Generating and Interpreting Referring Expressions with Vague Spatial Language

Adam Richard-Bollans
mm15alrb@leeds.ac.uk

Background

- We aim to create a semantic model of spatial prepositions which:
  - Can be incorporated into a situated dialogue system to aid referring expression comprehension and generation
  - Support existing theories of spatial language

Complexity & Vagueness

- Many features influence spatial preposition usage [1] and there are no clear boundaries demarcating when a preposition is, or is not, appropriate to use
- As well as representing geometric concepts, spatial prepositions denote functional relationships [2,3]
- Figure 1 provides an illustrative example. In (a) the pear is generally considered to be 'in' the bowl, whereas in (b) it is not

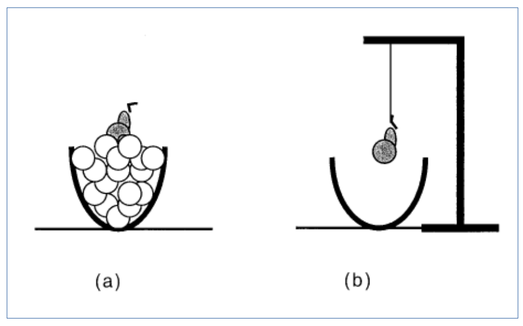


Figure 1 (from [2]). When is the pear 'in' the bowl?

Data Collection

- Lack of rich data including variety of salient features
- We have constructed a framework for data collection which allows easy feature extraction and the creation of varied environments and tasks
- We are currently running a study online[1]

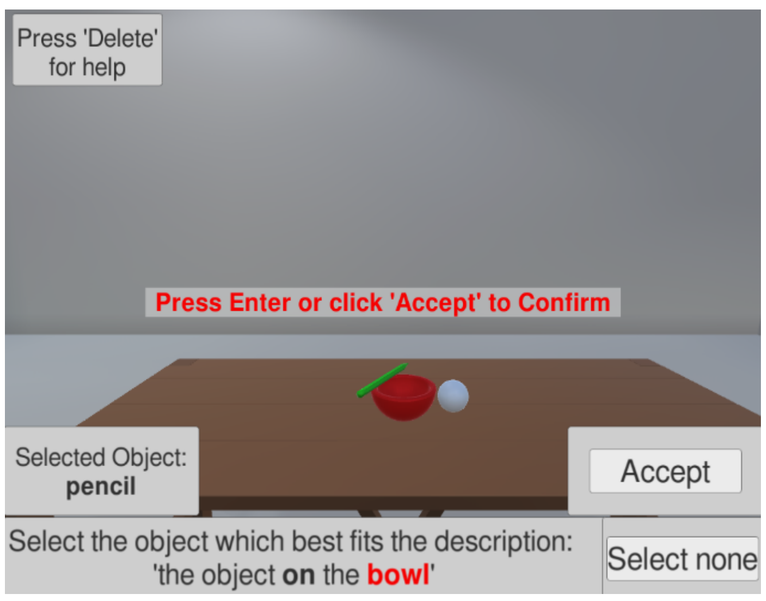


Figure 2. Data Collection Environment

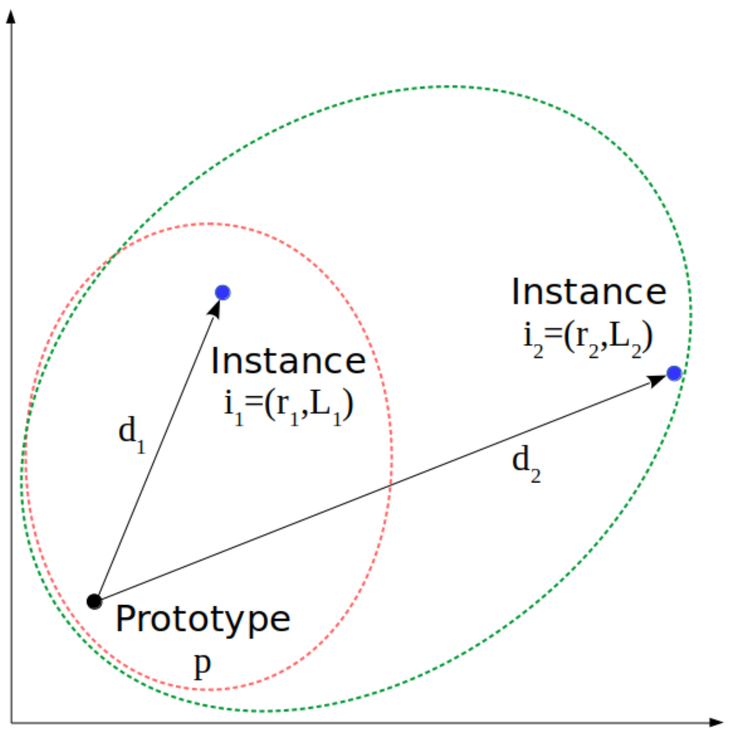
1. adamrichard-bollans.co.uk/spatial_language_project.html

Testing

- Currently in the process of constructing a virtual game environment for testing the dialogue model
- In order to achieve a goal, such as collecting specific objects in a scene, users must provide descriptions for the model to interpret; and vice versa
- This will allow for refining and testing pragmatic aspects of the model

Semantic Similarity

- Typicality is an important notion in language generation and interpretation
- We measure typicality of an instance (pair of objects) with respect to a description, D (a preposition, object pair), as the semantic similarity to a prototype for D
- First, semantic distance is calculated in a feature space using a weighted metric



- Semantic similarity is calculated as a decaying function of distance, where α is a constant representing the specificity of D:

$$s_D(i_1, p) = e^{(-\alpha \cdot d_1)}$$

Dialogue Model

- From the semantic similarity we can assess which objects in a scene best fit a description
- However, there are various *pragmatic* considerations to make
- A probabilistic model is appropriate for vague language [4]
- Naive model of interpretation --- pick object, r, which maximizes the following:

$$P(r|D) = s_D(i, p) \times \lambda(r)$$

where i denotes the instance representing r and the object given in D and $\lambda(\mathrm{r})$ denotes the salience of object r

- Pragmatic principles guiding collaborative communication suggest we should also account for:

$$P(D|r) = \frac{P(r|D) \times P(D)}{\sum_{D' \in \Pi} P(r|D') \times P(D')}$$

$\mathrm{P(D)}$ denotes the overall likelihood of providing description D and Π is the set of possible descriptions

- When generating and interpreting expression we aim to maximize both probabilities

References

[1] Richard-Bollans, A. (2018). Towards a Cognitive Model of the Semantics of Spatial Prepositions. In ESSLLI Student Session Proceedings. Springer. **[2]** Garrod, S., Ferrier, G., & Campbell, S. (1999). In and on: investigating the functional geometry of spatial prepositions. *Cognition*, 72(2), 167-189. **[3]** Herskovits, A. (1987). Language and spatial cognition. Cambridge University Press. **[4]** Mast, V., Falomir, Z., & Wolter, D. (2016). Probabilistic reference and grounding with PRAGR for dialogues with robots. Journal of Experimental & Theoretical Artificial Intelligence, 28(5), 889-911.