

PHYSICALLY SITUATED DIALOGUE

Natalie Parde
parde@uic.edu

CS 594: Language and Vision
Spring 2019



What is physically situated dialogue?

Dialogue for which the surrounding environment provides physical context to the language employed.

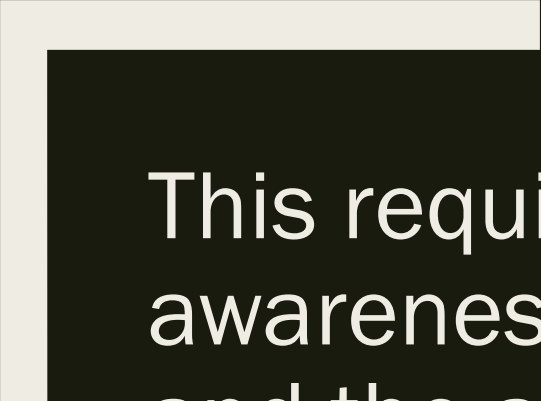


What are some example situational descriptors?

- Prepositions:
 - *Above, across, against, around, before, below, near, on, over, under, within....*
- Adjectives:
 - *Tall, short, long, wide, large, small, dark, light*
 - *Red, blue, orange, pink, fluffy, loud*
- Quantifiers:
 - *More, less*

Situational descriptors can change depending on context.





This requires awareness of and the ability to reason about objects and actors in the environment.

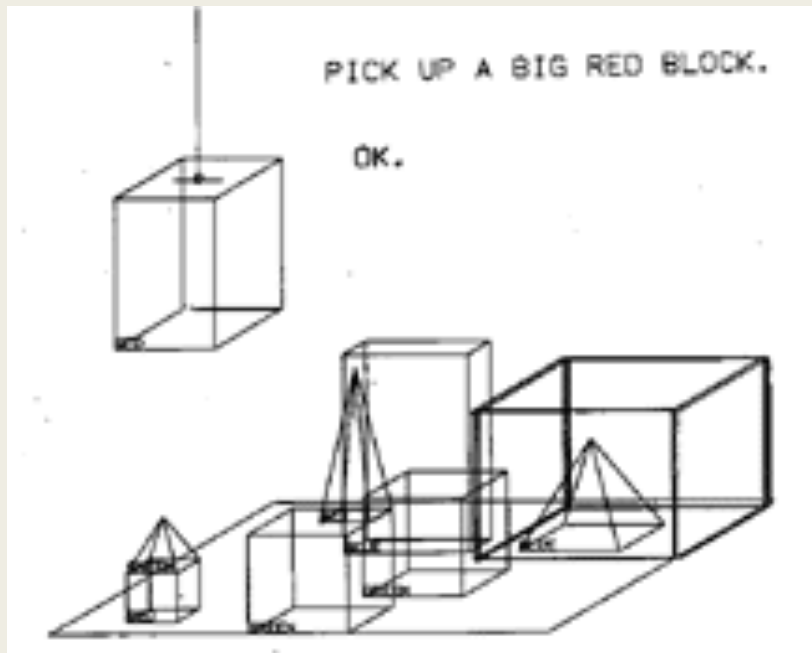
- Questions we may need to answer:
 - *What is relevant to understanding these terms?*
 - *What do we already know about this location?*
 - *What relationships can we reasonably expect to hold true?*
- <https://youtu.be/13KsDgmV0vE?t=27>

Early Research on Physically Situated Dialogue

- SHRDLU¹
 - *Program in which a human carried on a conversation with a computer about blocks in a virtual world*
 - *Limited world (blocks on a table) with language defined by a formal grammar*

¹Winograd, Terry. *Procedures as a representation for data in a computer program for understanding natural language*. No. MAC-TR-84. Massachusetts Institute of Technology Artificial Intelligence Technical Report, 1971.

SHRDLU



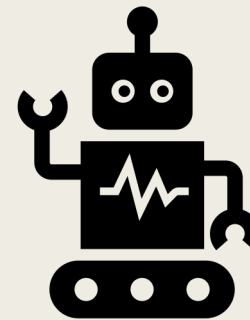
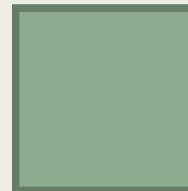
<https://hci.stanford.edu/winograd/shrdlu/shrdlu-original.gif>

- Sample dialogue:
 - *PICK UP A BIG RED BLOCK.*
 - *OK.*
 - *GRASP THE PYRAMID.*
 - *I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.*
 - *FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.*
 - *BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.*
 - *OK.*
 - *WHAT DOES THE BOX CONTAIN?*
 - *THE BLUE PYRAMID AND THE BLUE BLOCK.*

- Video footage:
<https://youtu.be/QAJz4YKUwqw>

Key Attributes of Physically Situated Dialogue

- Occurs between dialogue partners inhabiting a shared physical space
- Refers to the shared surroundings of those dialogue partners
- Involves dialogue partners (e.g., robotic agents) capable of performing actions in this space



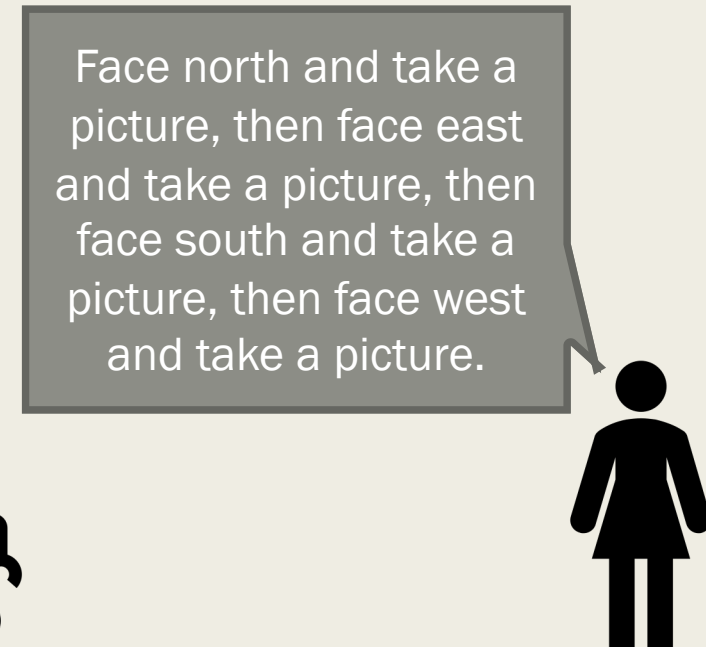
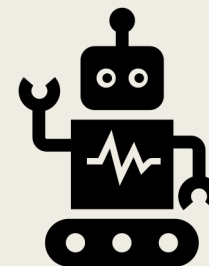
Important Considerations

■ Turn-taking

- *How quickly should a robot respond to dialogue or actions?*
 - When should it assume the dialogue partner has finished a turn?
- *Should a robot ever interrupt its dialogue partner?*
- *How should a robot handle interruptions to its speech or actions?*
- *Human partners engaging in task-oriented situated dialogue often exhibit speech overlap (beginning to speak before the previous speaker has finished).*
- https://youtu.be/Gx3_W6GI7eY

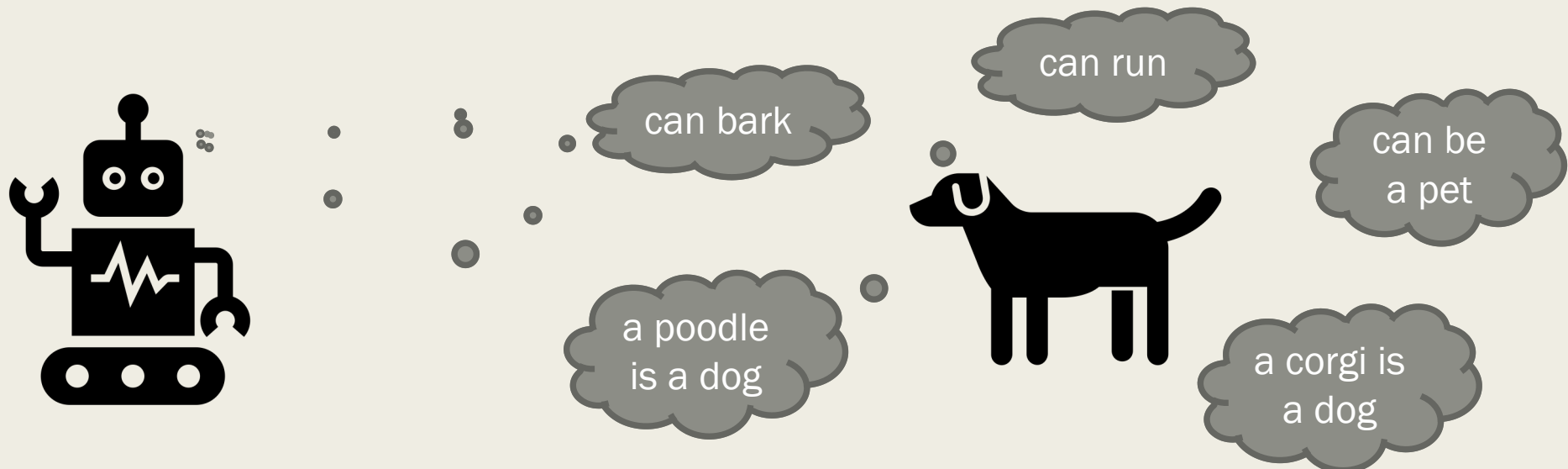
Important Considerations

- Stylistic Differences
 - *Some users are much more specific than others!*



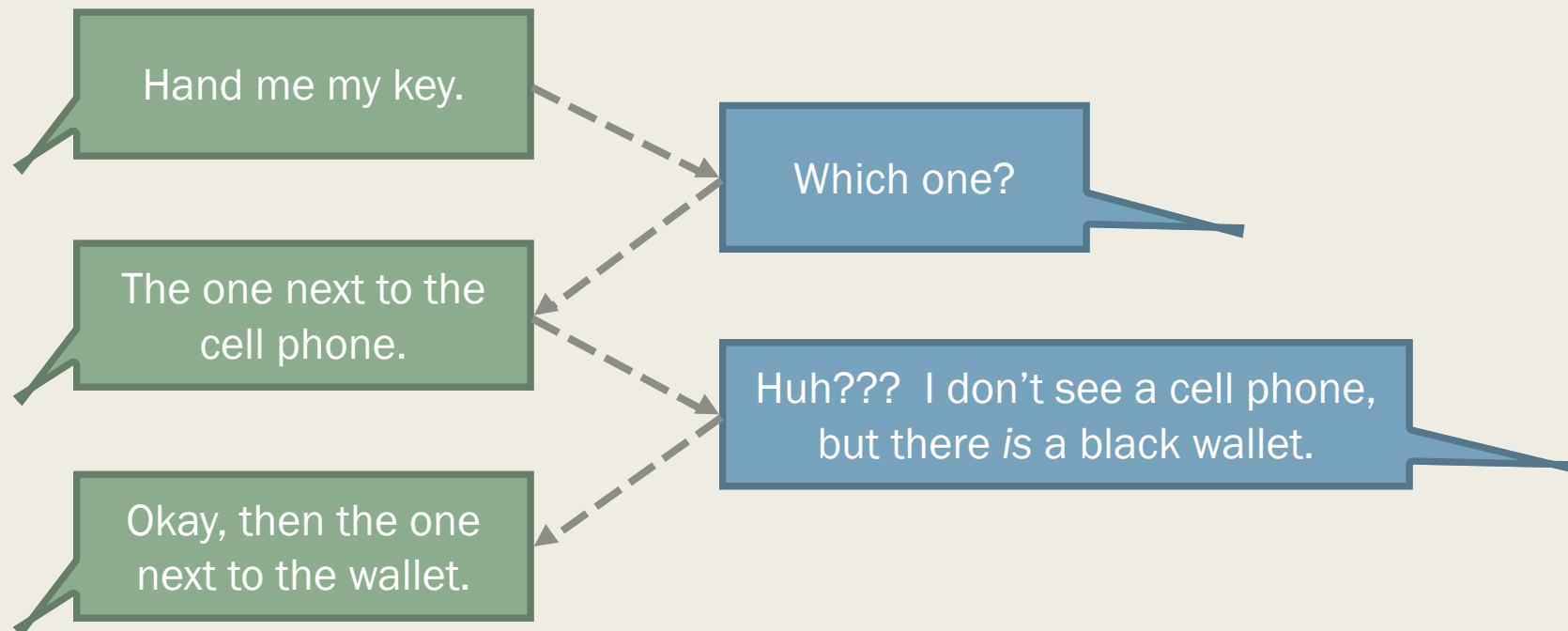
Important Considerations

- Commonsense reasoning
 - *How should commonsense knowledge bases be built?*
 - *How specific should these knowledge bases be?*
 - <http://conceptnet.io/>



Important Considerations

- Handling discrepancies
 - *Two dialogue partners are unlikely to have completely consistent worldviews*



Important Considerations

- World modeling
 - *What is the best way to represent an agent's physical surroundings?*
 - One option: Simulated models
 - <https://youtu.be/gVj-TeIJfrk>
 - Recent interest in predicting scene representations from limited observations
 - <https://youtu.be/oSZkDuDoFAI>

Important Considerations

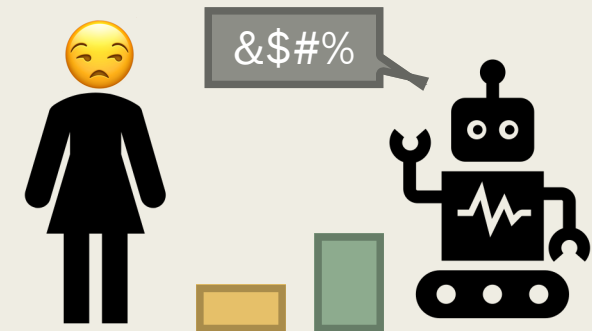
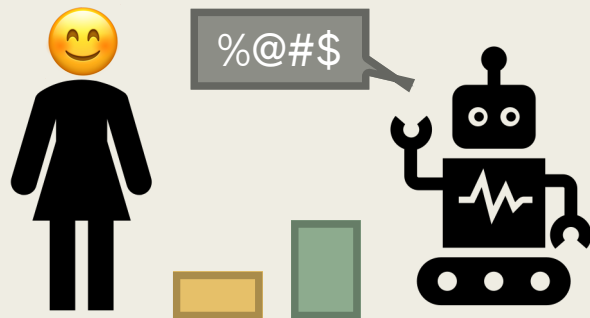
- Referring expressions
 - *How to identify appropriate referential terms when information is uncertain and may or may not be readily accessible?*



I want the **yellow**
M&M on the top left.

Important Considerations

- Social expectations
 - *Some related to turn-taking:*
 - Who should start talking next?
 - When should the next person/agent begin speaking?
 - *Others related to dialogue management and language generation:*
 - When should it be assumed that dialogue repairs need to be made?
 - How should the next turn be phrased, given what has taken place?





Why is physically situated dialogue important?

- Human-human interaction is highly coordinated with respect to the physical environment
...human-machine interaction should be too!
- Context matters
 - *Often changes meaning*
 - *Impacts language recognition*

Resources

Recent Publication Venues:

- *SIGDIAL 2018 Special Session on Physically Situated Dialogue*: <https://robodial.github.io/>
- *NAACL 2018 Workshop on Spatial Language Understanding*: https://spatial-language.github.io/old_SpLU_workshops/SpLU_2018/

Interesting lectures:

- *Mediating Shared Bases towards Situated Dialogue*, by Joyce Chai: <https://youtu.be/sun8UPMGE2k>
- *Simultaneously Learning and Grounding Object Features and Spatial Relations*, by Anthony Cohn: <https://youtu.be/HGEDFY15zM0>

Datasets:

- <https://cs.stanford.edu/people/jcjohns/clevr/>
- <http://lic.nlp.cornell.edu/nlvr/>

Wrapping up....

- Overview
- Early research on physically situated dialogue
- Key attributes
- Important considerations
- Resources