Referring as a collaborative process: learning to ground language through language games

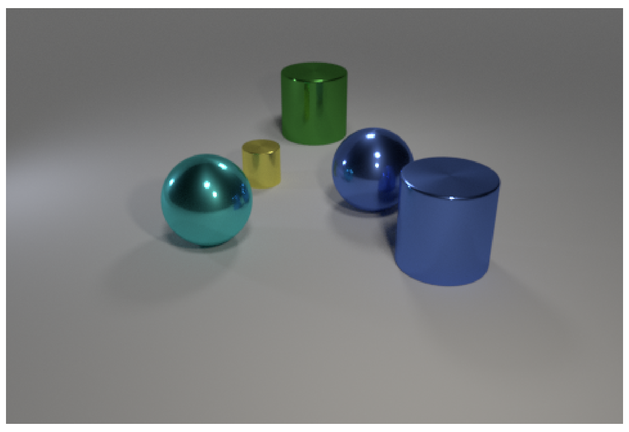
Dominik Künkele and **Simon Dobnik**
University of Gothenburg, Sweden
dominik.kuenkele@outlook.com and simon.dobnik@gu.se

AIMS

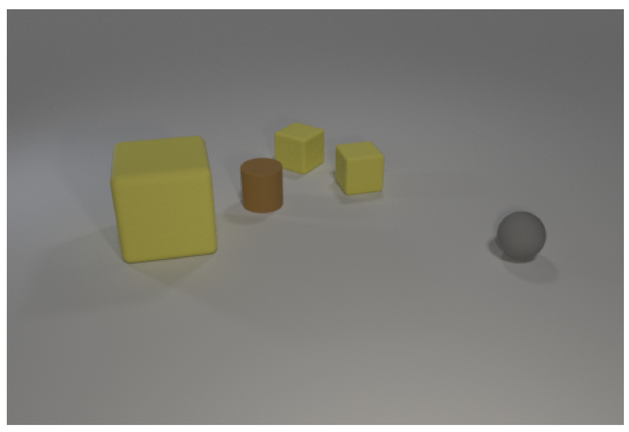
- Referential grounding in language games
- 3-d scenes of objects with different attributes and in different spatial configurations

DATASETS ‘DALE-5’ AND ‘DALE-2‘

Dale-5

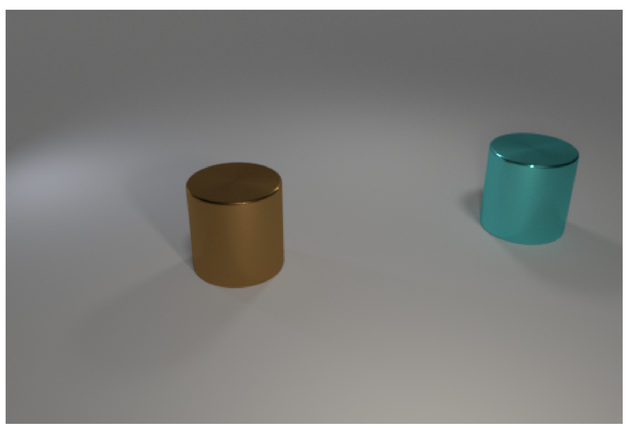


the *yellow cylinder*

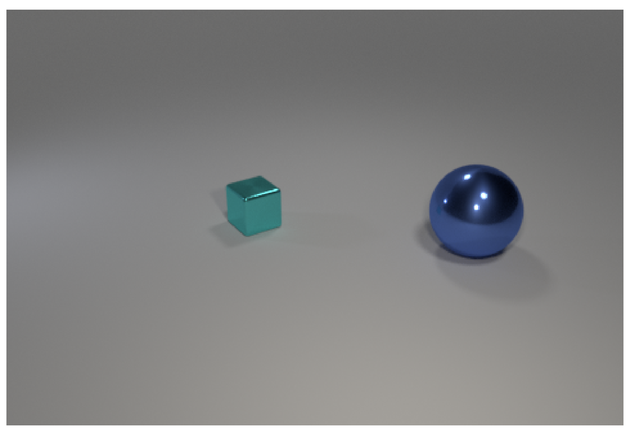


the *large yellow cube*

Dale-2



the *turquoise cylinder*

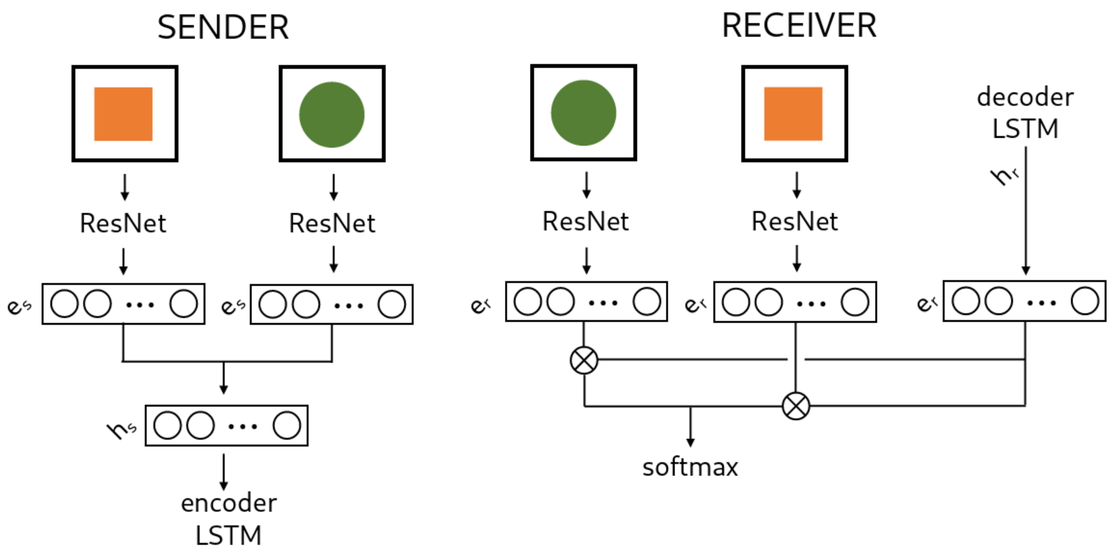


the *sphere*

- Two new datasets based on CLEVR (Johnson et al. (2017)
- Target object uniquely identifiable by 1-3 attributes (size, shape, colour)
- Objects in scenes are generated as anticipated by the GRE algorithm (Dale and Reiter, 1995) with ranking shape > colour > size
- Extract visual features of objects as bounding boxes of fixed size
- Agents negotiate on a language to refer to target objects

LANGUAGE GAMES

- Objects presented to sender and receiver in random order, except for the sender’s target
- Message is a sequence of symbols of vocabulary *V*
- Both sender and receiver are updated based on the receiver’s success to identify the target
- Loss: incremental sum of losses over sequences of symbols



RESULTS

Dataset	h_s	e_s	h_r	e_r	$\|V\|$	Acc.
Dale-2	10	10	10	10	10	95%
Dale-2	50	50	128	128	10	50%
Dale-5	10	10	10	10	10	23%
Dale-5	10	10	10	10	20	23%
Dale-5	10	10	10	10	100	41%

- |*V*| smaller than the set of attributes (3 + 8 + 2 = 13) enough for ‘Dale-2’
- |*V*| smaller than the possible set of unique objects (3 * 8 * 2 = 48) indicates learning of discriminative descriptions
- More symbols are required with more distractors