



Question: Why is person on the right pointing to the person on the left?

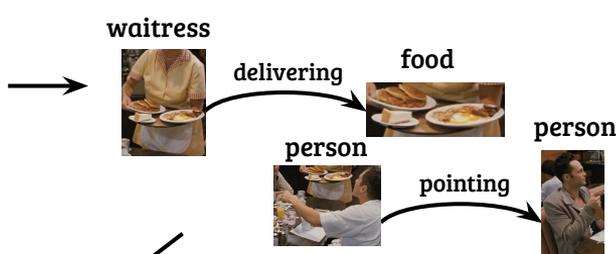
Answer: He is telling the waitress that the person on the left ordered the pancakes.

Natural language rationale: [The answer is true because] she is delivering food to the table and she doesn't know whose order is whose.

Raw features



Relations (semantics)



Inferences (pragmatics)

The waitress doesn't know whose order is whose.

NATURAL LANGUAGE RATIONALES WITH FULL-STACK VISUAL REASONING

What are natural language rationales?

Why natural language rationales?

What is the key challenge of visual-textual rationalization?

What is our approach?

How does it work?

What's next?

<https://arxiv.org/abs/2010.07526>