around** and **between** denote **intervals**, but in the case of **around** the half-width of the interval is parametrized by a **free variable** i.
- A Bayesian listener $\mathcal{L}$ processing **around n** draws inferences about both i and k. An **around n**-update then redistributes more weight to values of k closer to n.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{around n}] \propto \mathcal{P}_{\mathcal{L}}[k] \sum_{i=|n-k|}^{n} \mathcal{P}_{\mathcal{L}}[i]$$

- A **between x and y**-update homogeneously rescales the weights for values of $k \in [x; y]$, and assigns a 0-weight to values of $k \notin [x; y]$.

$$\mathcal{P}_{\mathcal{L}}[k \mid \textbf{between x and y}] \propto \begin{cases} \mathcal{P}_{\mathcal{L}}[k] & \text{if} \quad k \in [x; y] \\ 0 & \text{if} \quad k \notin [x; y] \end{cases}$$

Prediction [Égré et al., 2022]

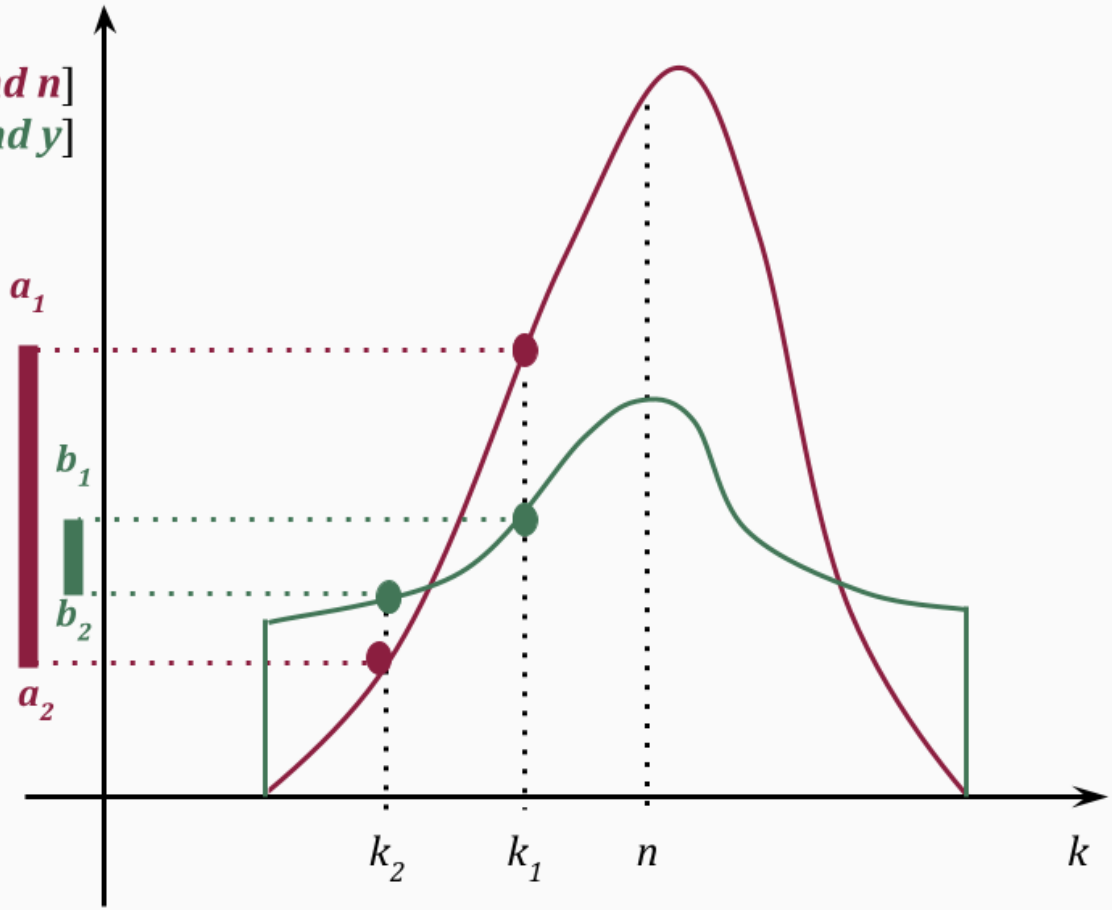
$\mathbf{P}_L[k \mid \textit{around } n]$
$\mathbf{P}_L[k \mid \textit{between } x \textit{ and } y]$

$\forall\ k_1, k_2\ .\ |n-k_1| < |n-k_2|$:

$$\frac{a_1}{a_2} \geq \frac{b_1}{b_2}$$

Average over all possible k_1, k_2

$$R_a \geq R_b$$



Experiment

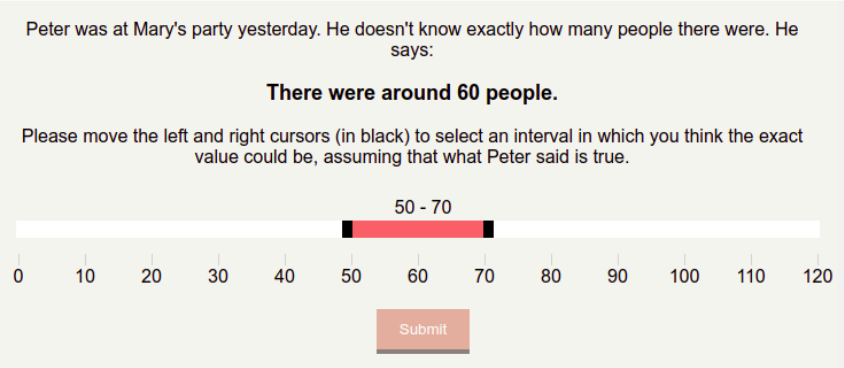
- Preregistered experiment conducted on 240 MTurk participants presented with matching **around n** (n=40, 50 or 60) and **between x and y** statements.
- Two tasks designed to elicit R_a and R_b:
  - Define a suitable **interval** of possible numerical values compatible with the expression (**around** or **between**);
  - Assign **weights** to the selected values according to how likely each value is to be the exact one.

Experiment

- Preregistered experiment conducted on 240 MTurk participants presented with matching **around n** (n=40, 50 or 60) and **between x and y** statements.
- Two tasks designed to elicit R_a and R_b:
  - Define a suitable **interval** of possible numerical values compatible with the expression (**around** or **between**);
  - Assign **weights** to the selected values according to how likely each value is to be the exact one.

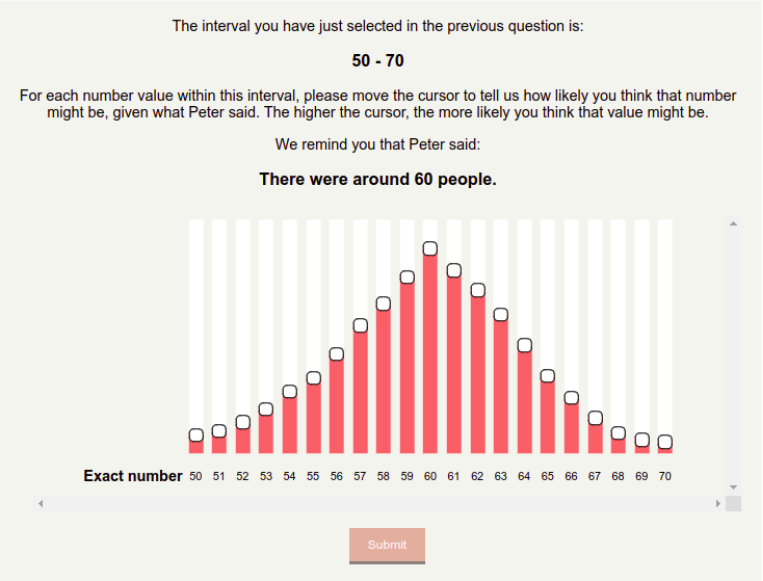
Experiment

- Preregistered experiment conducted on 240 MTurk participants presented with matching **around n** (n=40, 50 or 60) and **between x and y** statements.
- Two tasks designed to elicit R_a and R_b:
  - Define a suitable **interval** of possible numerical values compatible with the expression (**around** or **between**);
  - Assign **weights** to the selected values according to how likely each value is to be the exact one.



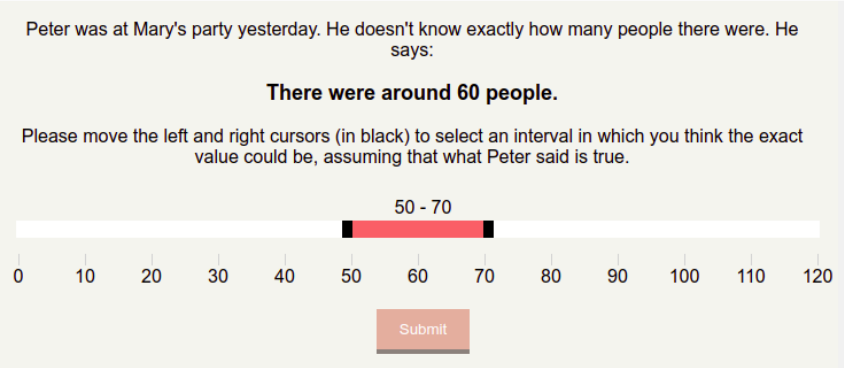
Experiment

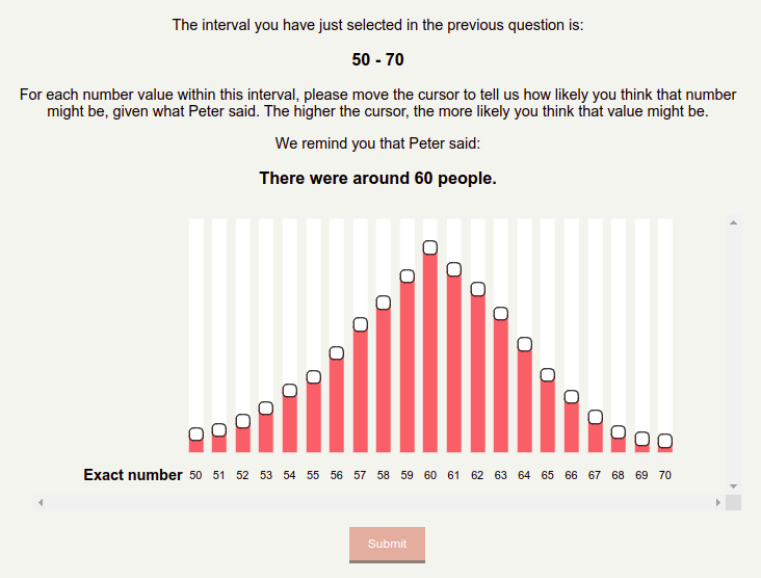
- Preregistered experiment conducted on 240 MTurk participants presented with matching **around n** (n=40, 50 or 60) and **between x and y** statements.
- Two tasks designed to elicit R_a and R_b:
  - Define a suitable **interval** of possible numerical values compatible with the expression (**around** or **between**);
  - Assign **weights** to the selected values according to how likely each value is to be the exact one.



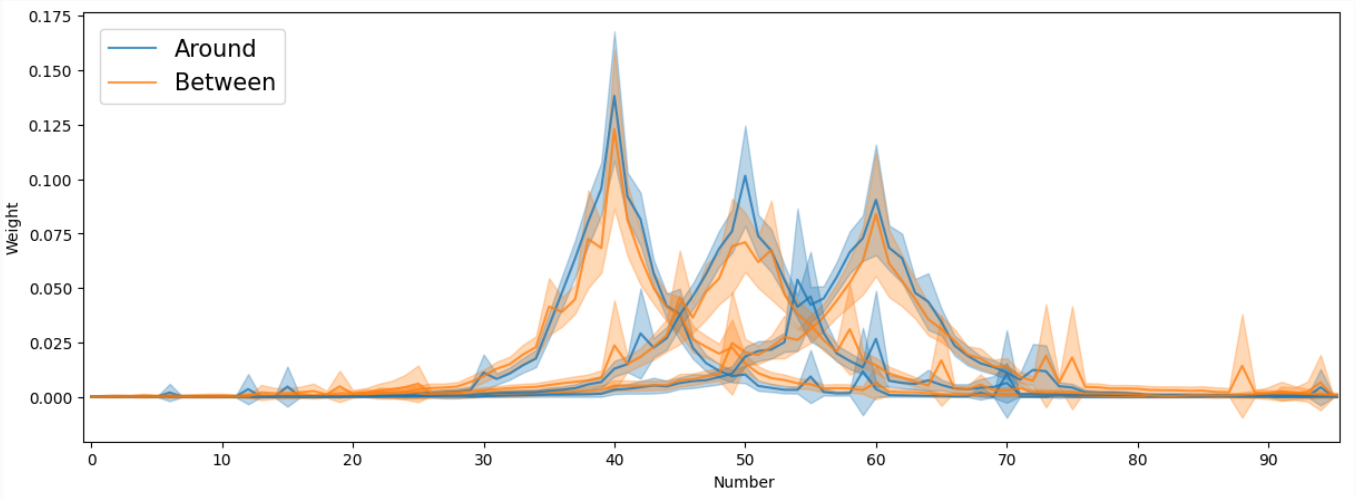
Experiment

- Preregistered experiment conducted on 240 MTurk participants presented with matching **around n** (n=40, 50 or 60) and **between x and y** statements.
- Two tasks designed to elicit R_a and R_b:
  - Define a suitable **interval** of possible numerical values compatible with the expression (**around** or **between**);
  - Assign **weights** to the selected values according to how likely each value is to be the exact one.





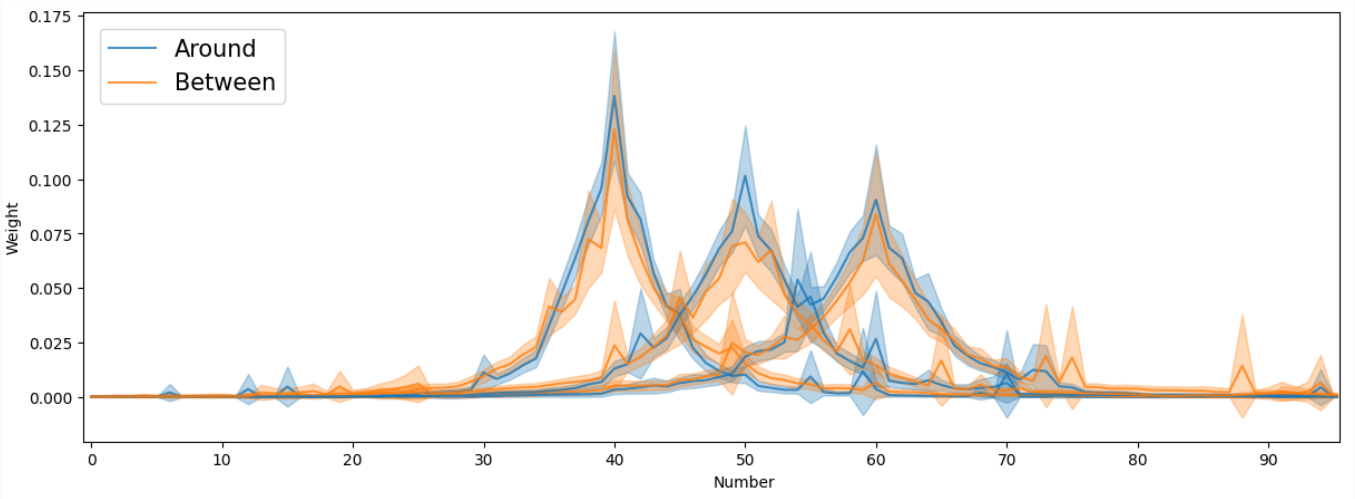
Results



Curious about the results ?

Come see our poster !

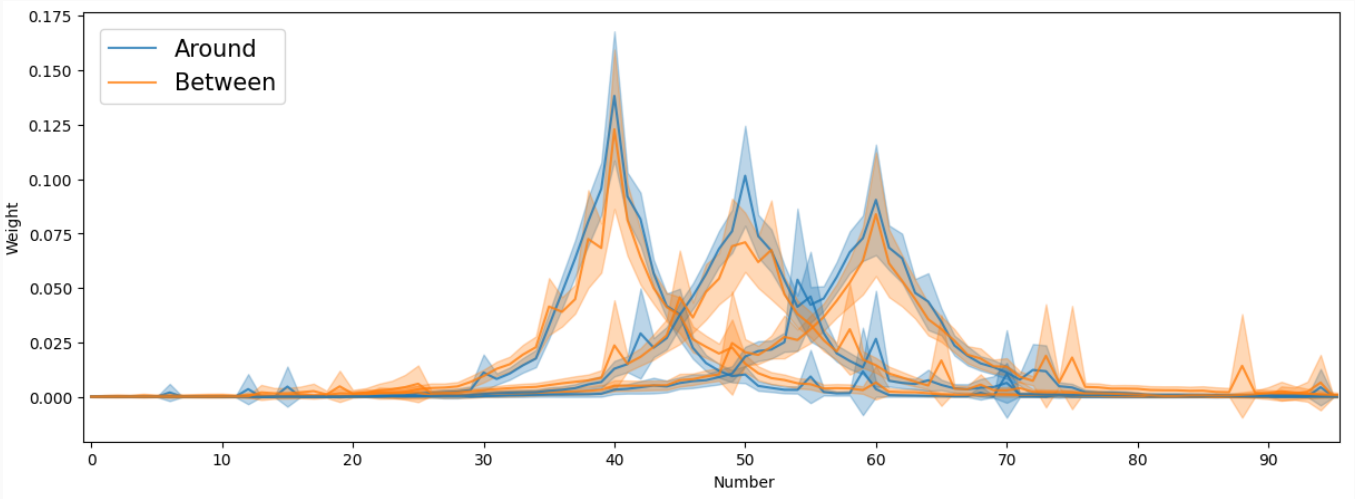
Results



Curious about the results ?

Come see our poster !

Results



Curious about the results ?

Come see our poster !

Selected references i

Bergen, L., Levy, R., and Goodman, N. (2016).

Pragmatic reasoning through semantic inference.

Semantics and Pragmatics.

Lassiter, D. and Goodman, N. D. (2013).

Context, scale structure, and statistics in the interpretation of positive-form adjectives.

SALT.

Qing, C. and Franke, M. (2015).

Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model.

SALT.

Égré, P., Spector, B., Mortier, A., and Verheyen, S. (2022).

On the optimality of vagueness: "around", "between", and the gricean maxims.

L&P.