Referring Expression Generation: What Can We Learn from Human Data?

Jette Viethen and Robert Dale

MACQUARIE UNIVERSITY | FACULTY OF SCIENCE

The Message of this Talk

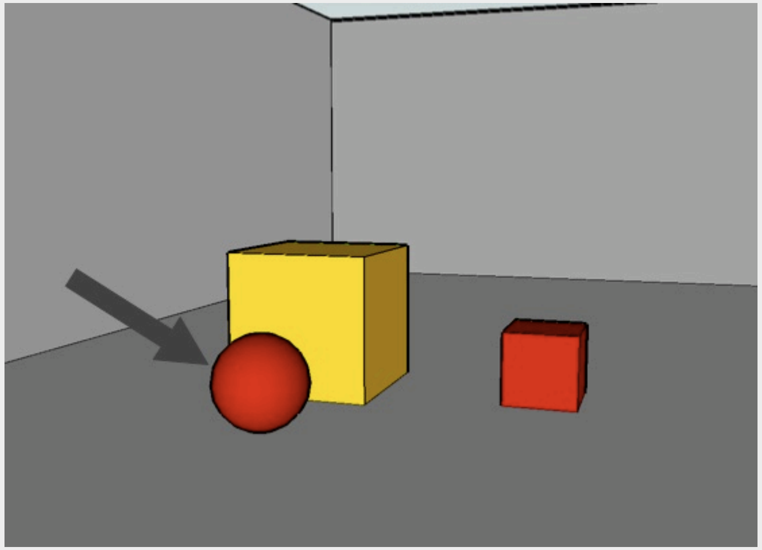
- We want to model human referring behaviour.
- Existing algorithms don’t explain the data.
- We suggest a new model as a starting point.

Overview

1. GRE3D3: a corpus of referring expressions
2. Existing algorithms and their problems
3. A new perspective on GRE
4. Open questions

Collecting the GRE3D3 Corpus

- Web-based production experiment
  - stimuli: simple 3D scenes
  - 3 objects: target, landmark, distractor
  - fully distinguishing descriptions encouraged
  - no time constraints
  - 10 stimuli pairs
  - 63 participants



The GRE3D3 Corpus

- 623 *distinguishing* descriptions
- 18 different *content patterns*

for example:

<tg_col, tg_type>	The blue cube
<tg_col, tg_type>, rel, lm_size, lm_col, lm_type>	The blue cube that's lying on top of the large red ball
<tg_type>, rel, lm_type>	The ball in front of the cube

Distribution of Patterns across Scenes

	Pattern	Scene # 1	2	3	4	5	6	7	8	9	10
A	tg_col, tg_type	17	24			36	32	26			40
B	tg_col, tg_type, rel, lm_col, lm_type	14	8	3		16	7	8	3		10
C	tg_col, tg_type, rel, lm_size, lm_col, lm_type		4		1			3		1	
D	tg_col, tg_type, rel, lm_size, lm_type						1	1		1	
E	tg_col, tg_type, rel, lm_type	4		1			2				
F	tg_size, tg_col, tg_type	2	1	15	44	5	3	2	25	40	8
G	tg_size, tg_col, tg_type, rel, lm_col, lm_type	1		14		2		1	14		1
H	tg_size, tg_col, tg_type, rel, lm_size, lm_col, lm_type		1	1	13	2	1	2	1	17	2
I	tg_size, tg_col, tg_type, rel, lm_size, lm_type				3					1	
J	tg_size, tg_col, tg_type, rel, lm_type			1			1				1
K	tg_size, tg_type			12					15		
L	tg_size, tg_type, rel, lm_size, lm_type				1						
M	tg_size, tg_type, rel, lm_type	1		7					4		
N	tg_type	11	13				14	14			
O	tg_type, rel, lm_col, lm_type		4					1			
P	tg_type, rel, lm_size, lm_col, lm_type							1			
Q	tg_type, rel, lm_size, lm_type		3					2			
R	tg_type, rel, lm_type	13	5	9			2	2	1		

Overview

1. GRE3D3: a corpus of referring expressions
2. Existing algorithms and their problems
3. A new perspective on GRE
4. Open questions

Current 'Serial Dependence' Model

Given:

- an intended referent R
- a set of distractors C
- a set of properties L_R
- the set of properties D to use in a description

let D = {}

repeat:

given C, select one property $p \in L_R$

add p to D

given D, recompute C

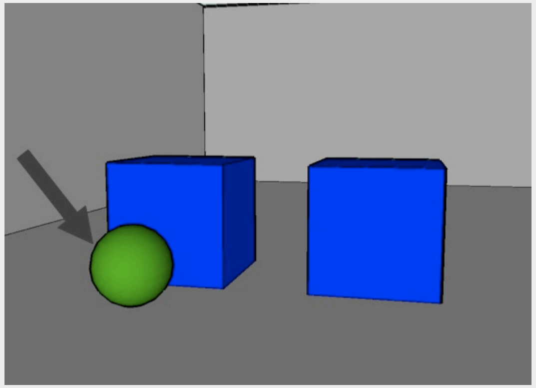
until C = {}

return D

What’s Wrong With That?

- Humans produce more redundancy than even the IA can.

the small green ball



- Unexplained variation:
  - Different people produce different descriptions in the same situation.
  - The same person produces different descriptions for similar situations.
- ‘Choose one then check’ seems computationally expensive and cognitively implausible.

Overview

1. GRE3D3: a corpus of referring expressions
2. Existing algorithms and their problems
3. A new perspective on GRE
4. Open questions

Learning Attribute-centric Heuristics

1. Identify relevant characteristics of scenes.
2. See if a decision tree learner can correlate these to the use of individual attributes.

Learned Item	Baseline (most frequent)	Using Scene Characteristics Only	Using Scene Characteristics and Participant
full content pattern	28.01%	**47.99%**	**57.62%**
tg_col inclusion	78.33%	78.33%	**89.57%**
tg_size inclusion	57.46%	**90.85%**	90.85%
rel inclusion	64.04%	65.00%	**81.22%**
lm_col inclusion	74.80%	**87.31%**	**93.74%**
lm_size inclusion	88.92%	**95.02%**	95.02%

Two Types of Reference

1. Reference ‘at a glance’:
   - Apply simple heuristics for immediately accessible properties.

The big white truck...

2. Deliberative reference:
   - Check whether the description satisfies the task.
   - Decide on a useful property to add next.

...in the left...

...with the massive bullbar.

A More General Model for GRE

Given:

- an intended referent R
- a set of task satisfaction criteria S
- a set of properties L_R
- the set of properties D to use in a description

let D = {}

repeat:

  select one or more properties $\in L_R$ via ‘reference at a glance’ or ‘deliberative reference’

  add them to D

until D satisfies S

return D

The Advantages of This Model

- ‘Reference at a glance’ is less computationally expensive than ‘deliberative reference’.
- Serial dependence is possible but not necessary:
  - Properties can be included independently of each other.
  - Deliberate one-by-one reference is still possible.
- ‘Distinguishingness’ is not the only satisfaction criterion.

Open Questions

- When is ‘reference at a glance’ preferred over ‘deliberative reference’ and vice versa?
- Which satisfaction criteria apply in which situations?
- How do we explain person-specific differences?