Referring Expression Generation through Attribute-Based Heuristics

Robert Dale and Jette Viethen
rdale | jviethen@science.mq.edu.au

The Aim of This Talk

- To argue that existing approaches to referring expression generation are missing an important dimension:
  - attempting to reduce every decision to a uniform measure like ‘discriminatory power’ is inappropriate
  - the reasons for the inclusion of any given property are specific to what that property is

Outline

- Referring Expression Generation: The Current Paradigm
- A Look at Some Human-Produced Data
- Learning Heuristics
- Where Next?

The Referent Identification Task

- Given
  - an intended referent R
  - a contrast set C consisting of the potential distractor entities
  - knowledge of the properties of the entities
- ... find a set of properties true of R that, together, are not true of any entity in C.
- The result is a distinguishing description of R.

The General Form of Algorithms

Given an intended referent R, a set of distractors C, a set of properties L_R, and the set of properties D to use in a description:

let D = $\emptyset$
repeat
    add a selected property $\in L_R$ to D
    recompute C given D
until C = $\emptyset$

How Algorithms Differ: The Selection of Properties

- The Greedy Algorithm [Dale 1989]
  - Check all the properties, see which one has the greatest discriminatory power
- The Incremental Algorithm [Dale and Reiter 1995]
  - Take the next property from a predetermined list, provided it has some discriminatory value
- The Graph-Based Algorithm [Krahmer et al 2003]
  - Take the property that has the greatest weight

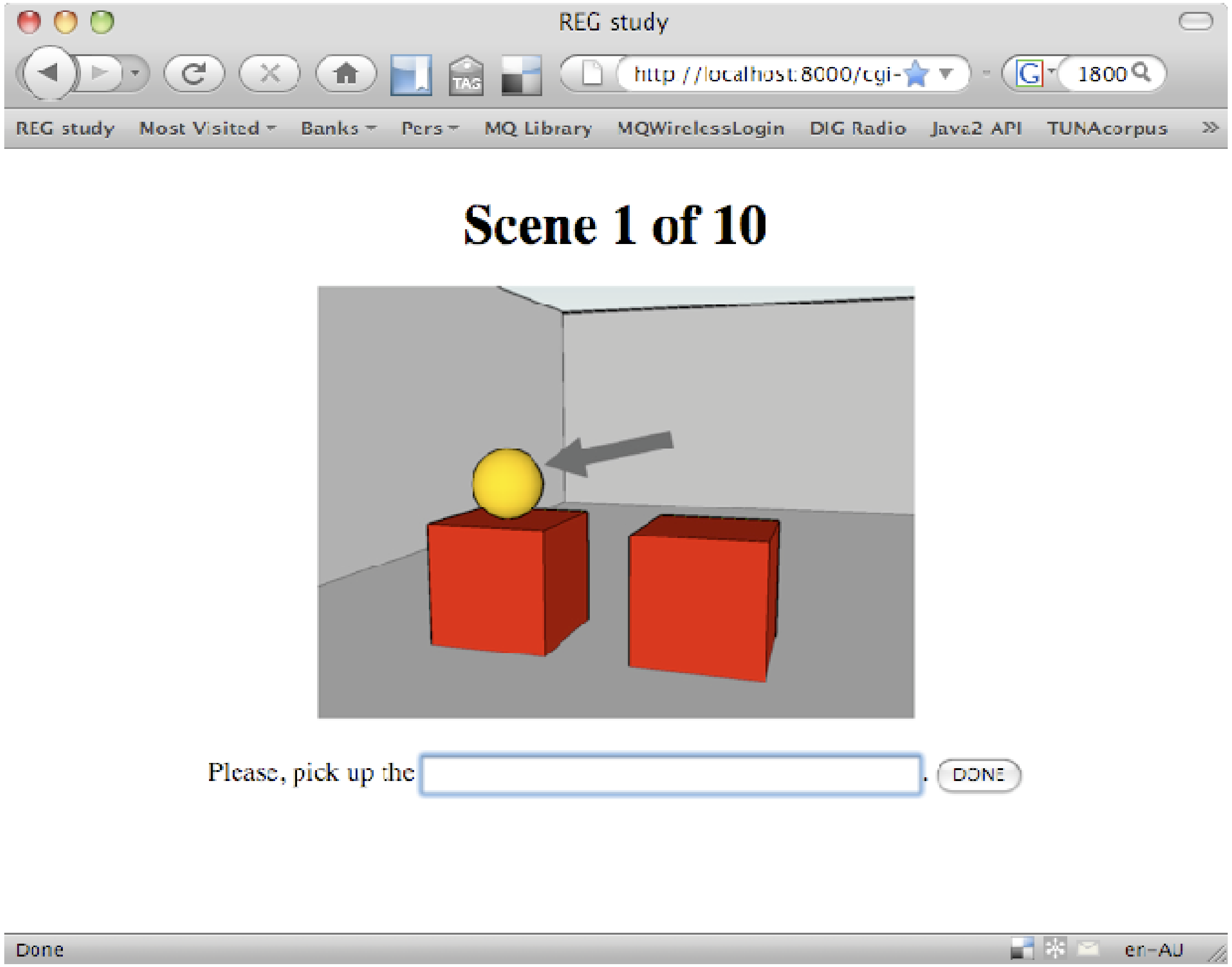
Why Is This Not a Good Model of What People Do?

1. People often produce redundant descriptions
2. People don’t always produce distinguishing descriptions
3. Different people do different things
4. The ‘add a property, check how we’re doing’ model seems too computationally expensive to be plausible

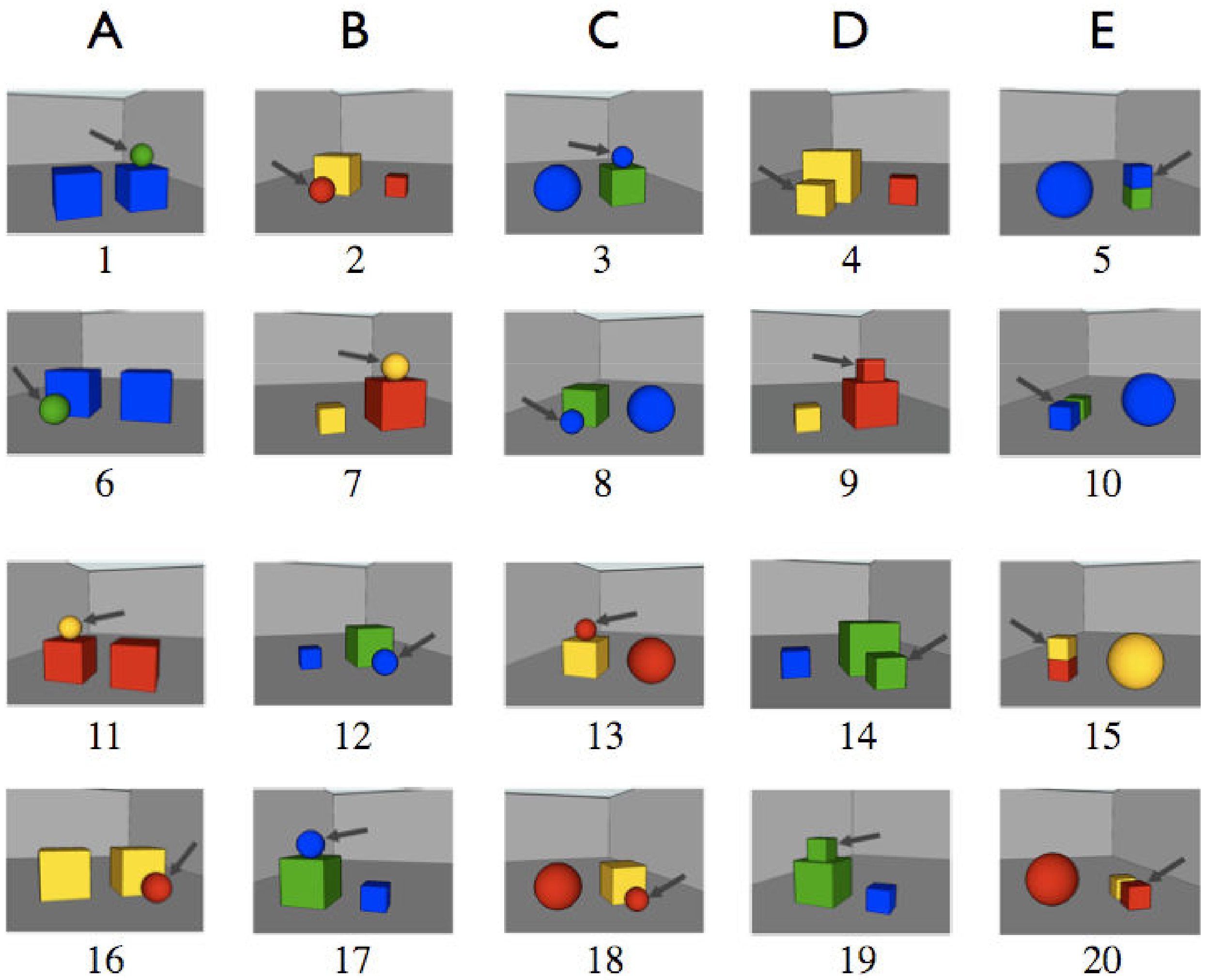
Outline

- Referring Expression Generation: The Current Paradigm
- A Look at Some Human-Produced Data
- Learning Heuristics
- Where Next?

The Experimental Setup



The Stimulus Scenes



Data Filtering and Normalisation

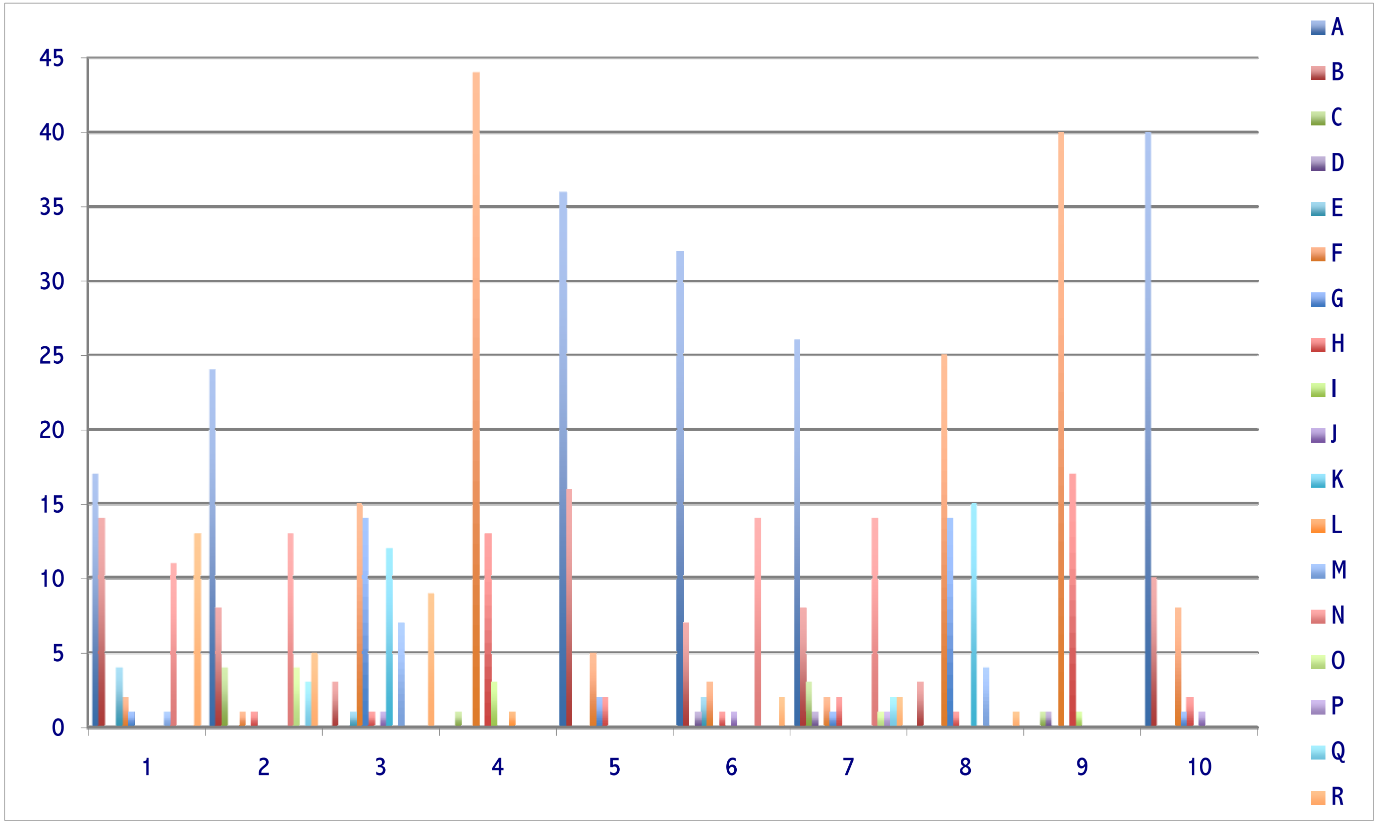
- 74 participants:
  - One asked for data to be discarded; one reported as being colour blind; one used very long referring expressions referring to the onlooker; eight participants only used type in their descriptions
- Normalisation:
  - Spelling mistakes corrected; colour names and head nouns normalised; complex syntactic structures simplified

→ 623 scene descriptions

Description Patterns

Label	Pattern	Example
A	⟨tg_col, tg_type⟩	*the blue cube*
B	⟨tg_col, tg_type, rel, lm_col, lm_type⟩	*the blue cube in front of the red ball*
C	⟨tg_col, tg_type, rel, lm_size, lm_col, lm_type⟩	*the blue cube in front of the large red ball*
D	⟨tg_col, tg_type, rel, lm_size, lm_type⟩	*the blue cube in front of the large ball*
E	⟨tg_col, tg_type, rel, lm_type⟩	*the blue cube in front of the ball*
F	⟨tg_size, tg_col, tg_type⟩	*the large blue cube*
G	⟨tg_size, tg_col, tg_type, rel, lm_col, lm_type⟩	*the large blue cube in front of the red ball*
H	⟨tg_size, tg_col, tg_type, rel, lm_size, lm_col, lm_type⟩	*the large blue cube in front of the large red ball*
I	⟨tg_size, tg_col, tg_type, rel, lm_size, lm_type⟩	*the large blue cube in front of the large ball*
J	⟨tg_size, tg_col, tg_type, rel, lm_type⟩	*the large blue cube in front of the ball*
K	⟨tg_size, tg_type⟩	*the large cube*
L	⟨tg_size, tg_type, rel, lm_size, lm_type⟩	*the large cube in front of the large ball*
M	⟨tg_size, tg_type, rel, lm_type⟩	*the large cube in front of the ball*
N	⟨tg_type⟩	*the cube*
O	⟨tg_type, rel, lm_col, lm_type⟩	*the cube in front of the red ball*
P	⟨tg_type, rel, lm_size, lm_col, lm_type⟩	*the cube in front of the large red ball*
Q	⟨tg_type, rel, lm_size, lm_type⟩	*the cube in front of the large ball*
R	⟨tg_type, rel, lm_type⟩	*the cube in front of the ball*

Distribution of Patterns Across Scenes



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		

Outline

- Referring Expression Generation: The Current Paradigm
- A Look at Some Human-Produced Data
- Learning Heuristics
- Where Next?

Can We Learn How to Refer?

1. Identify relevant characteristics of scenes
2. See if these can be correlated with patterns via a machine learner

Characteristics of Scenes

Label	Attribute	Values
tg_type = lm_type	Target and Landmark share Type	TRUE, FALSE
tg_type = dr_type	Target and Distractor share Type	TRUE, FALSE
lm_type = dr_type	Landmark and Distractor share Type	TRUE, FALSE
tg_col = lm_col	Target and Landmark share Colour	TRUE, FALSE
tg_col = dr_col	Target and Distractor share Colour	TRUE, FALSE
lm_col = dr_col	Landmark and Distractor share Colour	TRUE, FALSE
tg_size = lm_size	Target and Landmark share Size	TRUE, FALSE
tg_size = dr_size	Target and Distractor share Size	TRUE, FALSE
lm_size = dr_size	Landmark and Distractor share Size	TRUE, FALSE
rel	Relation between Target and Landmark	on top of, in front of

Results

- Weka J48 pruned decision tree classifier
- Predicts actual form of reference in 48% of cases under 10-fold cross validation
- The rule learned:

```
if tg_type = dr_type
    then use pattern F (⟨tg_size, tg_col, tg_type⟩)
    else use pattern A (⟨ tg_col, tg_type⟩)
endif
```

Distribution of Patterns Across Scenes

	Scene #									
Pattern	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		

Interim Conclusions

- We can learn a ‘correct answer’ for every scene
- We can’t explain the diversity in forms of reference

What About Speaker Difference?

- As well as the characteristics of scenes, add participant ID as a feature
- Description pattern prediction increases to 57.62%
- So: it may be possible to learn individual differences from the data

Learning the Presence or Absence of Individual Properties

Attribute to Include	Baseline (0-R)	Using Scene Characteristics	Using Scene Characteristics and Participant
Target Colour	78.33%	78.33%	**89.57%**
Target Size	57.46%	**90.85%**	90.85%
Relation	64.04%	65.00%	**81.22%**
Landmark Colour	74.80%	**87.31%**	**93.74%**
Landmark Size	88.92%	**95.02%**	95.02%

Heuristics for Colour Inclusion

- Always use colour [37 participants]
- If the target and the landmark are of the same type, use colour [all the rest]
- If the target and the landmark are not of the same type then:
  – Ignore colour [19 participants]
  – Use colour if target and distractor are the same size [4]
  – Use colour if target and distractor share size and the target is on top of the landmark [2]
  – Use colour if target and distractor share colour [1]

What Does This Mean?

- Everybody's different, but we often have some things in common:
  - Each 'speaker profile' consists of a collection of attribute-specific heuristics
  - Speaker profiles can vary significantly but be based on a set of commonly used attribute-specific heuristics
- The heuristics a particular speaker uses in a given situation may depend on a variety of contextual and personal-history factors

Speaker Profiles

#	tg_col	tg_size	tg_size	rel	lm_size
13	TgCol-T	TgSize-1	Rel-F	n/a	n/a
10	TgCol-T	TgSize-1	Rel-T	LmCol-T	LmSize-1
9	TgCol-1	TgSize-1	Rel-F	n/a	n/a
2	TgCol-3	TgSize-1	Rel-4	LmCol-F	LmSize-1
2	TgCol-T	TgSize-1	Rel-2	LmCol-T	LmSize-1
2	TgCol-1	TgSize-1	Rel-T	LmCol-1	LmSize-1

- TgCol-T = always include tg colour
- TgSize-1 = include tg size if tg and dr share type
- Rel-F = never use a relation

Outline

- Referring Expression Generation: The Current Paradigm
- A Look at Some Human-Produced Data
- Learning Heuristics
- Where Next?

Implications for Algorithm Development

- Each property is different: reduction to a single metric of value (such as discriminatory power) is too simplistic
- Properties may be included independently of other properties
- An alternative to the ‘add one then check’ model:
  - A ‘read off the scene’ model: gestalt analysis of a scene results in several properties being chosen in parallel

Is This The Whole Story?

- No. Sometimes we do reflect on the referring expression constructed so far, and add more:
  - Uhm, I'm gonna transfer to the phone on the table by the red chair . . . [points in the direction of the phone] the . . . the red chair, against the wall, uh the little table, with the lamp on it, the lamp that we moved from the corner? . . . the black phone, not the brown phone . . .

[Lucy from Twin Peaks]

New Questions

- What properties of a scene just ‘jump out’?
- How do we decide if the first cut is good enough?
- What kinds of reasoning are involved in determining what else is needed in a referring expression?

Conclusions

- Existing algorithms, based on a cycle of ‘add a carefully-considered property then check how we’re doing’, don’t acknowledge ‘bounded rationality’
- Hypothesis: different speakers use different heuristics for property inclusion in different circumstances, based on individual history and other factors
- Our investigative focus needs to shift to the question of what conditions the use of specific properties
- Could this be the end of discrimination?