Speaker-Dependent Variation in Content Selection for Referring Expression Generation

Jette Viethen and Robert Dale

MACQUARIE UNIVERSITY

CENTRE FOR LANGUAGE TECHNOLOGY

ALTA 2010

Melbourne 10/12/2010

Outline

1. Background and Motivation
2. Two Corpora of Distinguishing Descriptions
3. Modelling Human Reference Behaviour
4. Summary and Outlook

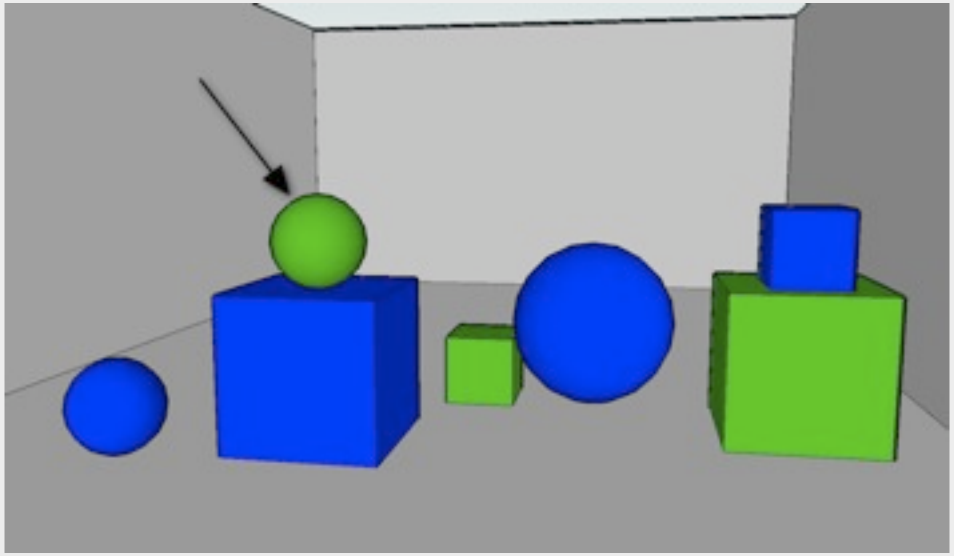
Outline

1. Background and Motivation
2. Two Corpora of Distinguishing Descriptions
3. Modelling Human Reference Behaviour
4. Summary and Outlook

Referring Expression Generation (REG)

... means automatically building distinguishing object descriptions

the small green ball



- ***Target Referent***: object to be described
- ***Distractors***: other objects in the environment that the target needs to be distinguished from
- ***Content Selection*** from the properties of the target and its relations to other objects (no linguistic realisation)

Different Goals of REG

Problem: There are always many different ways to describe a given object.

- The Minimalist Solution:
  always generate the shortest possible referring expression.
- The Application-Oriented Solution:
  generate one “good enough” referring expression.
- The Cognitive Science Solution:
  generate all referring expressions that we observe humans using.

The Goal of this Study

Observation: Not everyone describes the same thing
in the same way.

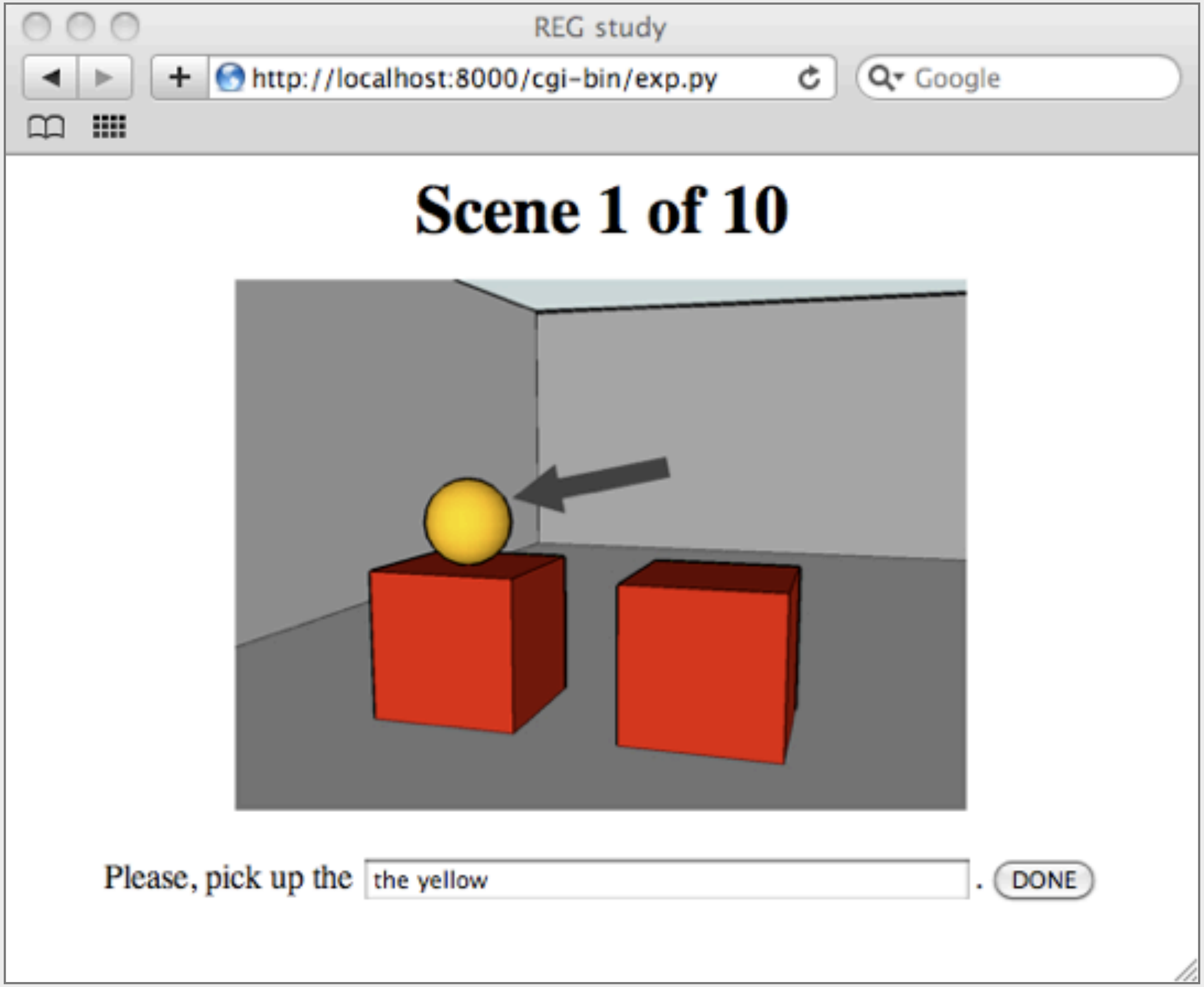
This has been ignored by existing approaches to REG.

→ How much does the content of a referring expression depend on personal preferences?

Outline

1. Background and Motivation
2. Two Corpora of Distinguishing Descriptions
3. Modelling Human Reference Behaviour
4. Summary and Outlook

Two Corpora: GRE3D3 and GRE3D7



Annotation of Semantic Content

- 9 Object Attributes
  - target type (`tg_type`)
  - target colour (`tg_col`)
  - target size (`tg_size`)
  - target location (`tg_loc`)
  - relation (`rel`)
  - landmark type (`lm_type`)
  - landmark colour (`lm_col`)
  - landmark size (`lm_size`)
  - landmark location (`lm_loc`)

- The attributes used in a description make up its **content pattern**:

<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

Some Numbers

	GRE3D3	GRE3D7
# original participants	74	318
# original descriptions	740	4689
# participants after filtering	63	280
# descriptions after filtering	630	4480
# different content patterns	32	27
# relational descriptions	224 (35.6%)	600 (13.4%)

Outline

1. Background and Motivation
2. Two Corpora of Distinguishing Descriptions
3. Modelling Human Reference Behaviour
4. Summary and Outlook

“Finding Patterns”

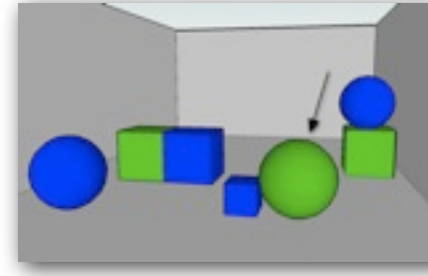
- Can we use machine learning to predict which content pattern people use in a given situation?

```
<tg_col, tg_type, rel, lm_size, lm_type>
                  OR
          <tg_col, tg_type>
                  OR
      <tg_size, tg_col, tg_type>
                  OR ...
```

1. based on characteristics of the scene only
2. based on the participant

- using C4.5 decision tree algorithm

Features



Feature	Explanation	Values	
TG_Size	size of the target (TG) object	small, large	direct
LM_Size	size of the landmark (LM) object	small, large	
Rel_Type	type of relation	horizontal, vertical	
Num_TG_Size	number of objects of same size as TG	numeric	comparative
Num_LM_Size	number of objects of same size as LM	numeric	
TG_LM_Same_Size	do TG and LM share size?	Boolean	
Num_TG_Col	number of objects of same colour as TG	numeric	
Num_LM_Col	number of objects of same colour as LM	numeric	
TG_LM_Same_Col	do TG and LM share colour?	Boolean	
Num_TG_Type	number of objects of same type as TG	numeric	
Num_LM_Type	number of objects of same type as LM	numeric	
TG_LM_Same_Type	do TG and LM share type?	Boolean	
Participant_ID	ID number of the description giver	alphanumeric	

Feature Recoding

Normalising Num_ features for overall number of objects:

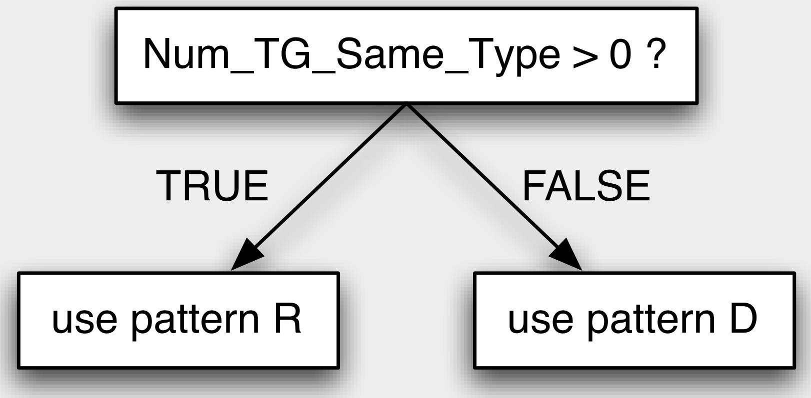
	number of objects sharing the property with the TG or LM				
GRE3D3	0		1		2
GRE3D7	0	1 2	3	4 5	6
general scale	none	few	half	most	all
joint coding	0	1	2	3	4

- Exact numbers are unlikely to be important
  - for visual salience.
  - for discriminatory power in our stimuli.
- Numerical values allow
  - elegant splits using inequalities.
  - cross-corpus testing.

Learning Content Patterns based on Scene Characteristics

training corpus	test method	accuracy of majority class baseline	accuracy of pruned decision tree
GRE3D3	10 fold X	(D) 27.3%	**46.5%**
	training set	(D) 27.3%	**46.5%**
	cross-corpus	(D) 36.7%	**47.9%**
GRE3D7	10 fold X	(R) 47.9%	**64.9%**
	training set	(R) 47.9%	**64.9%**
	cross-corpus	(R) 22.7%	**37.0%**

- Patterns predicted:
  - D: <tg_col, tg_type>
  - R: <tg_size, tg_col, tg_type>



- Decision trees beat the baseline.
- But much of the data remains unexplained.
- Training and testing on GRE3D7 achieves better results.

Learning Content Patterns based on Participant ID

training corpus	test method	+[scene features] –Participant_ID	–[scene features] +Participant_ID	+[scene features] +Participant_ID	
		pruned	n/a	pruned	unpruned
GRE3D3	training set	46.5%	41.9%	**91.3%**	**98.1%**
	10 fold X	46.5%	**31.1%**	54.4%	**57.6%**
GRE3D7	training set	64.9%	62.3%	**82.6%**	**93.8%**
	10 fold X	64.9%	57.1%	67.0%	63.7%

- Participant_ID as only feature does fairly well.
- The combination of [scene features] + Participant_ID accounts for most of the variation in the data.

Outline

1. Background and Motivation
2. Two Corpora of Distinguishing Descriptions
3. Modelling Human Reference Behaviour
4. Conclusions and Outlook

Summary and Conclusions

- Decision trees based on our scene features outperform a majority class baseline,

  but don't explain much of the data.
- Personal preferences have great impact on the content of referring expressions.
- Scene features combined with personal preferences explain most of the variation in the data.

⟶ Approaches to REG that aim to replicate human behaviour have to account for personal preferences.

Further Work

- Are all people really different?
  - Comparing individual models for each speaker
  - Automatic clustering of participants according to their referring behaviour
- Attribute-specific reference
  - Can more commonality be found at the level of individual attributes?
  - A new approach to REG: speaker profiles combine attribute-specific models to consider each attribute independently.

Thank You