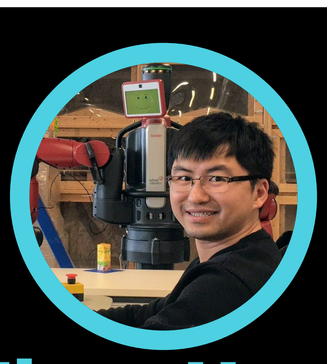
Evaluating **Referring Form Selection Models** *in* **Partially-Known Environments**



Zhao Han Polina Rygina Tom Williams

MIRRORLab, Department of Computer Science, Colorado School of Mines, USA

Read the Paper!

mirrorlab.mines.edu/publications/han2022inlg/

Zhao Han (he/him)

Follow: *@hanzhao*

Visit: *zhaohanphd.com*

Contact: *zhaohan@mines.edu*

MIRRORLab
Mines Interactive Robotics Research

Follow: *@MIRRORLab*

Visit: *mirrorlab.mines.edu*

Contact: *twilliams@mines.edu*

CS@Mines

Follow: *@CSatMines*

Visit: *cs.mines.edu*

Sponsor

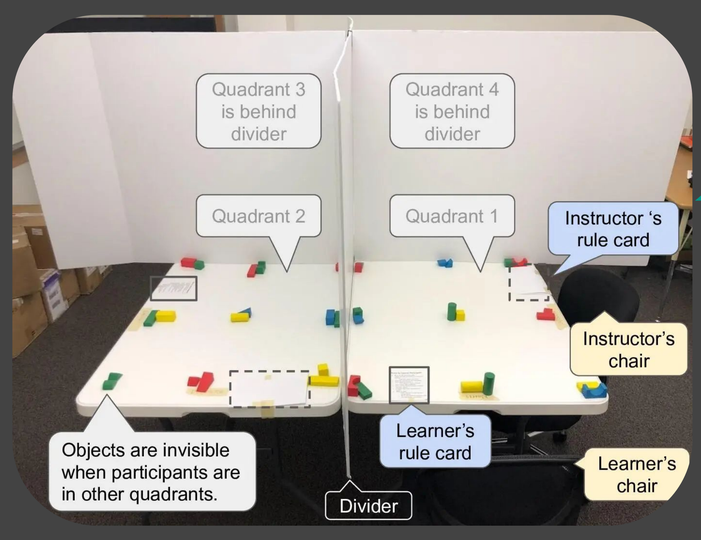
This work was funded in part by:

Motivation (Ecological Validity)

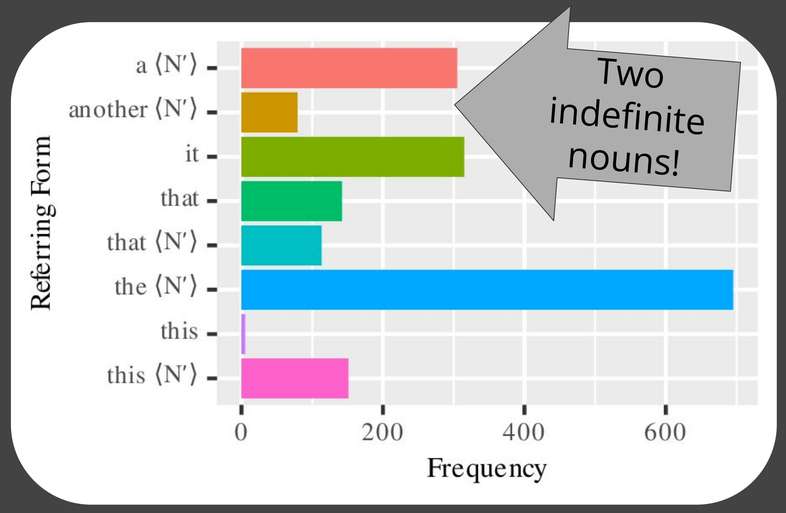
- To develop capable autonomous agents such as robots:
  - They need to **effectively communicate with humans**
  - They must be able to **refer to different entities** in ***situated*** contexts
- Recent attempts include **modelling the selection of referring forms on the basis of cognitive status (informed by Givenness Hierarchy)**
  - Previous approaches **showed promising results**: over 80% accuracy
  - Yet it **lacks ecological validity**. Task environment has few (11), constantly activated and always visible objects, encouraging only a subset of referring forms, e.g., “this” and proper nouns)

How can we **comprehensively model** the ***selection*** of **referring forms** and **evaluate** it?

- We designed a novel building-construction task with ***many (72),*** **repeated** objects in a ***partially-observable*** environment.



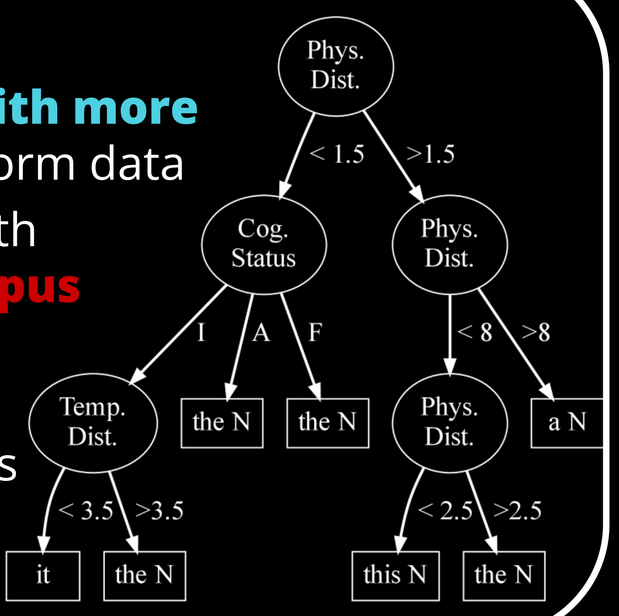
- We ran dyadic human-subjects study in **four quadrants** and **collected a wide variety of referring forms:**



- We **reassessed a decision tree model** (last figure) with our new comprehensive corpus and found **20% accuracy drop** to ~60%.

Takeaways & Future Work

- We proposed **a novel, situated task context with more and invisible objects** for full range of referring form data
- We re-assessed performance of existing model with the corpus and saw **20% drop with our new corpus**
- Performance drop show **more, non-uniquely identifiable, repeated, invisible objects are useful** to evaluate referring form selection models
- We plan to model gestures & explore other features unique in the task (e.g. visibility)



This work was funded in part by: Department of the Navy, ONR, Science & Technology