The Impact of Visual Context on the Content of Referring Expressions

Jette Viethen, Robert Dale and Markus Guhe

Our Question

Does the visual context matter
for reference in discourse?

The Answer

Visual context seems to matter less than previously thought.

Outline

1. Referring Expression Generation (in Dialogue)
2. The iMAP Corpus
3. Modelling human reference behaviour using ML
4. Investigating the size of the visual context
5. Results and Conclusions

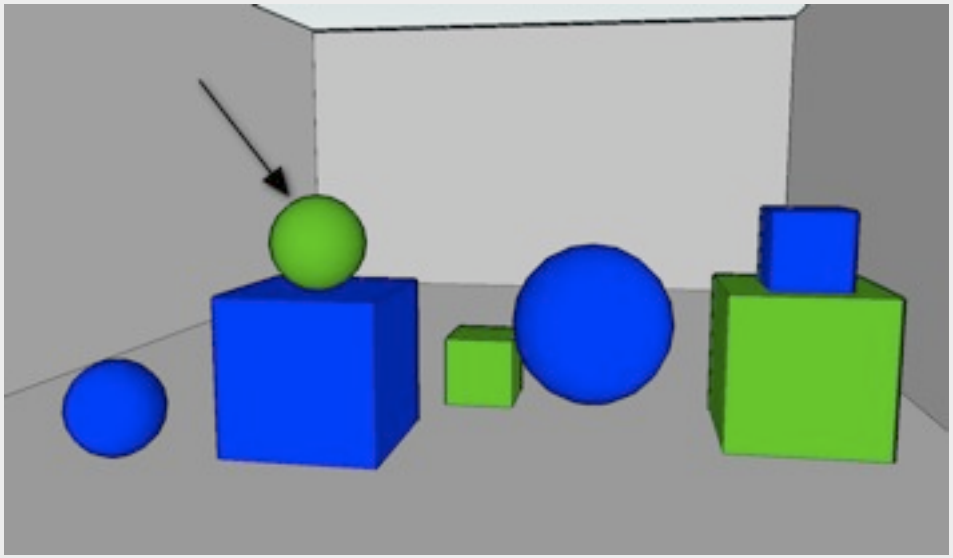
Referring Expression Generation (REG)

the green ball

the small green ball

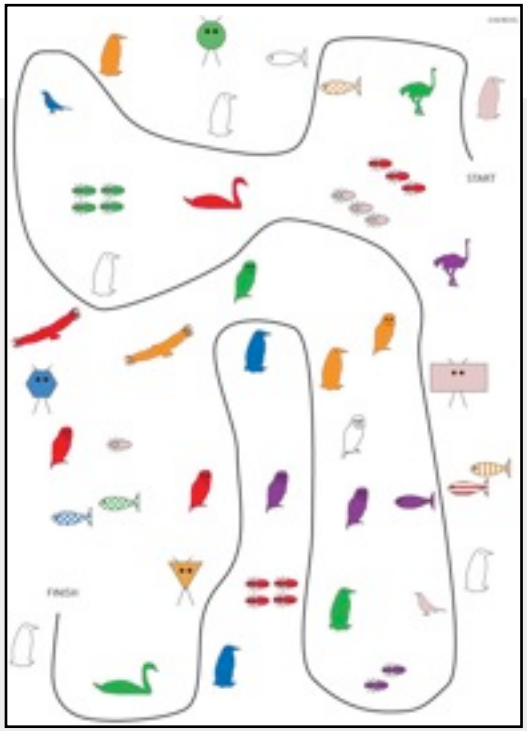
the green one

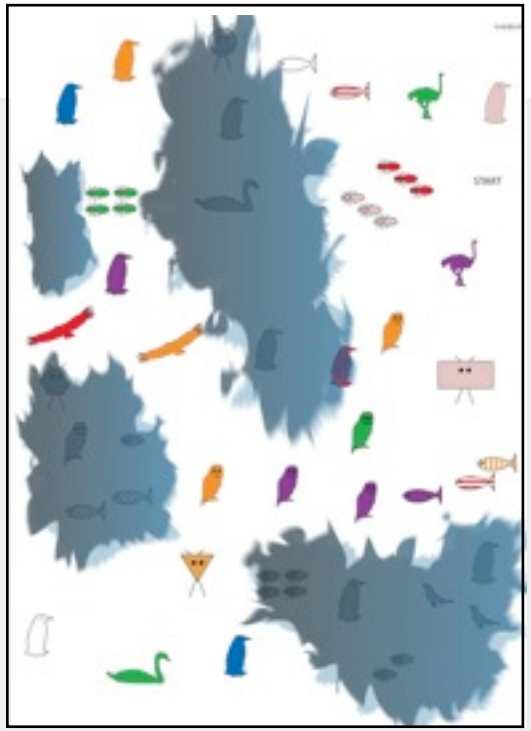
it



- ***Target Referent***: object to be described
- **(Visual) *Context***: set other objects (in the visual environment) that the target needs to be distinguished from
- **Successful reference**: a distinguishing description
- ***Content Selection*** from the attributes of the target and its relations to other objects (no linguistic realisation)

REG in Dialogue





- Objects get referred to repeatedly.
- The dialogue context (also) impacts on references.

Traditional REG approach:

- use mentions of other entities as distractors
- then compute a distinguishing description

Psycholinguistic findings:

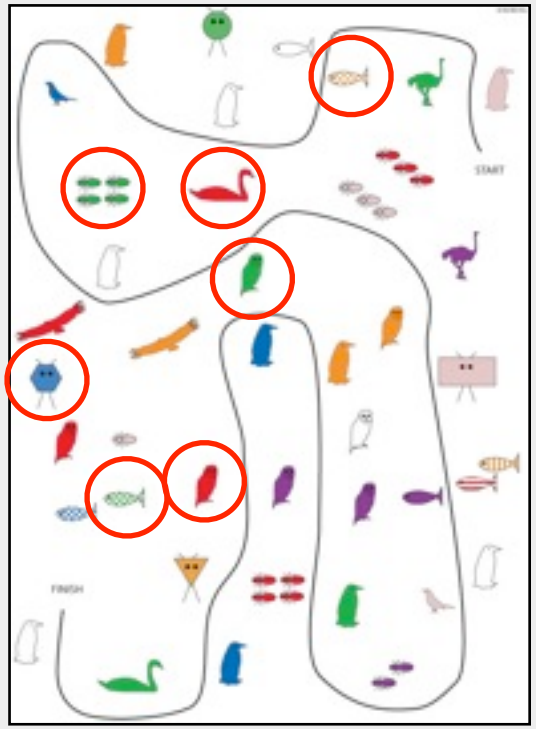
- semantic alignment of references
- conceptual pacts

Outline

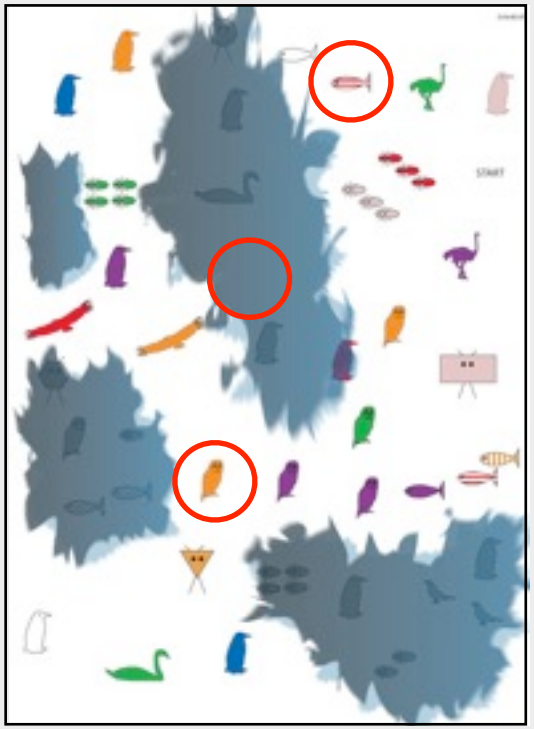
1. Referring Expression Generation (in Dialogue)
2. The iMAP Corpus
3. Modelling human reference behaviour using ML
4. Investigating the size of the visual context
5. Results and Conclusions

Referring Expressions in the iMAP Corpus

- landmarks distinguishable by
  - type
  - colour
  - one “other” attribute
- 4 map types



Instruction Giver



Instruction Follower

- 256 dialogues
- 34,126 REs in the corpus
- we exclude those that
  - use a spatial relation;
  - refer to more than one landmark;
  - don’t use any of the three ‘standard’ attributes.

⟶ 22,727 REs (6,369 initial; 16,358 subsequent)

Content Patterns

Go beneath the three pink bugs.

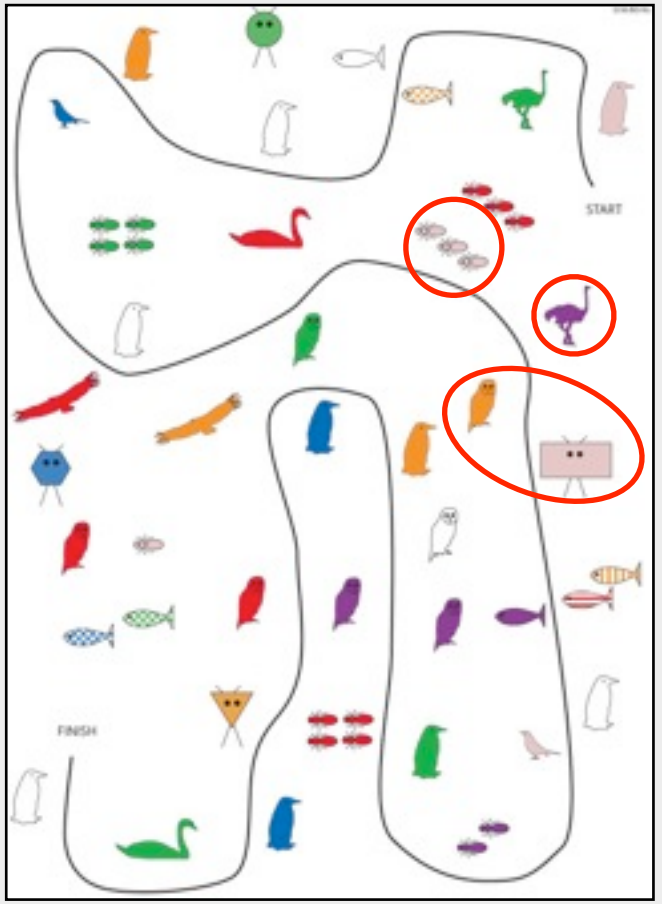
<other, col, type>

Turn towards the bottom just before the purple ostrich...

<other, col>

... between the owl and the alien.

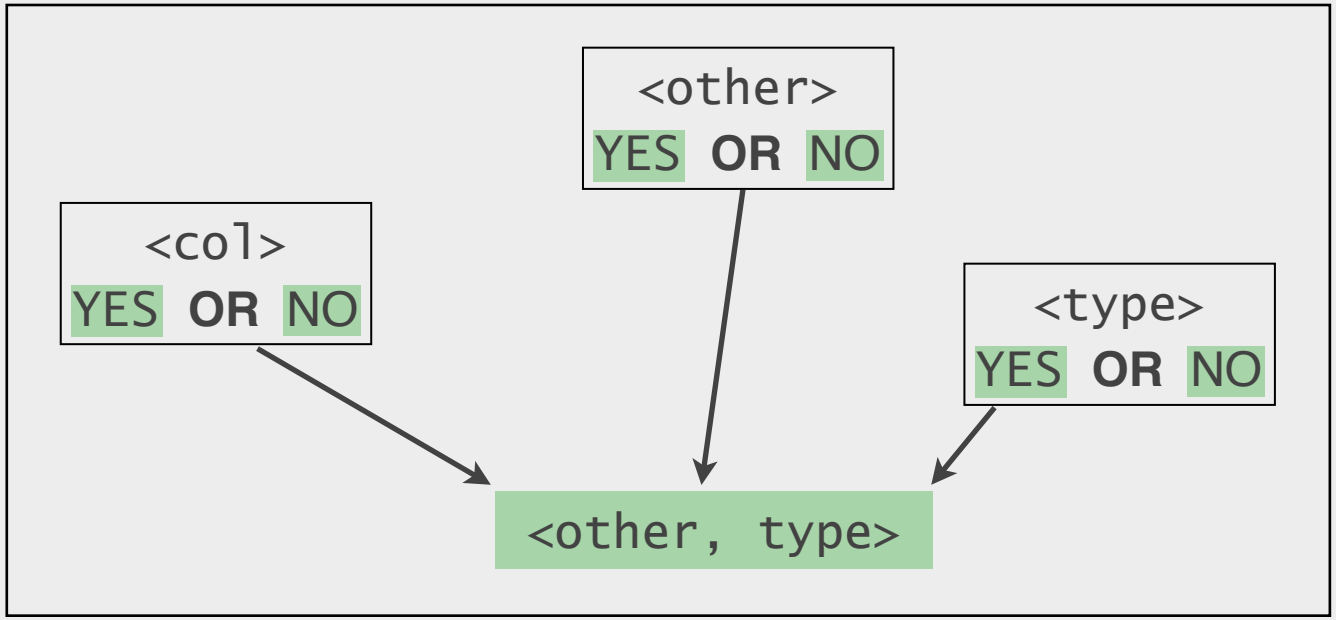
<other> <type>



Outline

1. Referring Expression Generation (in Dialogue)
2. The iMAP Corpus
3. Modelling human reference behaviour using ML
4. Investigating the size of the visual context
5. Results and Conclusions

Learning which Content Pattern to Use



- all attributes in parallel
- C4.5 decision trees
- 70–30 training–test set split

Features

- Traditional REG:
  - Count_Vis_Distractors, Prop_Vis_Same_*Att*
  - Distance_Closest, Closest_Same_*Att* ...

  Visual TradREG
  - Count_Disc_Distractors, Prop_Disc_Same_*Att*

  Discourse TradREG
- Alignment:
  - Last_Mention_*Att*, Distance_Last_Mention, Distance_Last_*Att*
  - Count_*Att*_Used, Quartile, Mention_No, Dialogue_No ...
- Independent:
  - Map_type, Ink_Orderliness, Mixedness
  - Target: other_Att, *Att*_Value, *Att*_Difference, Missing, Inked_Out
  - Dyad_ID, Speaker_ID, Speaker_Role

TradREG Features Don't Matter...

- Accuracy with which our system replicates the references in the corpus.

Features	Initial	Subsequent	All References
AllF	68.6%	58.8%	61.5%
AllF-TradREG	69.4%	58.2%	61.3%
AllF-Discourse TradREG	68.6%	58.4%	61.3%
AllF-Visual TradREG	69.4%	58.5%	61.6%

Outline

1. Referring Expression Generation (in Dialogue)
2. The iMAP Corpus
3. Modelling human reference behaviour using ML
4. Investigating the size of the visual context
5. Results and Conclusions

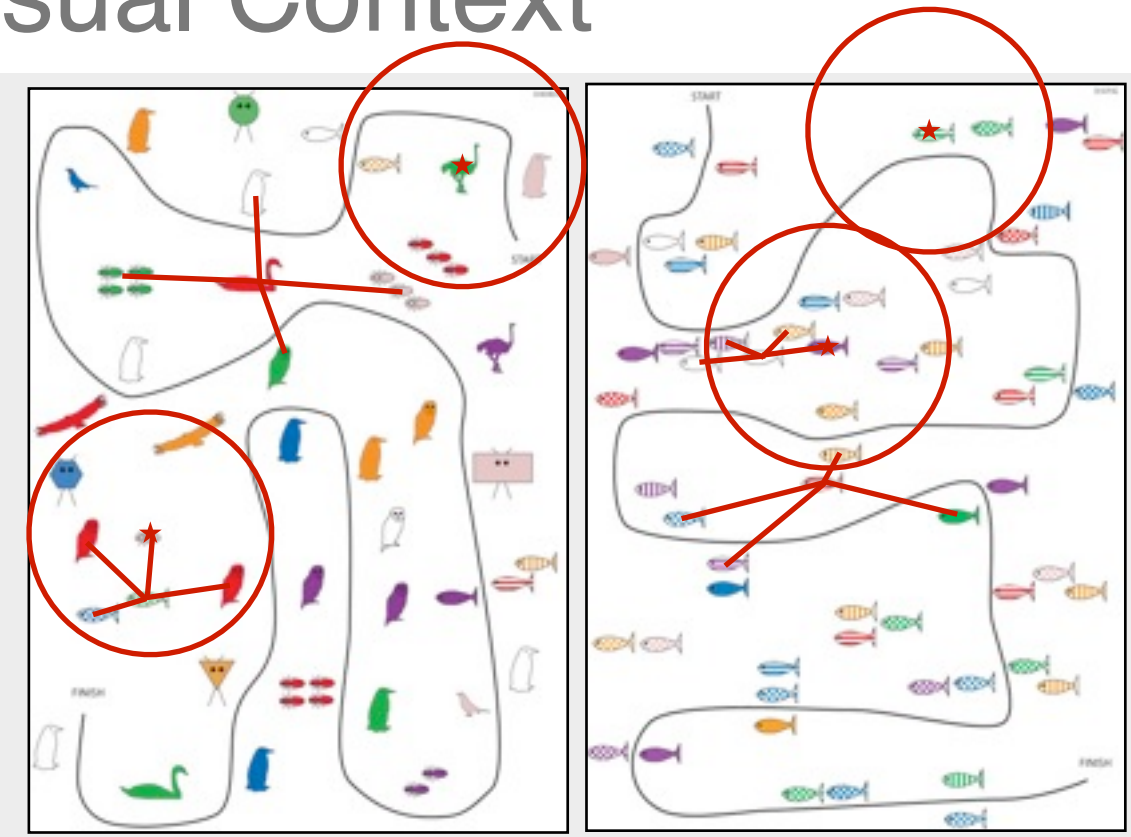
Features

- Traditional REG:
  - Count_Vis_Distractors, Prop_Vis_Same_*Att*
  - Distance_Closest, Closest_Same_*Att* ...
  - Count_Disc_Distractors, Prop_Disc_Same_*Att*
- Alignment:
  - Last_Mention_*Att*, Distance_Last_Mention, Distance_Last_*Att*
  - Count_*Att*_Used, Quartile, Mention_No, Dialogue_No ...
- Independent:
  - Map_type, Ink_Orderliness, Mixedness
  - Target: other_Att, *Att*_Value, *Att*_Difference, Missing, Inked_Out
  - Dyad_ID, Speaker_ID, Speaker_Role

Determining the Visual Context

original method:

average-6: a set radius with an average of 6 distractors



1. **count**: a set number of distractors per landmark
   (we try all counts between 0 – 61)
2. **distance**: a set radius around each landmark
   (we try all distances between 0 – 675 px in 15 px steps)

Outline

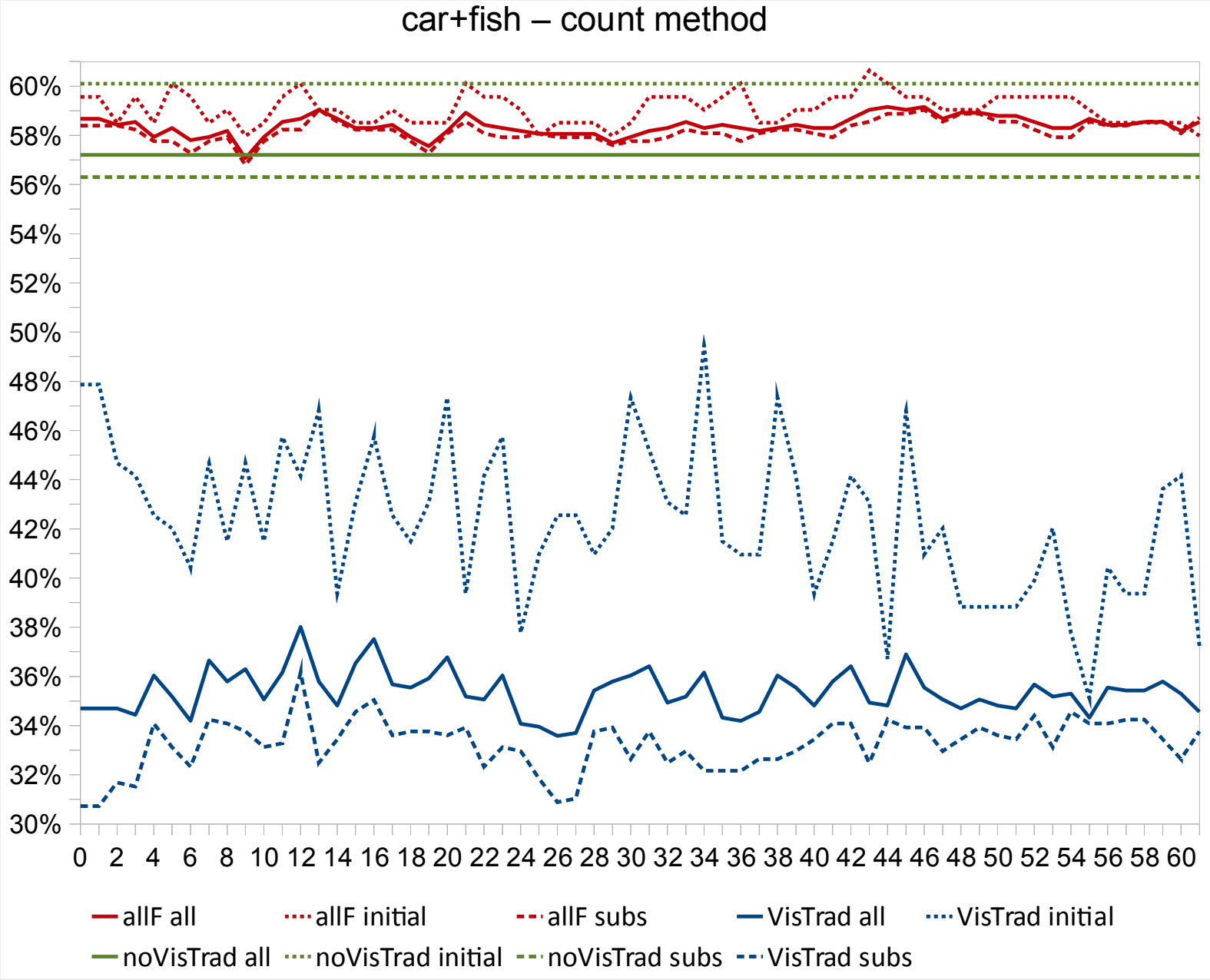
1. Referring Expression Generation (in Dialogue)
2. The iMAP Corpus
3. Modelling human reference behaviour using ML
4. Investigating the size of the visual context
5. Results and Conclusions

Choosing the Best Distractor Count

- using all features

Map type	Initial		Subsequent		All References	
	best counts	Acc	best counts	Acc	best counts	Acc
alien+sign	5	68.3%	43	62.5%	43	63.5%
fish+car	43	60.6%	13	59%	44, 46	59.2%
house+bird	22	75.6%	13, 19, 28	71.8%	3, 22	72.6%
trees+bugs	0, 1, 3, 11, 12	74.8%	33	68.4%	3	70.5%
weighted average		**71.1%**		**65.9%**		**67.1%**
average-6		**68.6%**		**58.8%**		**61.5%**

Comparing All Distractor Counts



Conclusions

- We can improve ML results by using the best visual context size.
- We can’t systematically determine the best visual context size.
  - Maybe our model is too coarse.
    - The path might play a role.
    - The discourse context might shape the visual context.

      go left until you get to the red alien
  - Maybe the map scenario is too simple.