Spatial References and Perspective in Natural Language Instructions for Collaborative Manipulation

Rosario Scalise, Shen Li
Henny Admoni, Stephanie Rosenthal, Siddhartha S. Srinivasa
“I am going to pick up the cup on the right!”
Key Issue: Ambiguity
Key Issue: Ambiguity

As scene complexity increases, so does the difficulty in specifying an object.
Key Issue: Ambiguity

As scene complexity increases, so does the difficulty in specifying an object.

Natural language is inherently ambiguous.
Forms of Ambiguity

Visual Appearance

“Pick up the coffee cup.”
Forms of Ambiguity

Visual Appearance

“Pick up the coffee cup.”

Which one?
Forms of Ambiguity

Perspective

“Pick up the coffee cup on the right.”
Forms of Ambiguity

Perspective

“Pick up the coffee cup on the right.”

Whose right?
Forms of Ambiguity

Proximity

“Pick up the coffee cup next to the donuts.”
Forms of Ambiguity

Proximity

“Pick up the coffee cup next to the donuts.”

How close is ‘next to’?
Can you uniquely describe this block?
How can we best **overcome ambiguity** when grounding our references **while** keeping communication natural?
Approach

Learn by observing what humans do and extract best-practices from the examples that are most successful.
Collect Corpus
Collect Corpus
Gain Insights

bender by Jordan Díaz Andrés from the Noun Project
Collect Corpus
Gain Insights

Evaluate Corpus

bender by Jordan Díaz Andrés from the Noun Project
Collect Corpus
Gain Insights

Evaluate Corpus
Extract Guidelines
Study 1: Collecting Instructions for Corpus

Scenario #1/14

You are facing the table just as it appears in the image, and on the other side of the table is a person represented by the silhouetted figure. How would you instruct the person to pick up the indicated block? (They cannot see the red arrow):

How difficult did you find it to answer this prompt?

- Very Difficult
- Difficult
- Neither difficult nor easy
- Easy
- Very Easy
Study 1 : Collecting Instructions for Corpus

Scenario #1/14

You are facing the table just as it appears in the image, and on the other side of the table is a person represented by the silhouetted figure.

How would you instruct the person to pick up the indicated block? (They cannot see the red arrow):

How would you instruct the person represented by the silhouetted figure to pick up the indicated block?

How difficult did you find it to answer this prompt?

- Very Difficult
- Difficult
- Neither difficult nor easy
- Easy
- Very Easy
Study 1: Collecting Instructions for Corpus

Scenario #1/14

You are facing the table just as it appears in the image, and on the other side of the table is a robot represented by the silhouetted figure. How would you instruct the person to pick up the indicated block? (They cannot see the red arrow):

How would you instruct the person represented by the silhouetted figure to pick up the indicated block?

How difficult did you find it to answer this prompt?

- Very Difficult
- Difficult
- Neither difficult nor easy
- Easy
- Very Easy
Study 1: Collecting Instructions for Corpus

Scenario #1/14

You are facing the table just as it appears in the image, and on the other side of the table is a robot represented by the silhouetted figure.

How would you instruct the person to pick up the indicated block? (They cannot see the red arrow):

How difficult did you find it to answer this prompt?

- Very Difficult
- Difficult
- Neither difficult nor easy
- Easy
- Very Easy

1400 Total
Evaluating

How do we tell how good any specific instruction is?

“Pick up the blue block”
Evaluating

Given an instruction and the stimulus it corresponds to, can people infer the correct block?

“Pick up the blue block”
Evaluating

Given an instruction and the stimulus it corresponds to, can people infer the correct block?

“Pick up the blue block”
Scenario #3/40

You are seated across the table from the silhouetted figure. You have just asked the silhouetted figure to pick up a block. The instructions you gave are shown in the blue text below:

Pick up the green block that is closest to you and on the right.

Which block do you expect the silhouetted figure to pick up? The red circle shows the block, click to show a black and white circle with your selection. Once you are satisfied with your selection, please click the next button to move on.
Metrics

For each instruction, we calculate:
Metrics

For each instruction, we calculate:

\[
\text{Accuracy:} \quad \frac{\text{# of successful block selections}}{\text{total # of times instruction is shown}}
\]
Metrics

For each instruction, we calculate:

**Accuracy**: \[rac{\text{# of successful block selections}}{\text{total # of times instruction is shown}}\]

**Avg. Completion time**: How long it takes to select the indicated block on average
Spatial Reference
Perspective
Pick up the box furthest to your left.

Partner perspective
Pick up the box furthest to your left.

Partner perspective
Pick up the box **furthest** to your left.

Partner perspective
Pick up the orange block closest to my right hand side.

Participant perspective
Pick up the orange block closest to my right hand side.

Participant perspective
Pick up the orange block closest to my right hand side.

Participant perspective
Please pick up the orange block that is closest to me.

Neither perspective
Please pick up the orange block that is **closest to me**.

Neither perspective
Pick up the rightmost orange block

Unknown perspective
Pick up the **rightmost** orange block

Right to ???

Unknown perspective
Neither Perspective is better
The bar chart shows the completion time (in seconds) for different conditions.

- **Unknown**: The completion time is significantly lower than for other conditions ($p < 0.001$).
- **Neither**: The completion time is significantly lower than for other conditions ($p < 0.001$).
- **Participant**: The completion time is significantly lower than for other conditions ($p < 0.001$).
- **Partner**: The completion time is significantly lower than for other conditions ($p < 0.001$).
Prefer Neither Perspective
Other Factors
Pick the blue block that is closer to you and right next to the yellow block

Neither perspective
Pick the **blue block** that is closer to you and right next to the yellow block.

Neither perspective
Pick the blue block that is closer to you and right next to the yellow block.

Neither perspective
Pick up the blue block on your far right.

Partner perspective
Pick up the **blue** block on your far right.

Partner perspective
Tradeoff
Neither Perspective
Be Explicit
Robot Partner vs Human Partner
Pick up the third blue block from your left.
Partner Perspective
Thank You!

Our dataset will be made available soon!
Spatial References and Perspective in Natural Language Instructions for Collaborative Manipulation

Rosario Scalise, Shen Li
rscalise@andrew.cmu.edu, shenli@cmu.edu