

A pilot event for a  
GRE Shared Task Evaluation Campaign

# The Attribute Selection for GRE Challenge

# Overview

## 1. Introduction

### 1. Background to ASGRE Challenge

### 2. Generation of Referring Expressions

### 3. The Attribute Selection for GRE Challenge

- Data and Task
- Evaluation
- Organisation

## 2. System presentations by participants

## 3. Evaluation results

# Background to ASGRE Challenge

- UCNLG'05, Birmingham – tabled question on shared NLG task put to panel on data-driven NLG sparks off interest
- ENLG'05, Aberdeen – discussions, huge interest in topic
- INLG'06, Sydney – special session on sharing data and comparative evaluation; 4 papers, 6 open-mic presentations
- 2006 ACL SIGGEN board elections – nearly all position statements mention evaluation
- NSF Workshop on Shared Tasks and Comparative Evaluation in NLG (Apr 2007, Arlington, US) – 15 position papers, 20+ participants
  - break-out group on feasibility of shared task evaluation for GRE
- First NLG Shared Task Evaluation event at UCNLG+MT

# Generation of Referring Expressions

- GRE is a field with a long tradition, going back at least to work in the mid-1980s by Appelt, Grosz, Joshi, McDonald and others
- Field as it is today was shaped by Dale's work in late 1980s and early 1990s
- General GRE task: given a domain of entities and a target entity, create a referring expression for the target entity
- Additional requirements include that RE should be
  - distinguishing
  - minimal

# Generation of Referring Expressions

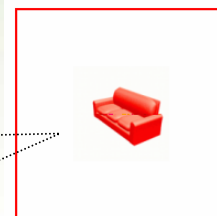
- GRE is a field with a long tradition, going back at least to work in the mid-1980s by Appelt, Grosz, Joshi, McDonald and others
- Field as it is today was shaped by Dale's work in late 1980s and early 1990s
- General GRE task: given a domain of entities and a target entity, create a referring expression for the target entity
- Additional requirements include that RE should be
  - distinguishing
  - minimal



# Generation of Referring Expressions

- GRE is a field with a long tradition, going back at least to work in the mid-1980s by Appelt, Grosz, Joshi, McDonald and others
- Field as it is today was shaped by Dale's work in late 1980s and early 1990s
- General GRE task: given a domain of entities and a target entity, create a referring expression for the target entity
- Additional requirements include that RE should be
  - distinguishing
  - minimal

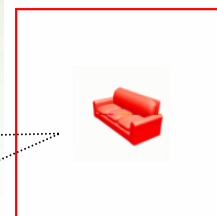
*the red sofa, it, the object,  
the red sofa facing left, the sofa,  
the object facing left, the red object,  
the object in the top left corner,  
etc.*



# Generation of Referring Expressions

- GRE is a field with a long tradition, going back at least to work in the mid-1980s by Appelt, Grosz, Joshi, McDonald and others
- Field as it is today was shaped by Dale's work in late 1980s and early 1990s
- General GRE task: given a domain of entities and a target entity, create a referring expression for the target entity
- Additional requirements include that RE should be
  - **distinguishing**
  - minimal

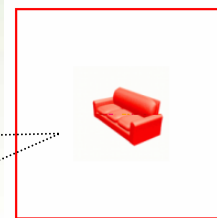
*the red sofa, it, the object,  
the red sofa facing left, the sofa,  
the object facing left, the red object,  
the object in the top left corner,  
etc.*



# Generation of Referring Expressions

- GRE is a field with a long tradition, going back at least to work in the mid-1980s by Appelt, Grosz, Joshi, McDonald and others
- Field as it is today was shaped by Dale's work in late 1980s and early 1990s
- General GRE task: given a domain of entities and a target entity, create a referring expression for the target entity
- Additional requirements include that RE should be
  - distinguishing
  - **minimal**

*the red sofa, it, the object,  
the red sofa facing left, **the sofa**,  
the object facing left, the red object,  
the object in the top left corner,  
etc.*



# Generation of Referring Expressions

- GRE research looks less at realisation than selection of semantic content, in particular selection of properties to describe referent
- Influential algorithms:
  - Full Brevity Algorithm (Dale, 1992)
  - Incremental Algorithm (Dale & Reiter, 1995)

# Attribute Selection for GRE Challenge

- Pilot event:
  - gauge community interest
  - start with small-scale low-risk pilot event
  - relaxed, collaborative atmosphere
  - if successful, grow into larger-scale, longer-term evaluation initiative (series of evaluation events)
- Choice of task for pilot event:
  - GRE: lively and well-defined NLG subfield
  - Sizeable research community working same clearly defined task: attribute selection; with similar inputs/outputs
  - availability of TUNA corpus designed for attribute selection task (van Deemter, Gatt, van der Sluis)

# ASGRE Challenge – Overview

- Data:
  - corpus of paired inputs and outputs derived from TUNA corpus
  - divided into training, development and test data
- Task:
  - implement attribute selection method that maps inputs to outputs
  - use training and development data to develop methods
- Participation requirements:
  - submission of report describing method and giving evaluation results for development data
  - after report submission, download test set inputs and submit outputs within 1 week
- Evaluation:
  - participants perform evaluation on development set
  - organisers perform evaluation for test set outputs

# ASGRE Challenge – Data

## Corpus data:

- Input: sets of attributes for domain entities (target referent and distractors)
- Output: set of attributes for target referent derived from human-produced descriptions of target referent (TUNA elicitation experiment)
- Two subdomains: people and furniture

# ASGRE Challenge – Data

This is scenario 4 of 38



Which object is in a red box?

submit

# ASGRE Challenge – Data

This is scenario 1 of 38



Which object is in a red box?

submit

# ASGRE Challenge – Data

```
<TRIAL ID="s101t4">
<DOMAIN>
<ENTITY ID="121" TYPE="target">
  <ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
  <ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
  <ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
  <ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
  <ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
  <ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" /> </ENTITY>
<ENTITY ID="122" TYPE="distractor"> ... </ENTITY>
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" /> </ENTITY>
<ENTITY ID="57" TYPE="distractor"> ... </ENTITY>
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" /> </ENTITY>
<ENTITY ID="52" TYPE="distractor"> ... </ENTITY>
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" /> </ENTITY>
<ENTITY ID="106" TYPE="distractor"> ... </ENTITY>
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" /> </ENTITY>
<ENTITY ID="41" TYPE="distractor"> ... </ENTITY>
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" /> </ENTITY>
<ENTITY ID="36" TYPE="distractor"> ... </ENTITY>
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" /> </ENTITY>
</DOMAIN>
<DESCRIPTION>
  <ATTRIBUTE ID="a150" NAME="colour" VALUE="blue" />
  <ATTRIBUTE ID="a151" NAME="size" VALUE="small" />
  <ATTRIBUTE ID="a152" NAME="type" VALUE="fan" />
</DESCRIPTION>
</TRIAL>
```

# ASGRE Challenge – Data

```
<TRIAL ID="s101t4">  
<DOMAIN>
```

```
<ENTITY ID="121" TYPE="target">  
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue">  
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="up">  
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small">  
<ATTRIBUTE NAME="dimension" TYPE="gradable" VALUE="3">
```

```
</ENTITY>  
<ENTITY ID="122" TYPE="distractor">  
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue">  
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="up">  
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small">  
<ATTRIBUTE NAME="dimension" TYPE="gradable" VALUE="3">
```

```
</ENTITY>  
<ENTITY ID="57" TYPE="distractor">  
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue">  
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="up">  
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small">  
<ATTRIBUTE NAME="dimension" TYPE="gradable" VALUE="3">
```

```
</ENTITY>  
<ENTITY ID="52" TYPE="distractor">  
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue">  
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="up">  
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small">  
<ATTRIBUTE NAME="dimension" TYPE="gradable" VALUE="3">
```

```
</ENTITY>  
<ENTITY ID="106" TYPE="distractor">  
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue">  
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="up">  
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small">  
<ATTRIBUTE NAME="dimension" TYPE="gradable" VALUE="3">
```

```
</ENTITY>  
<ENTITY ID="41" TYPE="distractor">  
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue">  
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="up">  
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small">  
<ATTRIBUTE NAME="dimension" TYPE="gradable" VALUE="3">
```

```
</ENTITY>  
<ENTITY ID="36" TYPE="distractor">  
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue">  
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="up">  
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small">  
<ATTRIBUTE NAME="dimension" TYPE="gradable" VALUE="3">
```

```
</ENTITY>  
</DOMAIN>  
<DESCRIPTION>  
...  
</DESCRIPTION>
```

```
</TRIAL>
```

```
<TRIAL ID="s101t4">
```

```
<DOMAIN>
```

```
<ENTITY ID="121" TYPE="target">...</ENTITY>
```

```
<ENTITY ID="122" TYPE="distractor">...</ENTITY>
```

```
<ENTITY ID="57" TYPE="distractor">...</ENTITY>
```

```
<ENTITY ID="52" TYPE="distractor">...</ENTITY>
```

```
<ENTITY ID="106" TYPE="distractor">...</ENTITY>
```

```
<ENTITY ID="41" TYPE="distractor">...</ENTITY>
```

```
<ENTITY ID="36" TYPE="distractor">...</ENTITY>
```

```
</DOMAIN>
```

```
<DESCRIPTION>
```

```
...
```

```
</DESCRIPTION>
```

```
</TRIAL>
```

# ASGRE Challenge – Data

```
<TRIAL ID="s101t4">
<DOMAIN>
<ENTITY ID="121" TYPE="target">
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" />
</ENTITY>
<ENTITY ID="122" TYPE="distractor"> ... </ENTITY>
<ENTITY ID="57" TYPE="distractor"> ... </ENTITY>
<ENTITY ID="52" TYPE="distractor"> ... </ENTITY>
<ENTITY ID="106" TYPE="distractor"> ... </ENTITY>
<ENTITY ID="41" TYPE="distractor"> ... </ENTITY>
<ENTITY ID="36" TYPE="distractor"> ... </ENTITY>
</DOMAIN>
<DESCRIPTION>
...
</DESCRIPTION>
</TRIAL>
```

```
<TRIAL ID="s101t4">
```

```
<DOMAIN>
```

```
<ENTITY ID="121" TYPE="target">
```

```
<ENTITY ID="122" TYPE="distractor"> ... </ENTITY>
```

```
<ENTITY ID="57" TYPE="distractor"> ... </ENTITY>
```

```
<ENTITY ID="52" TYPE="distractor"> ... </ENTITY>
```

```
<ENTITY ID="106" TYPE="distractor"> ... </ENTITY>
```

```
<ENTITY ID="41" TYPE="distractor"> ... </ENTITY>
```

```
<ENTITY ID="36" TYPE="distractor"> ... </ENTITY>
```

```
</DOMAIN>
```

```
<DESCRIPTION>
```

```
...
```

```
</DESCRIPTION>
```

```
</TRIAL>
```

```
<ENTITY ID="121" TYPE="target">
```

```
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
```

```
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
```

```
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
```

```
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
```

```
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
```

```
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" />
```

```
</ENTITY>
```

# ASGRE Challenge – Data

```
<TRIAL ID="s101t4">
<DOMAIN>
<ENTITY ID="121" TYPE="target">
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" />
</ENTITY>
<ENTITY ID="41" TYPE="distractor"> ... </ENTITY>
</DOMAIN>
<DESCRIPTION>
<ATTRIBUTE ID="a150" NAME="colour" VALUE="blue" />
<ATTRIBUTE ID="a151" NAME="size" VALUE="small" />
<ATTRIBUTE ID="a152" NAME="type" VALUE="fan" />
</DESCRIPTION>
</TRIAL>
```

```
<TRIAL ID="s101t4">
```

```
<DOMAIN>
```

```
<ENTITY ID="121" TYPE="target">
```

```
<ENTITY ID="122" TYPE="target">
```

```
<ENTITY ID="57" TYPE="target">
```

```
<ENTITY ID="52" TYPE="target">
```

```
<ENTITY ID="106" TYPE="target">
```

```
<ENTITY ID="41" TYPE="distractor">
```

```
<ENTITY ID="36" TYPE="distractor">
```

```
</DOMAIN>
```

```
<DESCRIPTION>
```

```
...
```

```
</DESCRIPTION>
```

```
</TRIAL>
```

```
<ENTITY ID="121" TYPE="target">
```

```
<ATTRIBUTE NAME="colour" TYPE="literal" VALUE="blue" />
```

```
<ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left" />
```

```
<ATTRIBUTE NAME="type" TYPE="literal" VALUE="fan" />
```

```
<ATTRIBUTE NAME="size" TYPE="literal" VALUE="small" />
```

```
<ATTRIBUTE NAME="x-dimension" TYPE="gradable" VALUE="1" />
```

```
<ATTRIBUTE NAME="y-dimension" TYPE="gradable" VALUE="3" />
```

```
</ENTITY>
```

```
<ENTITY ID="41" TYPE="distractor"> ... </ENTITY>
```

```
<DESCRIPTION>
```

```
<ATTRIBUTE ID="a150" NAME="colour" VALUE="blue" />
```

```
<ATTRIBUTE ID="a151" NAME="size" VALUE="small" />
```

```
<ATTRIBUTE ID="a152" NAME="type" VALUE="fan" />
```

```
</DESCRIPTION>
```

# ASGRE Challenge – Task

- Implement an attribute selection method which maps input domains to attribute sets for the target referent
- Up to participants what the aim is:
  - “humanlikeness”
  - unique identification
  - minimality
  - something else?
- Implies several evaluation methods

# Evaluation criteria

1. *Uniqueness*: does the attribute set uniquely describe the target referent?
2. *Minimality*: is the attribute set one of the minimal attribute sets that describes the target referent?
3. *Human-like*: is the attribute set similar to the attribute sets in the corpus?
4. *Identification Accuracy*: does the attribute set enable subjects to identify the target referent correctly?
5. *Identification Speed*: does the attribute set enable subjects to identify a referent quickly?

# Organisation

- Based on other NLP shared-task evaluations, SEMEVAL and CoNLL in particular
- Novel aspects:
  - Participants' reports submitted *before* test data released
  - Participants perform part of the evaluation themselves
  - Use of several evaluation metrics
  - Submission of additional evaluation metrics invited

# Participation

- 19 registrations; 13 researchers formed 6 teams and submitted 22 systems by deadline
- Submitting teams:
  - CAM: Computer Lab, Cambridge University, UK
  - DIT: Dublin Institute of Technology, Ireland
  - GRAPH: Universities of Twente and Tilburg, NL, and Macquarie University, Australia
  - IS: University of Stuttgart, Germany
  - NIL: Universidad Complutense de Madrid, Spain
  - TITCH: Tokyo Institute of Technology, Japan