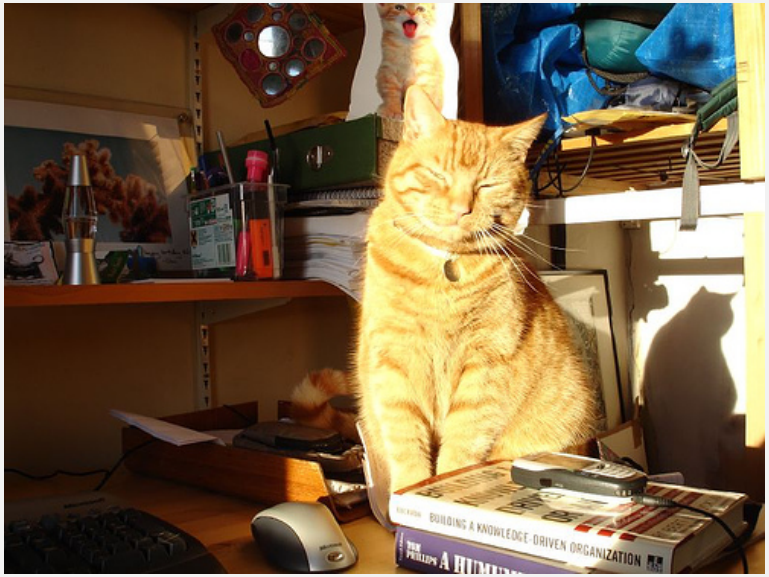
ShapeWorld for automatic language generation in a closed-world domain

Alexander Kuhnle

Computer Laboratory
University of Cambridge

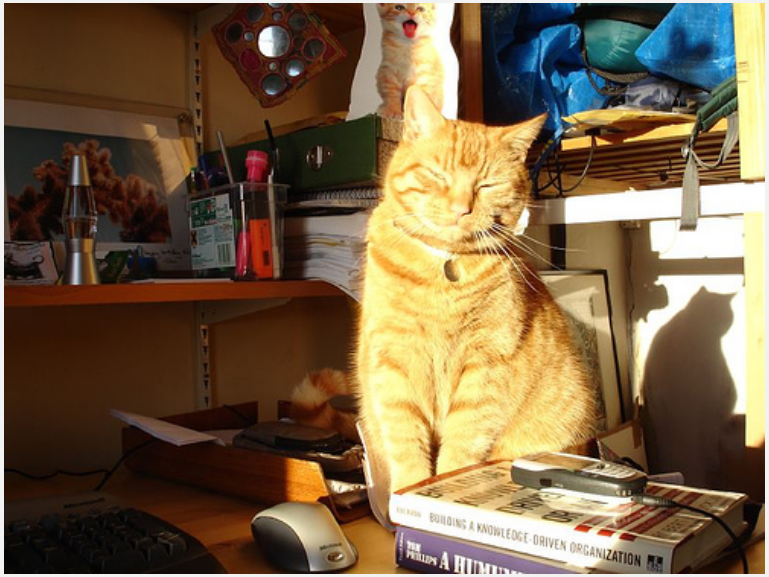
DELPH-IN Summit, 2017

Visual question answering

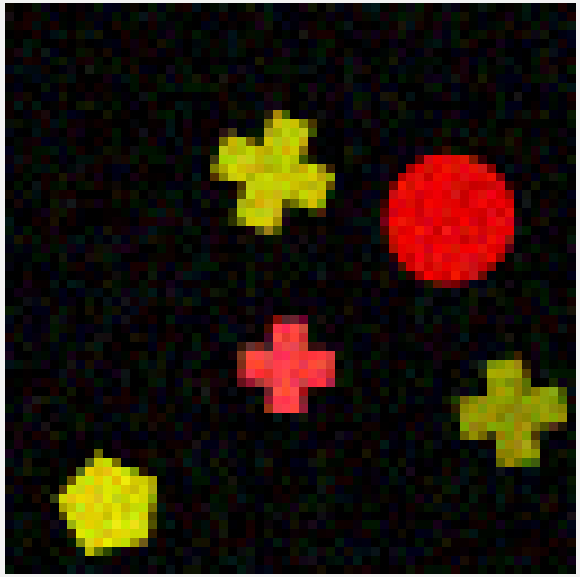


- What object is shining on the animal?
- What objects is the cat sitting behind?
- How many cats?

Visual question answering

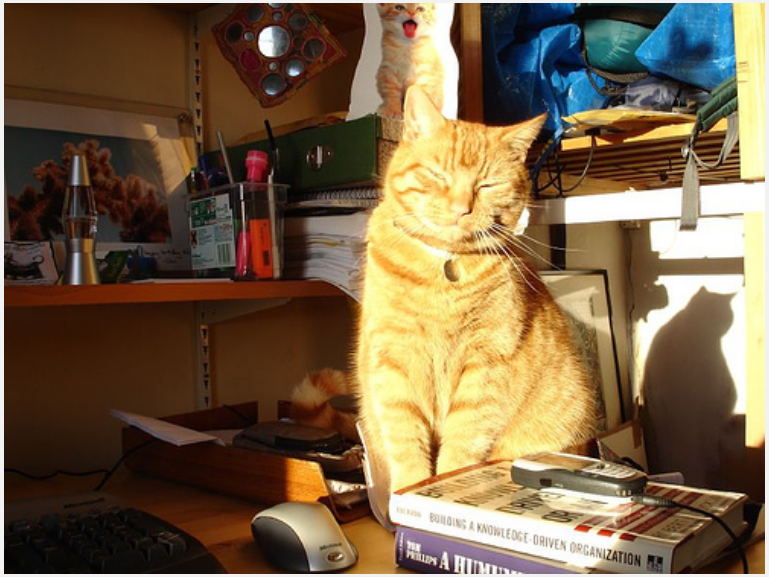


- What object is shining on the animal?
- What objects is the cat sitting behind?
- How many cats?

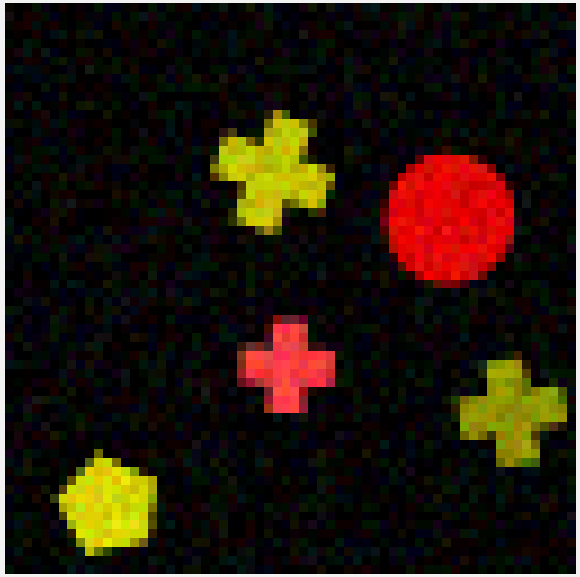


- Is there a yellow circle?
- Are most crosses yellow?
- Are all red crosses to the right of a pentagon?

Visual question answering



- What object is shining on the animal?
- What objects is the cat sitting behind?
- How many cats?

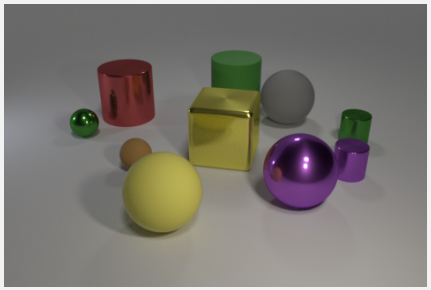


- Are most crosses yellow?
- Ist dem Kreis ein Quadrat am nächsten?
- Czy dwa żółte krzyżyki sa obok siebie?
- 有没有绿色的圆圈?
- etc...

Outline

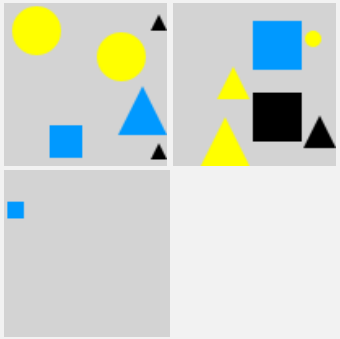
1. Background
2. SIG ‘hackathon’
3. Discussion points
4. More details on generation system

Background



CLEVR

- The yellow sphere has what size?
- How many things are either brown balls or objects in front of the red cylinder?
- What is the material of the cylinder behind the cylinder left of the large yellow rubber sphere?



NLVR

- There is a box with exactly two blue items and exactly two black items.

Sort-of-CLEVR

- Is the blue object on the top or on the bottom?
- What is the color of the object that is closest to the red object?

Background

All these datasets

- are fixed and cannot (easily) be modified/extended.
- use a simple *ad hoc* grammar or crowd-sourcing for language.

What differentiates my system?

- It is a configurable data generator.
- Language synthesis is based on (D)MRS formalism, hence easily extensible, e.g. to **multi-lingual data**.
- Evaluation can focus on specific understanding abilities in any desired detail.

Papers (on arXiv):

- **ShapeWorld: A new test methodology for multimodal language understanding**. April 2017. *Alexander Kuhnle and Ann Copestake*.
- **Deep learning evaluation using deep linguistic processing**. June 2017. *Alexander Kuhnle and Ann Copestake*.

SIG 'hackathon'

Integrating other languages into ShapeWorld

Motivation:

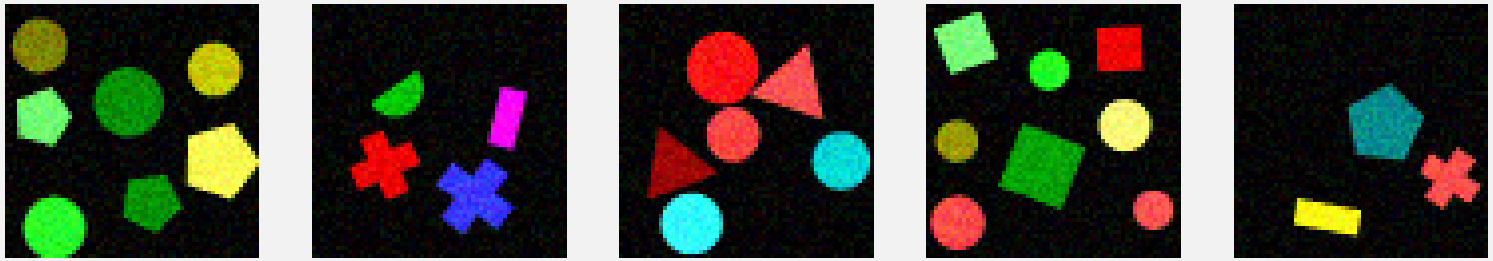
- Does the DMRS-based generation process generalize to other DELPH-IN grammars? What needs to be adapted?
- Controllable generation of visually grounded language data, for deep learning evaluation or whatever else...
- Ideal outcome: Paper on interesting evaluation findings for various languages.

SIG 'hackathon'

Integrating other languages into ShapeWorld

Motivation for grammar engineer:

- Non-trivial compositional language generation in action.
- Language synthesis for a limited visual domain with well-defined underlying semantics.
- If interest: Online demo platform which generates true/false statements about randomly sampled shape images:



Discussion points

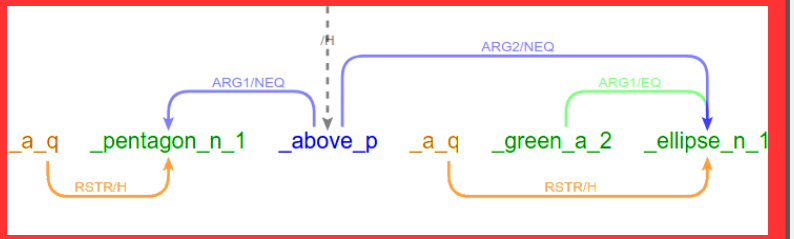
- Are other grammars compatible with this framework?
- What are the potential challenges?
- Is such a framework of more general interest, is it useful to grammar engineers?
- Is multi-lingual language synthesis done elsewhere?
- Any feedback, suggestions, comments welcome! ☺

ShapeWorld generation system

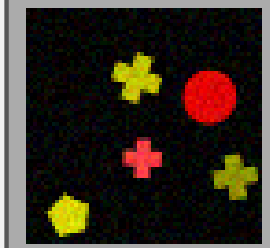
Generated world description

size: 64, **color**: {**name**: *black*, **shade**: 0.0}, **noise-range**: 0.1, **entities**:
[{ **shape**: {**name**: *cross*, **extent**: {**x**: 0.10, **y**: 0.10}}, **rotation**: 0.06,
color: {**name**: *yellow*, **shade**: -0.24}, **center**: {**x**: 0.47, **y**: 0.28} },
{ **shape**: {**name**: *cross*, **extent**: {**x**: 0.08, **y**: 0.08}}, **rotation**: 0.76,
color: {**name**: *red*, **shade**: 0.26}, **center**: {**x**: 0.49, **y**: 0.65} },
{ **shape**: {**name**: *pentagon*, **extent**: {**x**: 0.09, **y**: 0.08}}, **rotation**: 0.27,
color: {**name**: *yellow*, **shade**: -0.16}, **center**: {**x**: 0.15, **y**: 0.91} },
{ **shape**: {**name**: *circle*, **extent**: {**x**: 0.12, **y**: 0.12}}, **rotation**: 0.53,
color: {**name**: *red*, **shade**: -0.12}, **center**: {**x**: 0.80, **y**: 0.37} },
{ **shape**: **name**: *cross*, **extent**: {**x**: 0.09, **y**: 0.09}}, **rotation**: 0.73,
color: {**name**: *yellow*, **shade**: -0.42, **center**: {**x**: 0.92, **y**: 0.73} }]

Linguistic representation



World



Caption

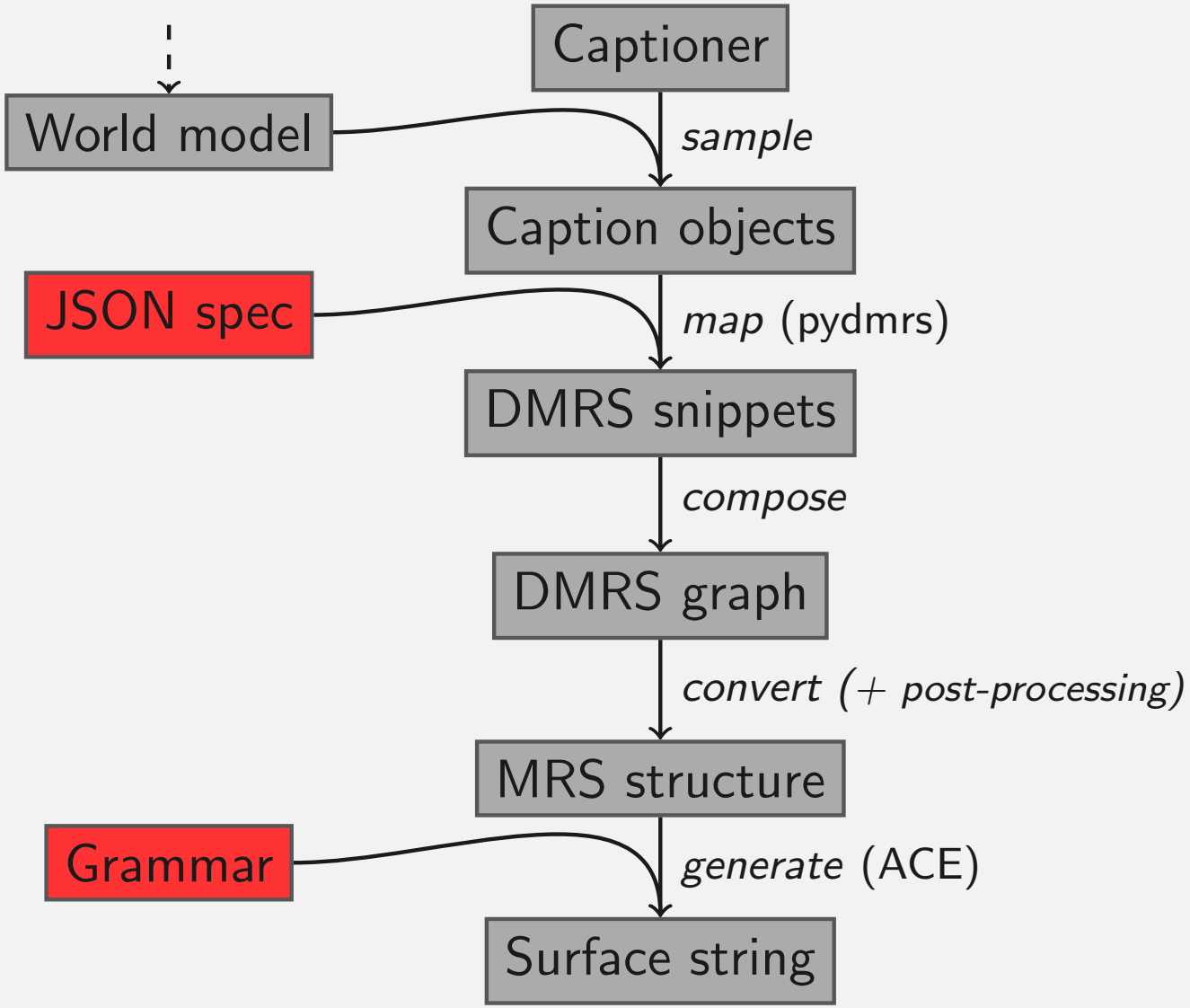
"There is a blue circle."

"Most crosses are yellow."

"A pentagon is below a cross."

Agreement?

Language generation in ShapeWorld



Captioner

Simple noun phrase captioner

`AttributeEntity` captioner:

1. Randomly pick an object
2. Extract its shape/color(/etc) attributes
3. If wrong statement, change attribute(s) in some way
4. Hypernym ratio: Randomly remove some attributes
5. Create `EntityType` object of `Attribute` list
6. Verify that the component dis-/agrees as expected

Examples:

"green ellipse", *"magenta shape"*, *"semicircle"*, *"shape"*, etc

Captioner

Counting quantifier captioner

`Counting` captioner:

- Randomly choose an 'absolute' quantifier
- Restrictor: `AttributeEntity`
- Body: `CaptionerMixer(AttributeRelation, SpatialRelation, ComparisonRelation)`

Examples:

- *"Four rectangles are blue."*
- *"Two shapes are above a red circle."*
- *"Both crosses are bigger than a pentagon."*

Realizer

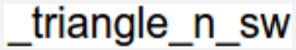
DMRS snippets defined in JSON file (per language)

```
"attributes": {
    "color": {
        "red": "[attr]:_red_a_1 e? =1=> [arg]:node", ...
    }, ...
},
"nouns": {
    "shape": {
        "square": "[noun]:_square_n_1 x?", ...
    }, ...
},
"relations": {
    "x-rel": {
        "-1": "[rel]:_to_p e? -2-> _left_n_of x[_s___] <-- _the_q;
               :_left_n_of <=1= _of_p e -2-> [ref]:node <-- _a_q",
        ...
    }, ...
}, ...
```

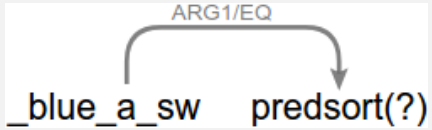
Realizer

DMRS snippet composition

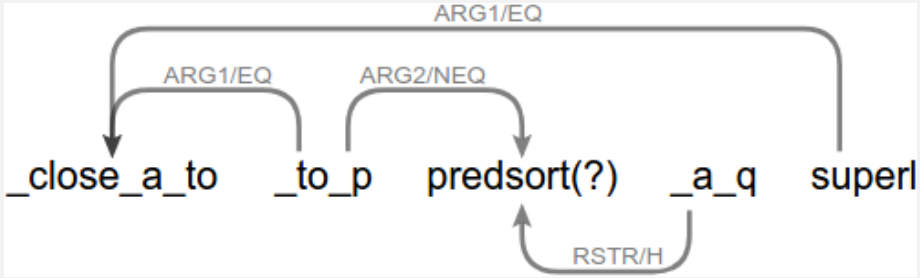
"triangle"



"blue"



"closest"




```
noun_dmrs.compose(mod_dmrs, fusion={'noun':'arg'})
```

1. Go through the fusion pairs and unify nodes, i.e. predicates and sortinfos (optionally using custom predicate hierarchy)
2. Add other nodes
3. Add missing links and overwrite links if necessary
4. Adopt handle/index

Realizer

Post-processing as paraphrase rules

- Required rules to make DMRS *'grammatical'*
- Optional rules for linguistic variety (e.g. hypernyms)
- Resolve *'inconsistencies'* with internal caption system (*"is a green shape"* $\rightarrow$ *"is green"*)
- Make captions sound more *'natural'* (*"a square is above a square"* $\rightarrow$ *"a square is above another square"*)

Example: hypernym for *"red square"*

```
{
  "search": "_red_a_1 e? =1=> [shape]:_square_n_1 x?",
  "replace": "[shape]:_red+square_n_1 x?"
}
```