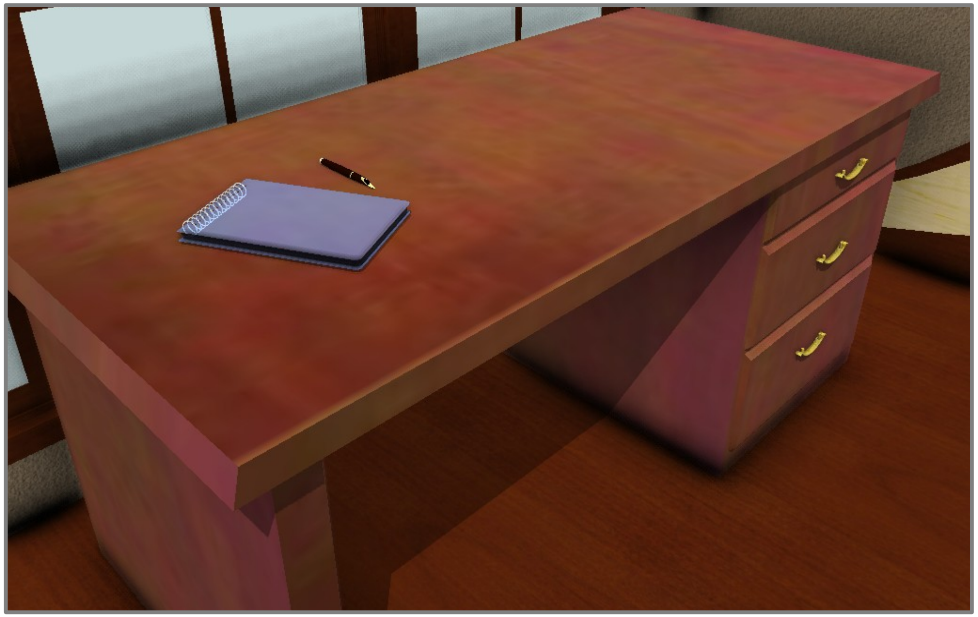
Text to 3D Scene Generation with Rich Lexical Grounding

Angel Chang Will Monroe Manolis Savva
Christopher Potts Christoper D. Manning
Stanford University



"There is a desk and there is a notepad on the desk. There is a pen next to the notepad."

ACL-IJCNLP July 27, 2015 Beijing, China

Outline

- Introduction and prior work
- Dataset
- Lexical learning
- Generation with lexical grounding
- Evaluation
- Challenges and Conclusion

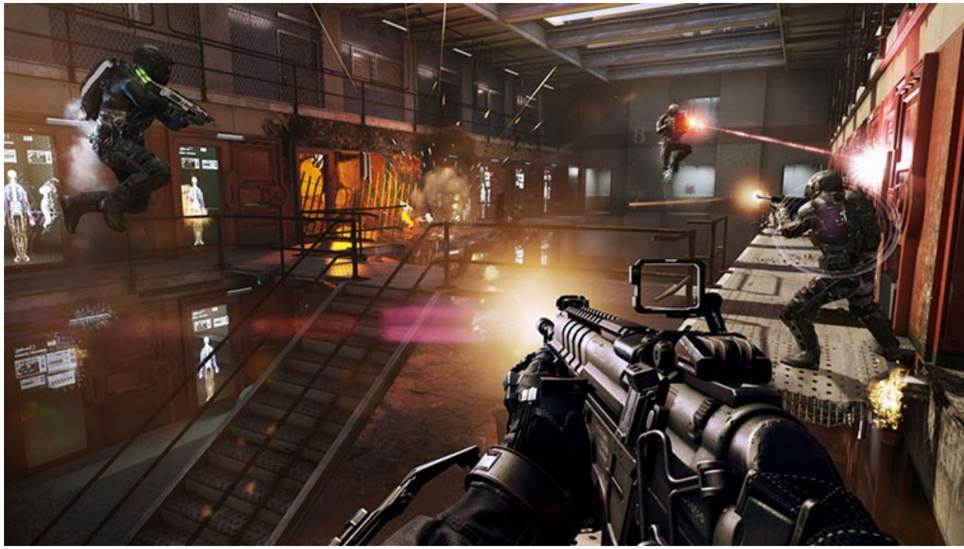
Outline

- Introduction and prior work
- Dataset
- Lexical learning
- Generation with lexical grounding
- Evaluation
- Challenges and conclusion

The art of 3D scene design

The art of 3D scene design

Call of Duty: Advanced Warfare
[Activision / Sledgehammer Games]



The art of 3D scene design

Call of Duty: Advanced Warfare
[Activision / Sledgehammer Games]

Toy Story 3
[Disney / Pixar]

The art of 3D scene design

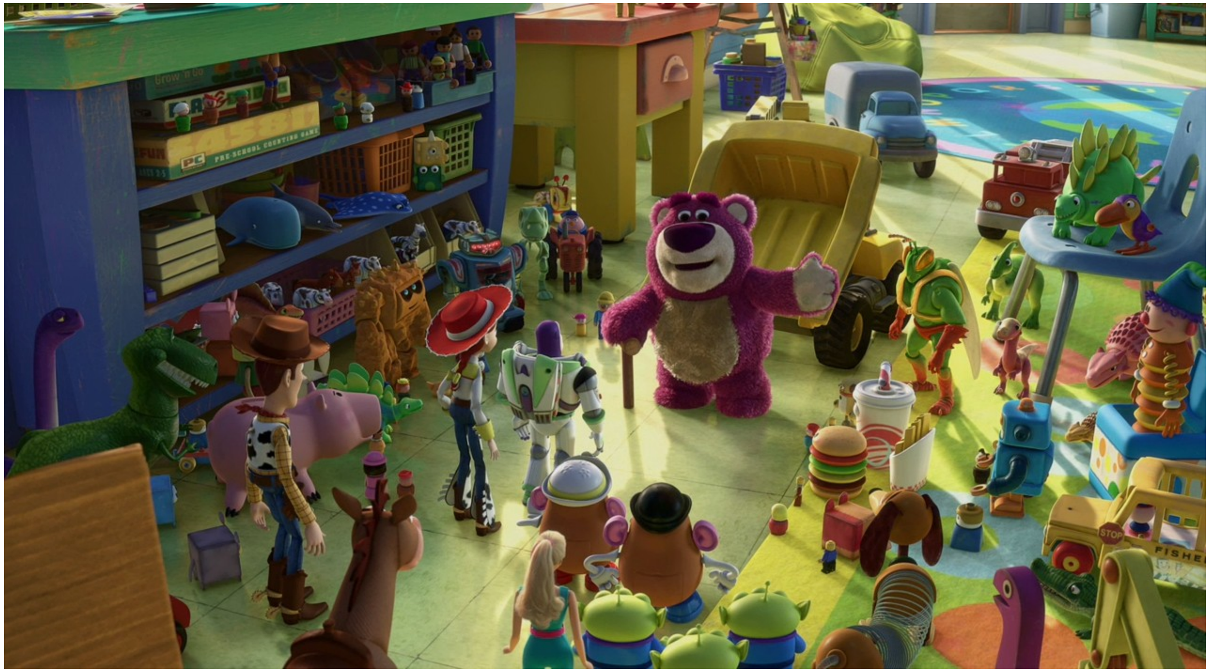
Call of Duty: Advanced Warfare
[Activision / Sledgehammer Games]

Toy Story 3
[Disney / Pixar]

“Modern: Plywood, Plastic & Polished Metal”
[Homedit Interior Design & Architecture]

Generating 3D scenes from text

Generating 3D scenes from text



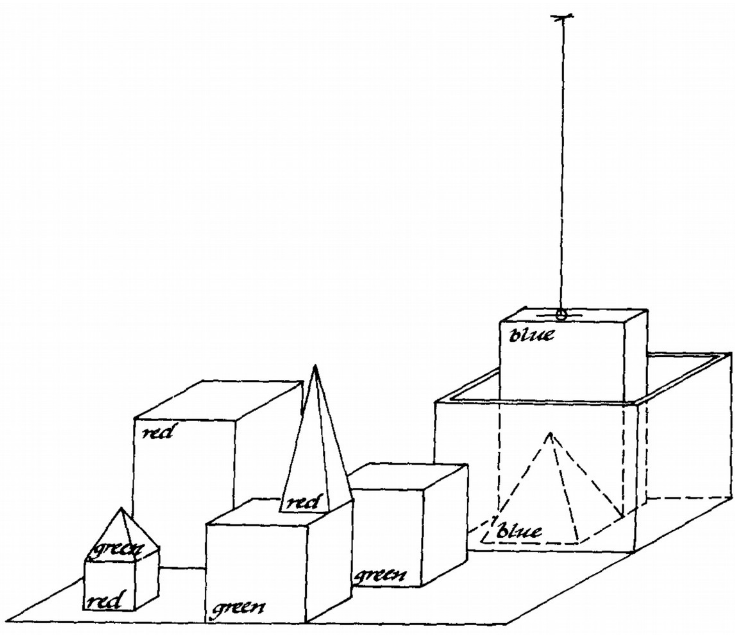
```
TOYS' POV -- An idyllic day care classroom, filled with the happy bustle
of four- and five-year-olds, playing with toys -- dinosaurs, a baby
doll, a pink Teddy bear, a Ken doll. ...

A Tonka Truck races forward, then backs up in a quick 180 arc, revealing
a large pink Teddy bear, LOTSO, in its bed. Lotso taps a Tinker Toy cane
and the truck bed rises, "dumping" him out. Like Bob Hope stepping off
the links in Palm Springs, Lotso exudes an easy, cheerful charisma.
```

(Screenplay by Michael Arndt)

Selected prior work

SHRDLU (Winograd, 1972)



WordsEye (Coyne and Sproat, 2001)

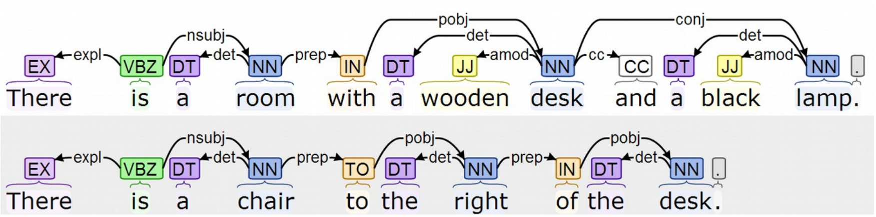
Scene generation pipeline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.

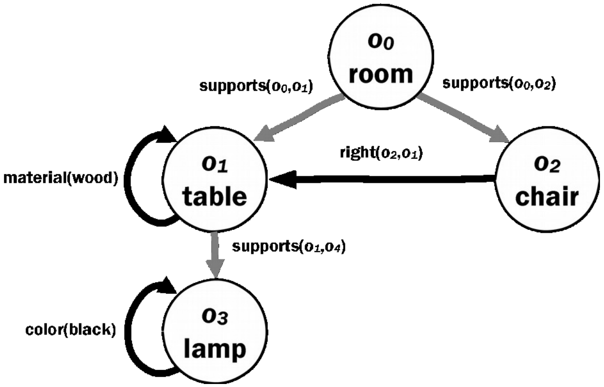
(Chang et al., 2014)

Scene generation pipeline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.


EX VBZ DT NN IN DT JJ NN CC DT JJ NN .
expl nsubj det prep pobj det amod cc conj det amod
There is a room with a wooden desk and a black lamp.
EX VBZ DT NN TO DT NN IN DT NN .
expl nsubj det prep pobj det prep pobj det
There is a chair to the right of the desk.

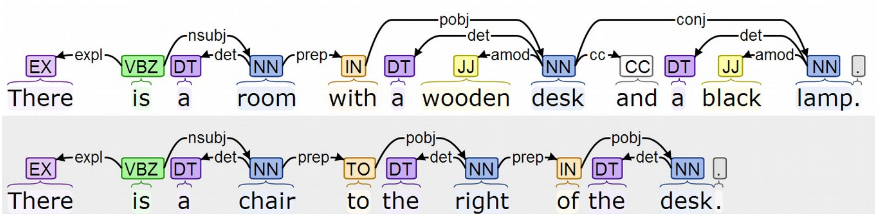
parsing



(Chang et al., 2014)

Scene generation pipeline

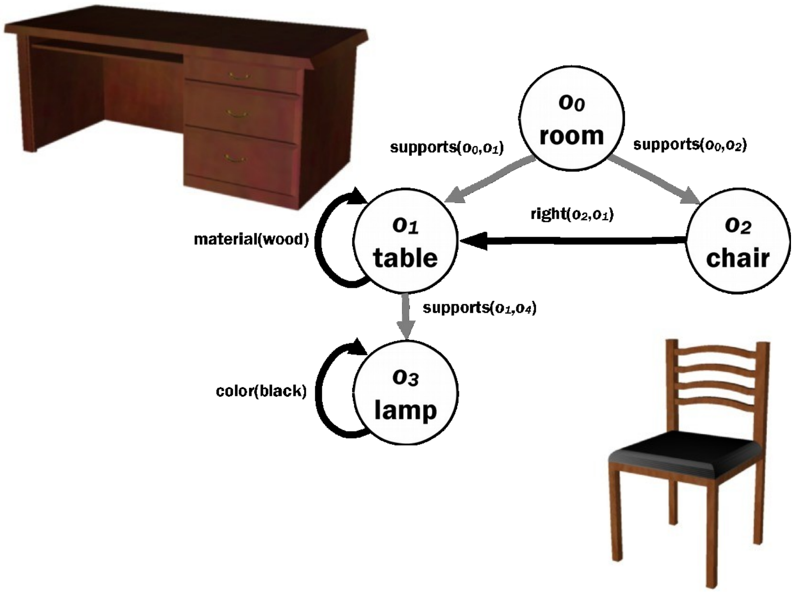
There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.



EX VBZ DT NN IN DT JJ NN CC DT JJ NN .
There is a room with a wooden desk and a black lamp.
expl nsubj det prep pobj det amod cc conj det amod
EX VBZ DT NN TO DT NN IN DT NN .
There is a chair to the right of the desk.
expl nsubj det prep pobj det prep pobj det

parsing

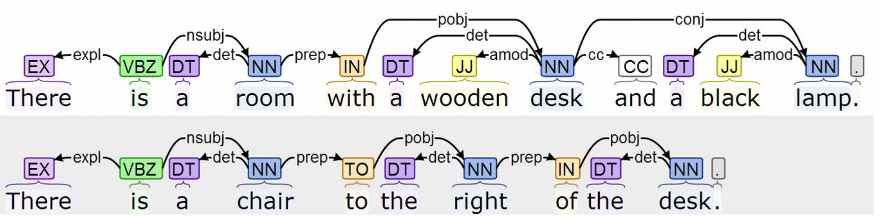
object selection



(Chang et al., 2014)

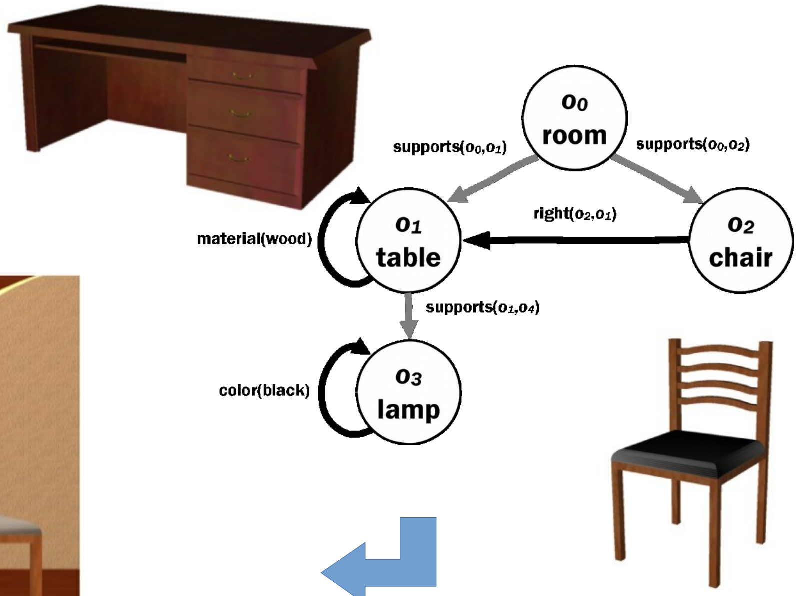
Scene generation pipeline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.

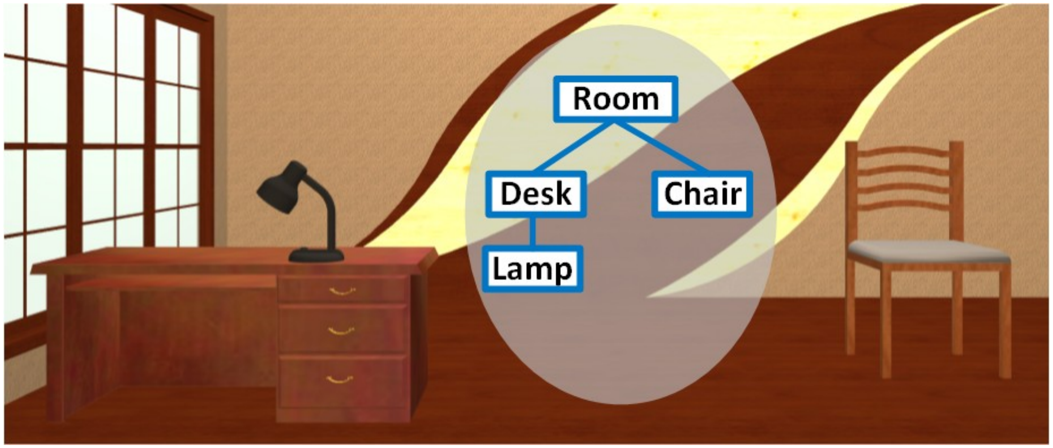

EX VBZ DT NN IN DT JJ NN CC DT JJ NN .
expl nsubj det prep pobj det amod conj cc det amod
There is a room with a wooden desk and a black lamp.
EX VBZ DT NN TO DT NN IN DT NN .
expl nsubj det prep pobj det prep pobj det
There is a chair to the right of the desk.

parsing

object selection



layout



(Chang et al., 2014)

Handling lexical variety

sofa

couch

loveseat

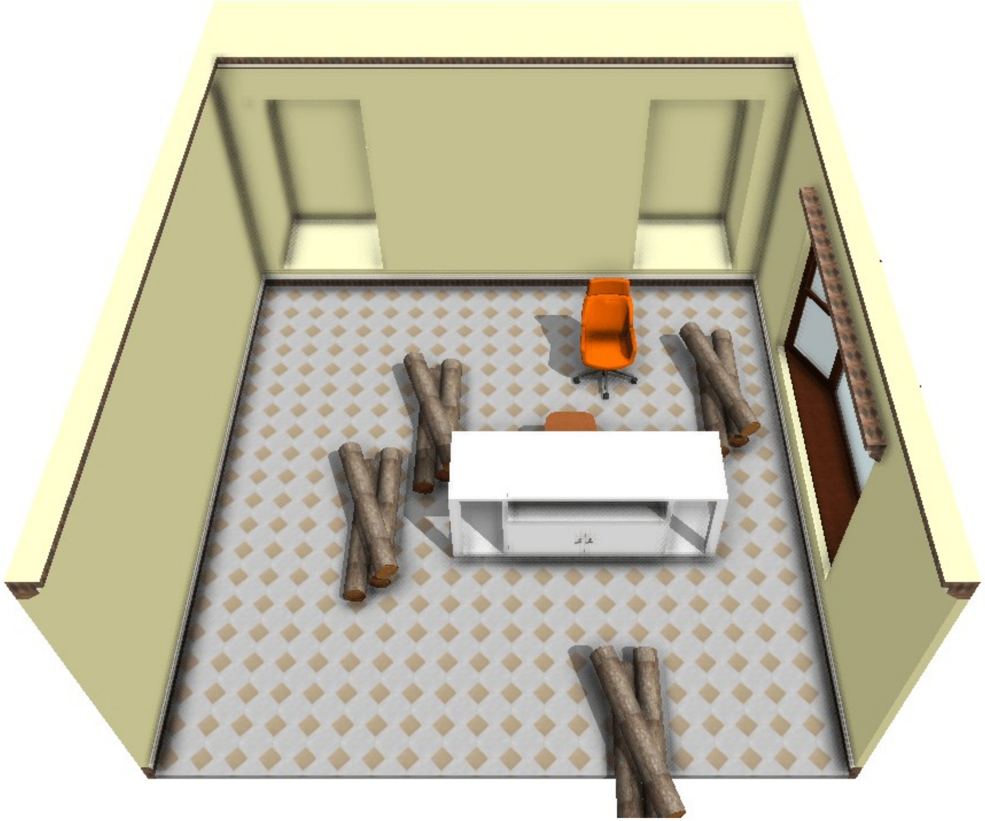
dresser

chest of drawers

cabinet

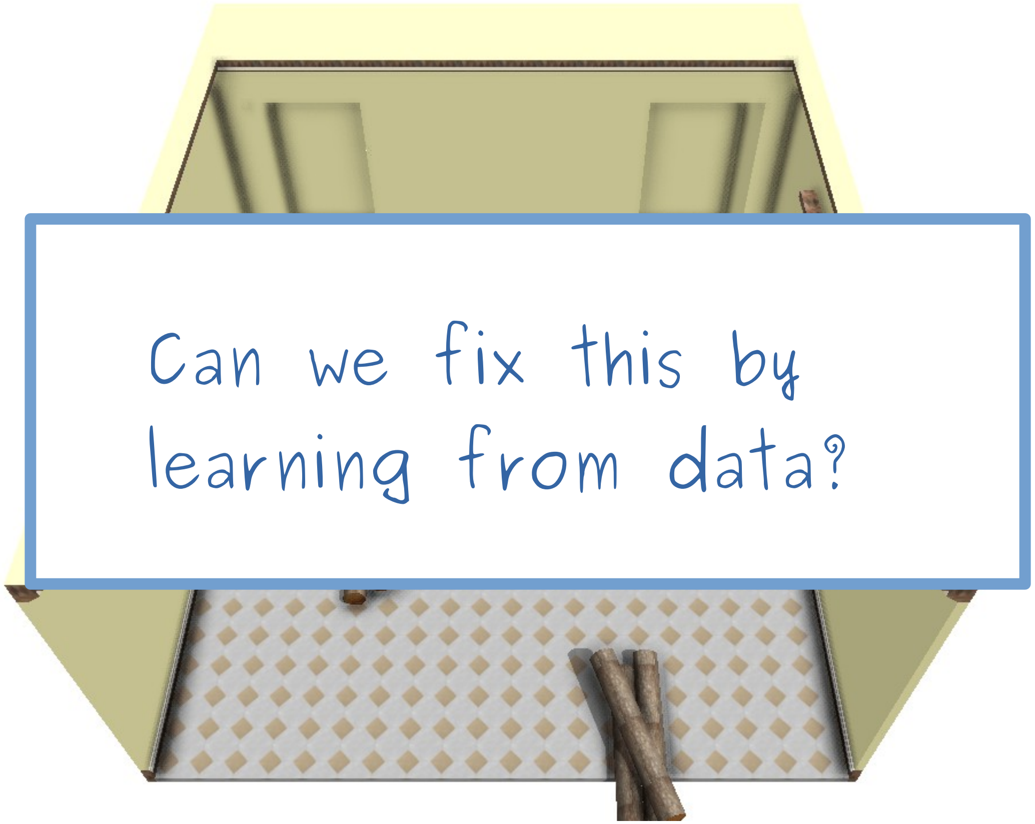
Identifying object mentions

*Wood table and **four wood** chairs in the center of the room*



Identifying object mentions

*Wood table and **four wood** chairs in the center of the room*



Can we fix this by learning from data?

Outline

- Introduction and prior work
- Dataset
- Lexical learning
- Generation with lexical grounding
- Evaluation
- Challenges and conclusion

Outline

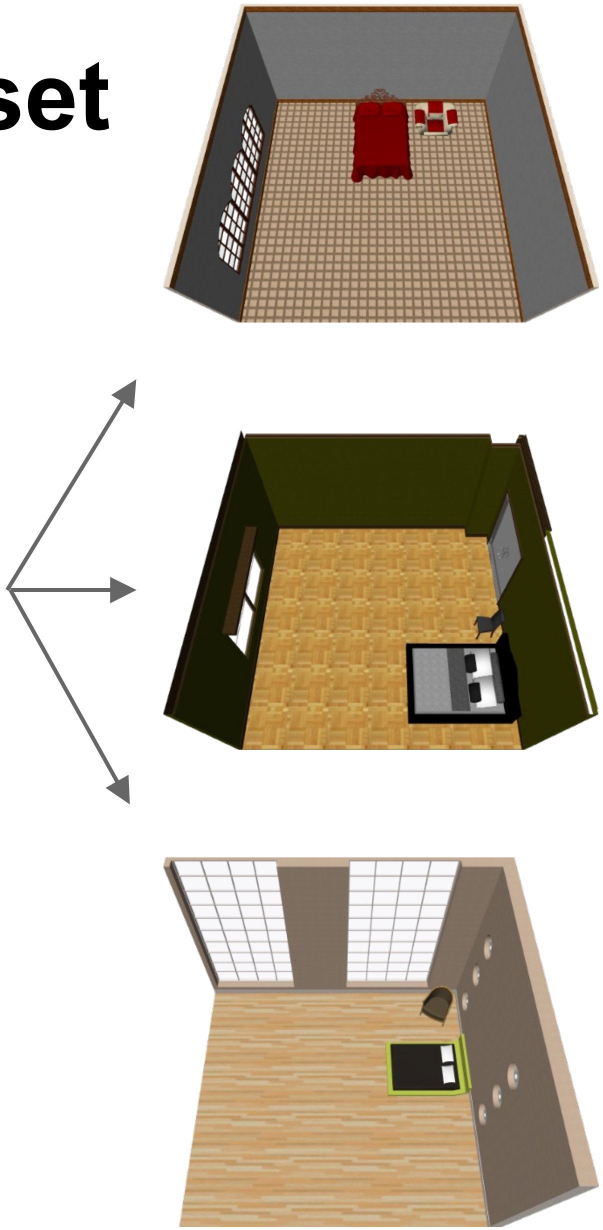
- Introduction and prior work
- Dataset
- Lexical learning
- Generation with lexical grounding
- Evaluation
- Challenges and conclusion

Dataset

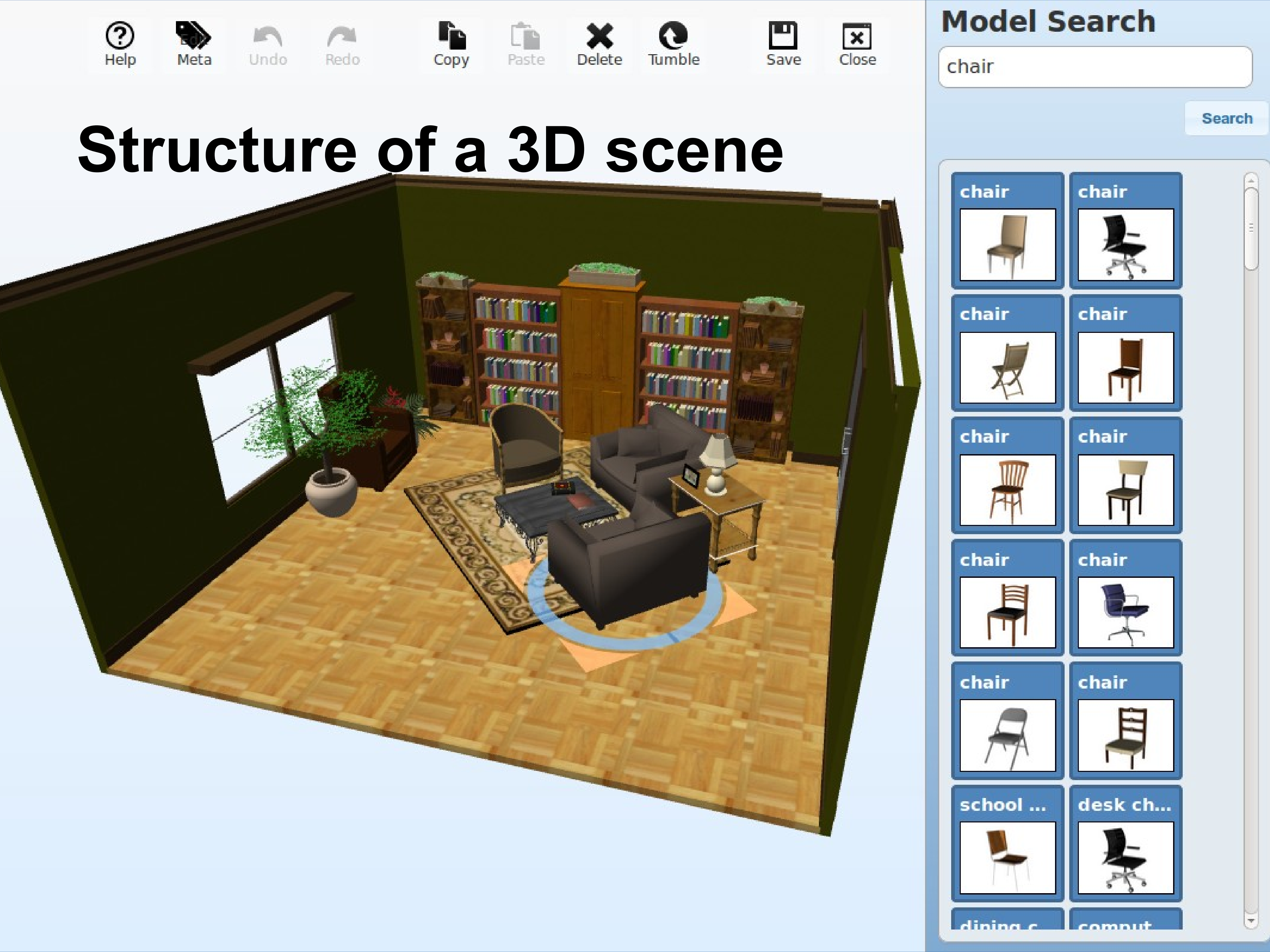
There is a bed and there is a chair next to the bed.

Dataset

There is a bed and there is a chair next to the bed.



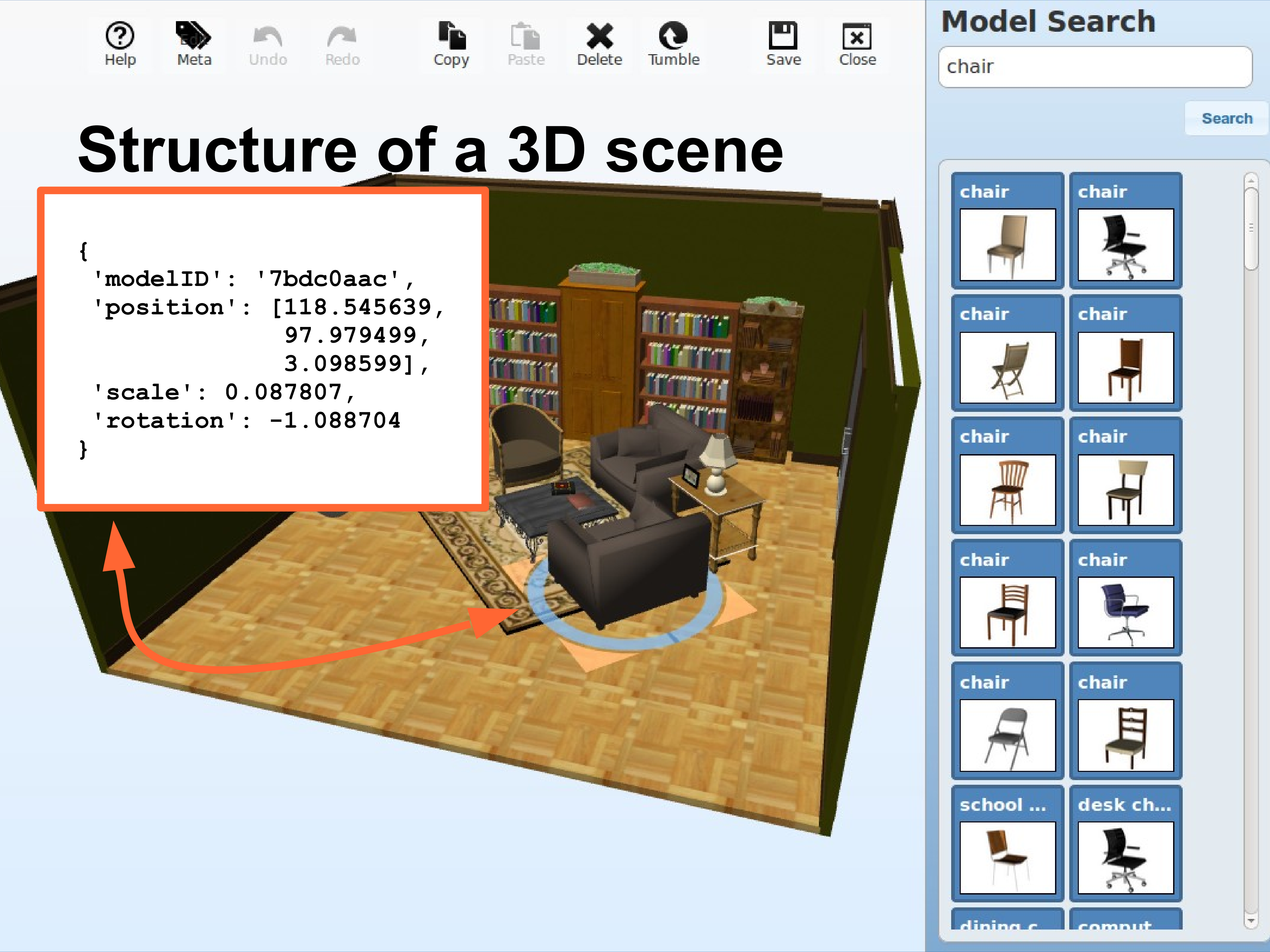
Structure of a 3D scene



Model Search

chair

Structure of a 3D scene



```
{
 'modelID': '7bdc0aac',
 'position': [118.545639,
              97.979499,
              3.098599],
 'scale': 0.087807,
 'rotation': -1.088704
}
```

Model Search

chair

Structure of a 3D scene

```
{
 'modelID': '7bdc0aac',
 'position': [118.545639,
              97.
              3.0
 'scale': 0.08780
 'rotation': -1.0
}
```

Field	Value
name	ellington armchair
id	7bdc0aac
tags	armchair, chair, ellington, haughton, sam, seating, woodmark
category	Chair
wnlemmas	armchair
unit	0.028974
up	[0, 0, 1]
front	[0, -1, 0]

Structure of a 3D scene

```
{
 'modelID': '7bdc0aac'
 'position': [118.545639,
              97.9
              3.09
 'scale': 0.087807
 'rotation': -1.08
}
```

human-tagged keywords & categories

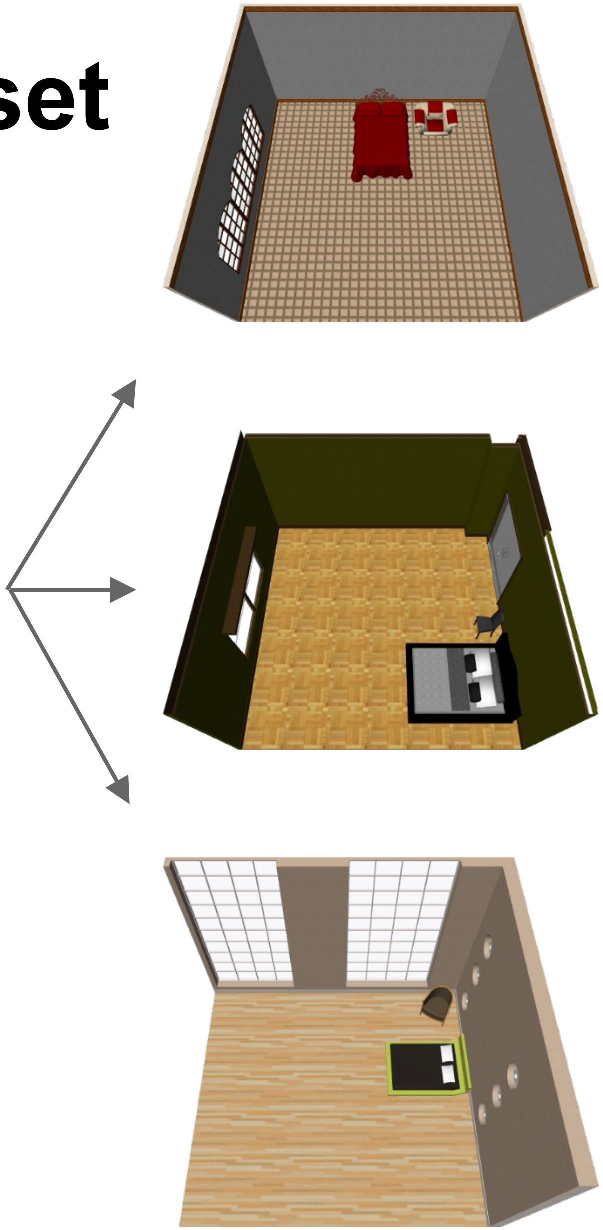
WordNet

size & orientation suggestions

Field	Value
name	ellington armchair
id	7bdc0aac
tags	armchair, chair, ellington, haughton, sam, seating, woodmark
category	Chair
wnlemmas	armchair
unit	0.028974
up	[0, 0, 1]
front	[0, -1, 0]

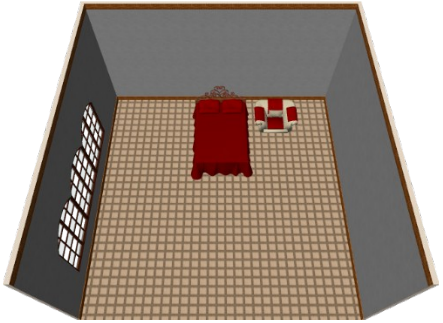
Dataset

There is a bed and there is a chair next to the bed.



Dataset

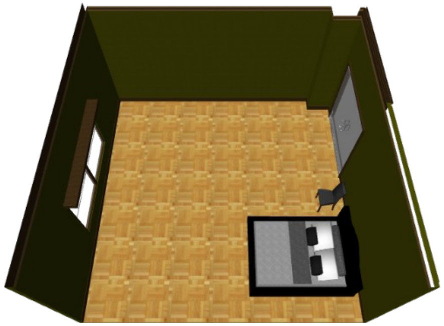
There is a bed and there is a chair next to the bed.



The room has three windows on one wall. There is a red bed in the back of the room. Along side the bed is a side chair that is red and white.

This room has a bed with red bedding against the wall. Next to the bed is a chair.

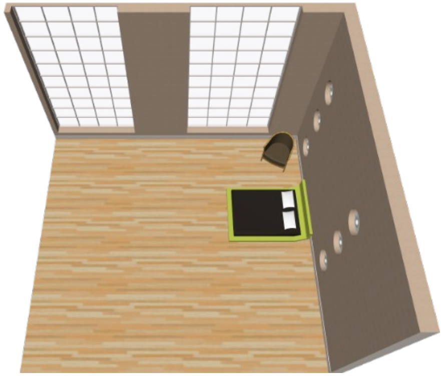
there is a antique looking bed with red covers and pillows in a room. next to it is a recliner chair with red padding. also there are windows.



there is a bed with five pillows on it, and next to it is a chair

There is a bed in the room with two pillows and a small chair near to the right side of it.

There is a large grey bed in the bottom right corner of the room. Above the bed is a small black chair.



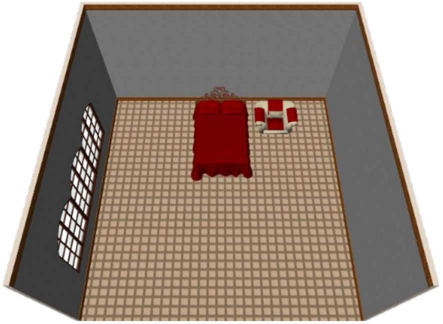
Floor to ceiling windows on back wall. Green bed with two pillows and black blanket. Lights recessed into right side wall. Light wood flooring. A chair is in the upper right hand corner

There is a bed on the side of the room. There is a chair in the corner, next to the windows.

I see a bed and a chair.

Dataset

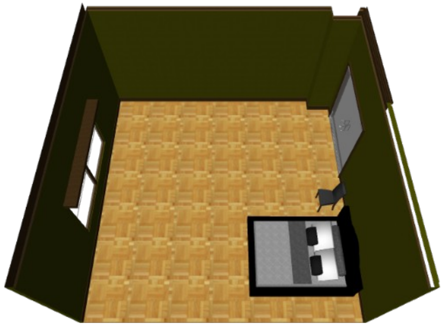
There is a bed and there is a chair next to the bed.



The room has three windows on one wall. There is a red bed in the back of the room. Along side the bed is a side chair that is red and white.

This room has a bed with red bedding against the wall. Next to the bed is a chair.

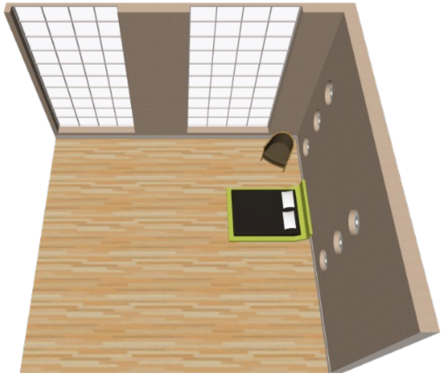
there is a antique looking bed with red covers and pillows in a room. next to it is a recliner chair with red padding. also there are windows.



there is a bed with five pillows on it, and next to it is a chair

There is a bed in the room with two pillows and a small chair near to the right side of it.

There is a large grey bed in the bottom right corner of the room. Above the bed is a small black chair.



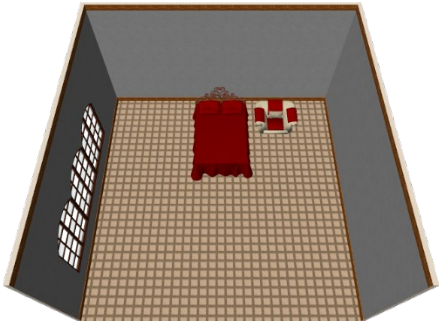
Floor to ceiling windows on back wall. Green bed with two pillows and black blanket. Lights recessed into right side wall. Light wood flooring. A chair is in the upper right hand corner

There is a bed on the side of the room. There is a chair in the corner, next to the windows.

I see a bed and a chair.

Dataset

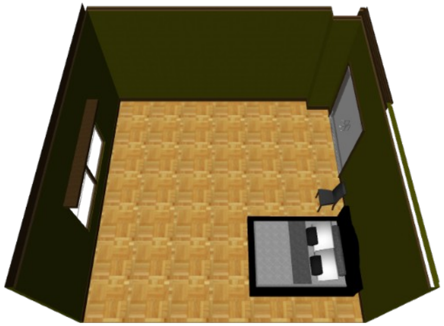
There is a bed and there is a chair next to the bed.



The room has three windows on one wall. There is a red bed in the back of the room. Along side the bed is a side chair that is red and white.

This room has a bed with red bedding against the wall. Next to the bed is a chair.

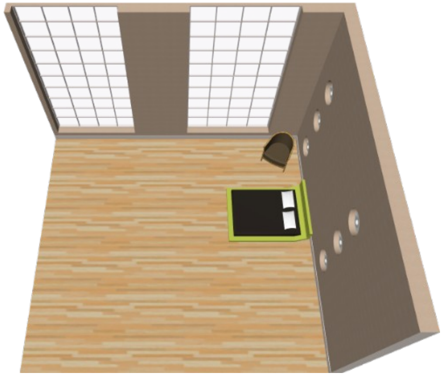
there is a antique looking bed with red covers and pillows in a room. next to it is a recliner chair with red padding. also there are windows.



there is a bed with five pillows on it, and next to it is a chair

There is a bed in the room with two pillows and a small chair near to the right side of it.

There is a large grey bed in the bottom right corner of the room. Above the bed is a small black chair.



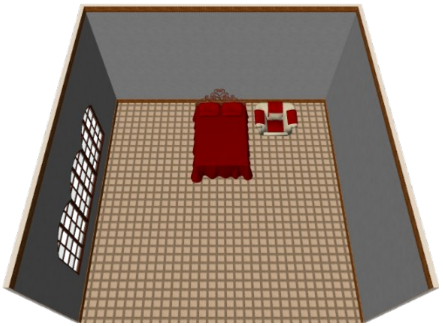
Floor to ceiling windows on back wall. Green bed with two pillows and black blanket. Lights recessed into right side wall. Light wood flooring. A chair is in the upper right hand corner

There is a bed on the side of the room. There is a chair in the corner, next to the windows.

I see a bed and a chair.

Dataset

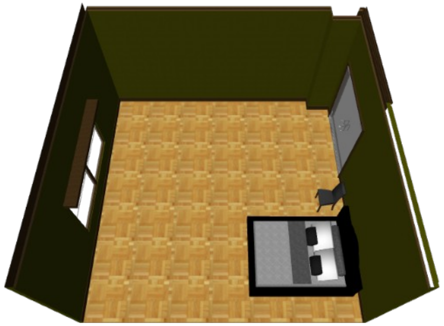
There is a bed and there is a chair next to the bed.



The room has three windows on one wall. There is a red bed in the back of the room. Along side the bed is a side chair that is red and white.

This room has a bed with red bedding against the wall. Next to the bed is a chair.

there is a antique looking bed with red covers and pillows in a room. next to it is a recliner chair with red padding. also there are windows.



there is a bed with five pillows on it, and next to it is a chair

There is a bed in the room with two pillows and a small chair near to the right side of it.

There is a large grey bed in the bottom right corner of the room. Above the bed is a small black chair.

60 seed sentences

1128 scenes

4284 scene descriptions

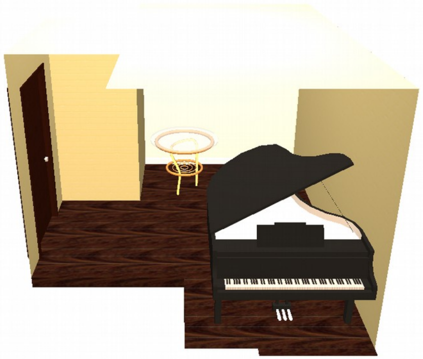
Outline

- Introduction and prior work
- Dataset
- Lexical learning
- Generation with lexical grounding
- Evaluation
- Challenges and conclusion

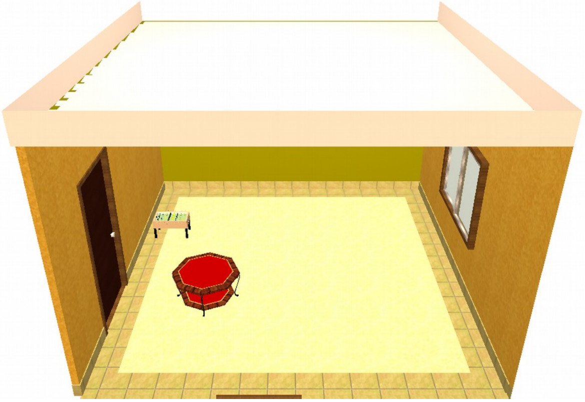
Discrimination task

brown room with a refrigerator in the back corner

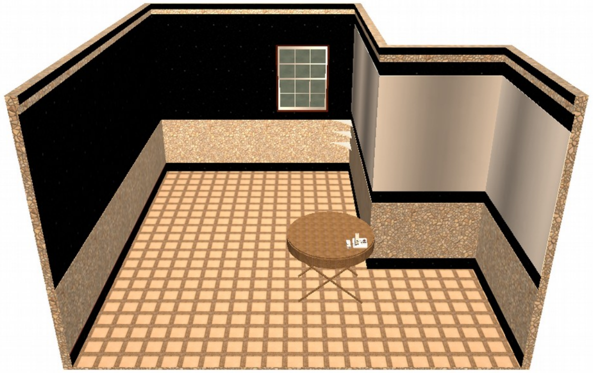
A



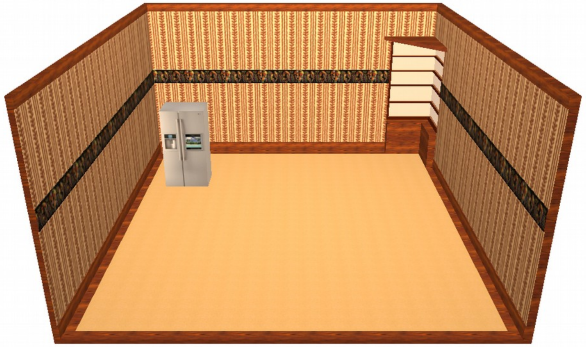
B



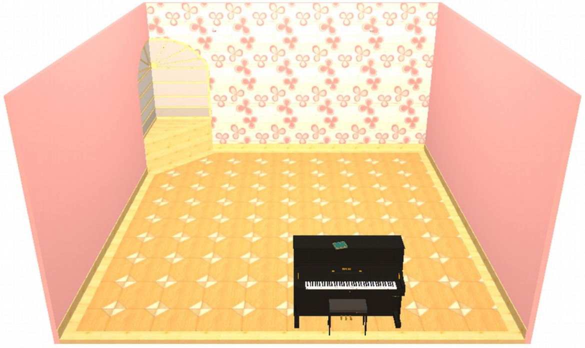
C



D



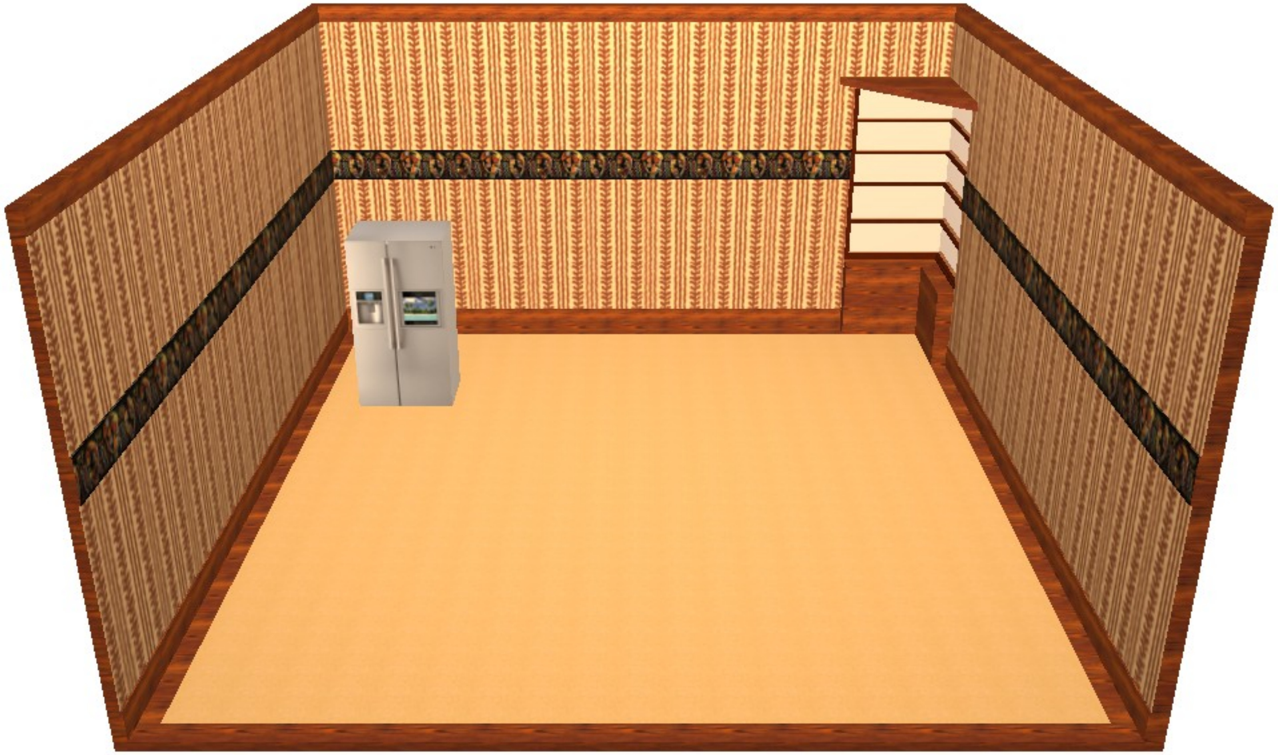
E



Discrimination task

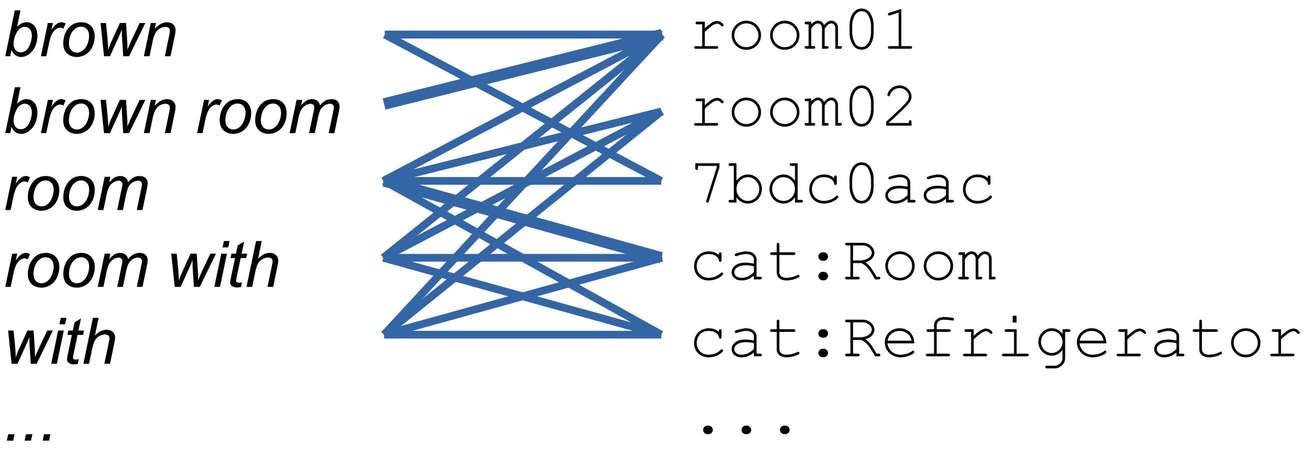
brown room with a refrigerator in the back corner

Ⓓ



Learning lexical items

- One-vs.-all logistic regression
- Features: **1**{(language, object)}
  - language: bag-of-words / bag-of-bigrams
  - object: model id / category



brown
brown room
room
room with
with
...

`room01`
`room02`
`7bdc0aac`
`cat:Room`
`cat:Refrigerator`
`...`

Discrimination results

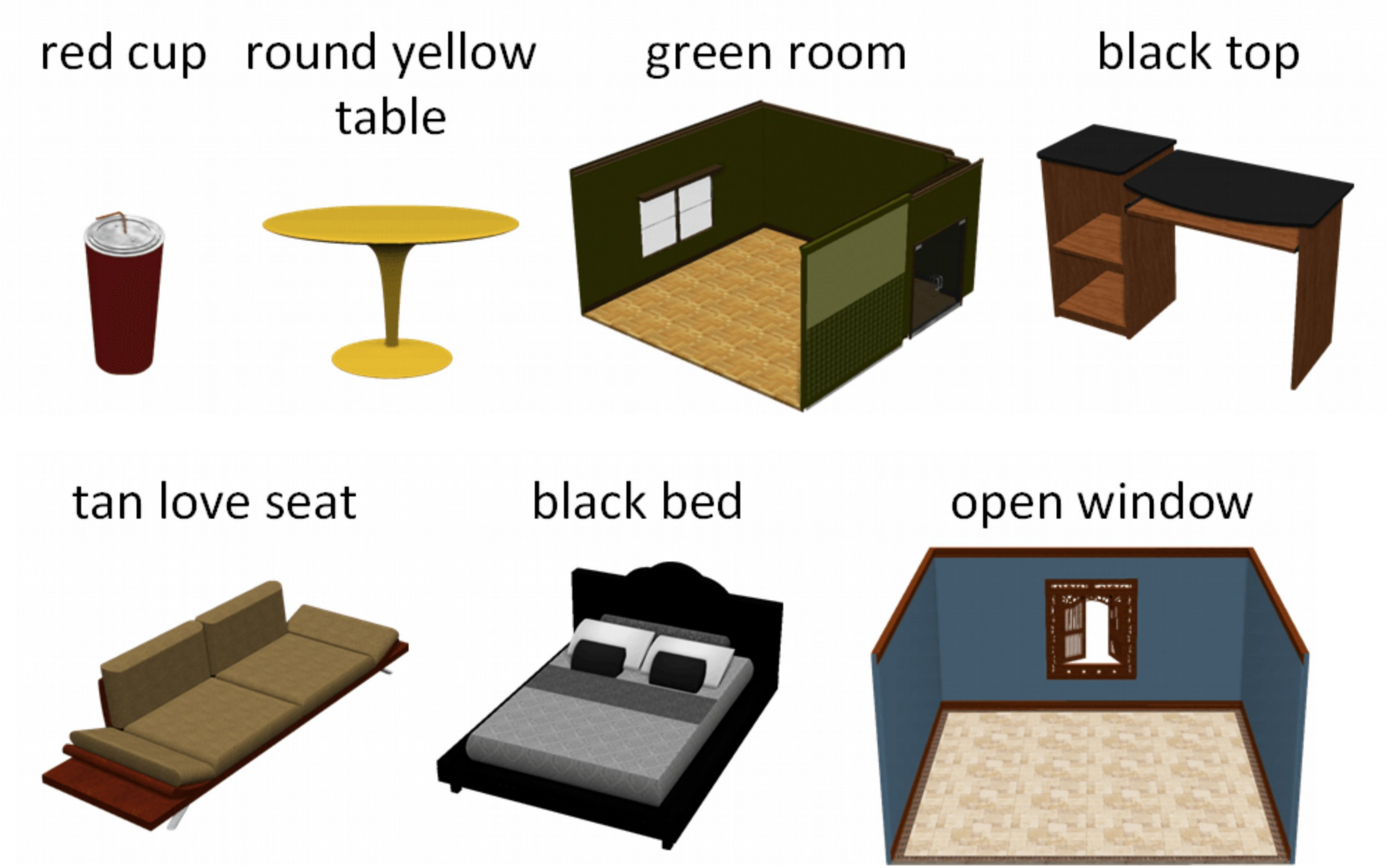
- Accuracy (% correct scenes identified)

	Random set
Model ids only	71.5%
Model ids + categories	**83.3%**

Lexical grounding examples

text	category
chair	Chair
couch	Couch
sofa	Couch
fruit	Bowl
bookshelf	Bookcase

Lexical grounding examples

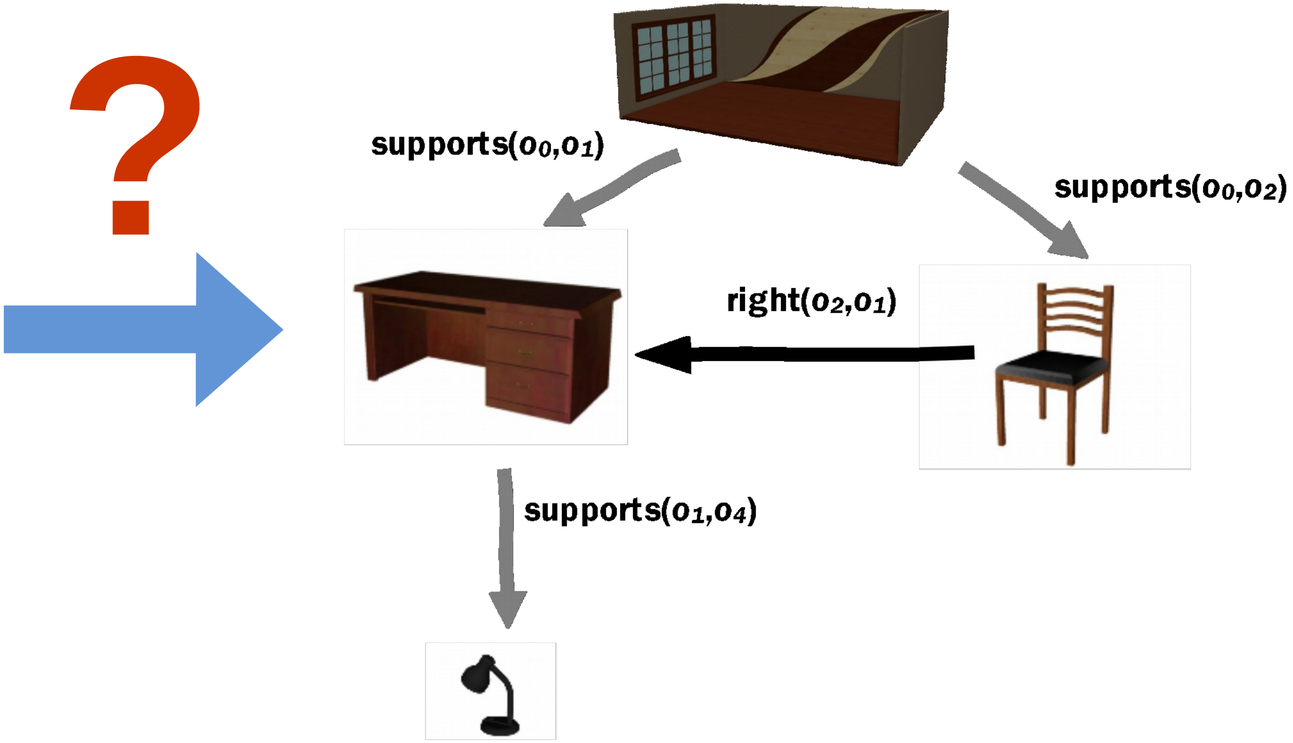


Outline

- Introduction and prior work
- Dataset
- Lexical learning
- Generation with lexical grounding
- Evaluation
- Challenges and conclusion

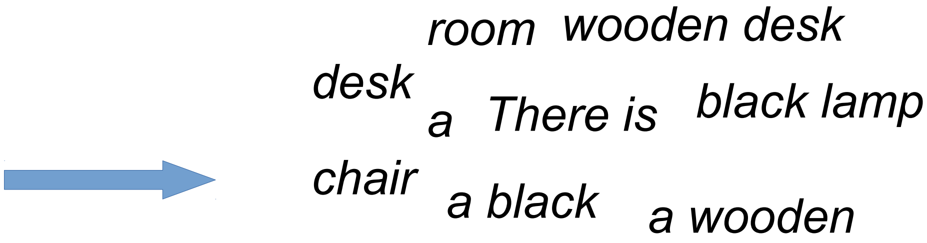
Generate!

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.



Baseline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.



room wooden desk

desk

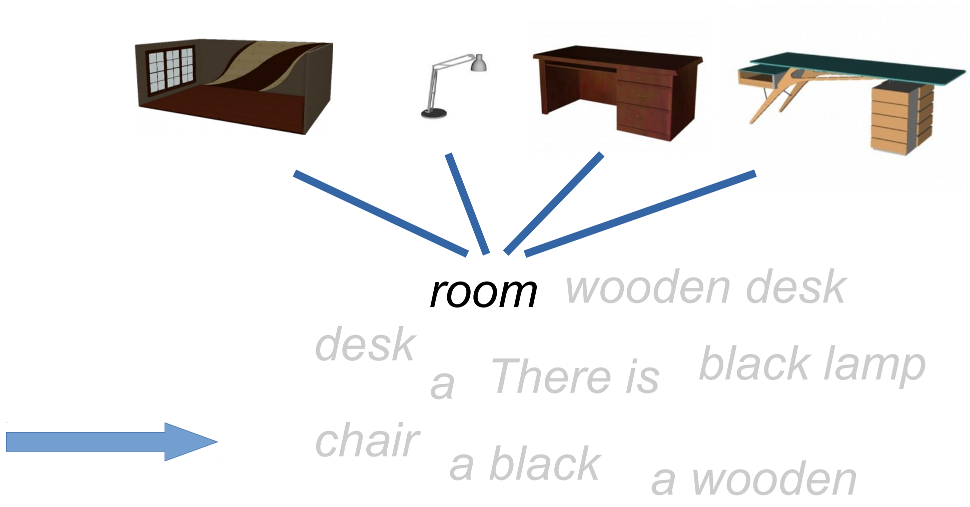
a There is black lamp

chair

a black a wooden

Baseline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.



room

wooden desk

desk

a

There is

black lamp

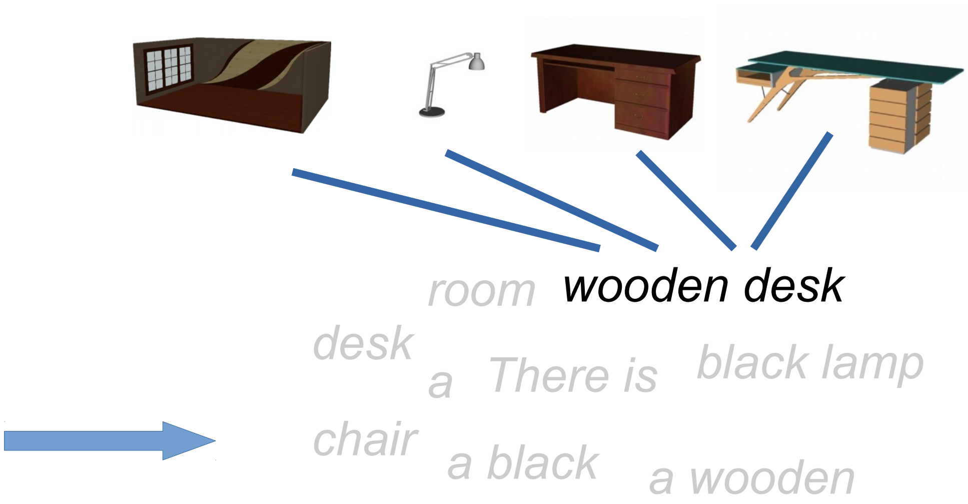
chair

a black

a wooden

Baseline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.



wooden desk

room

desk

a

There is

black lamp

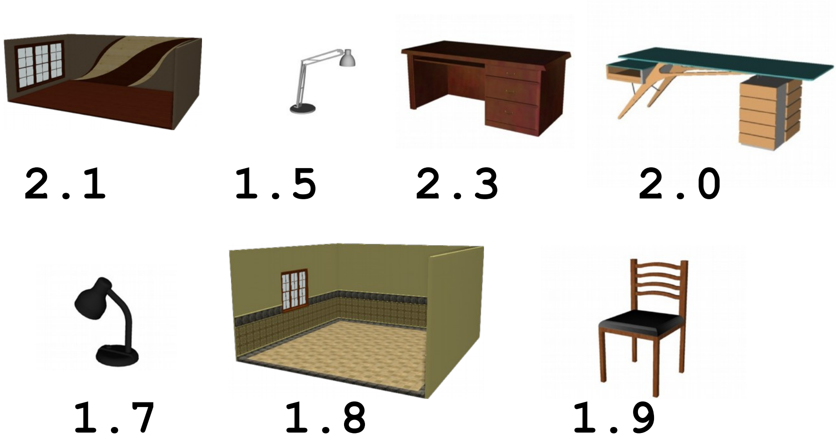
chair

a black

a wooden

Baseline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.

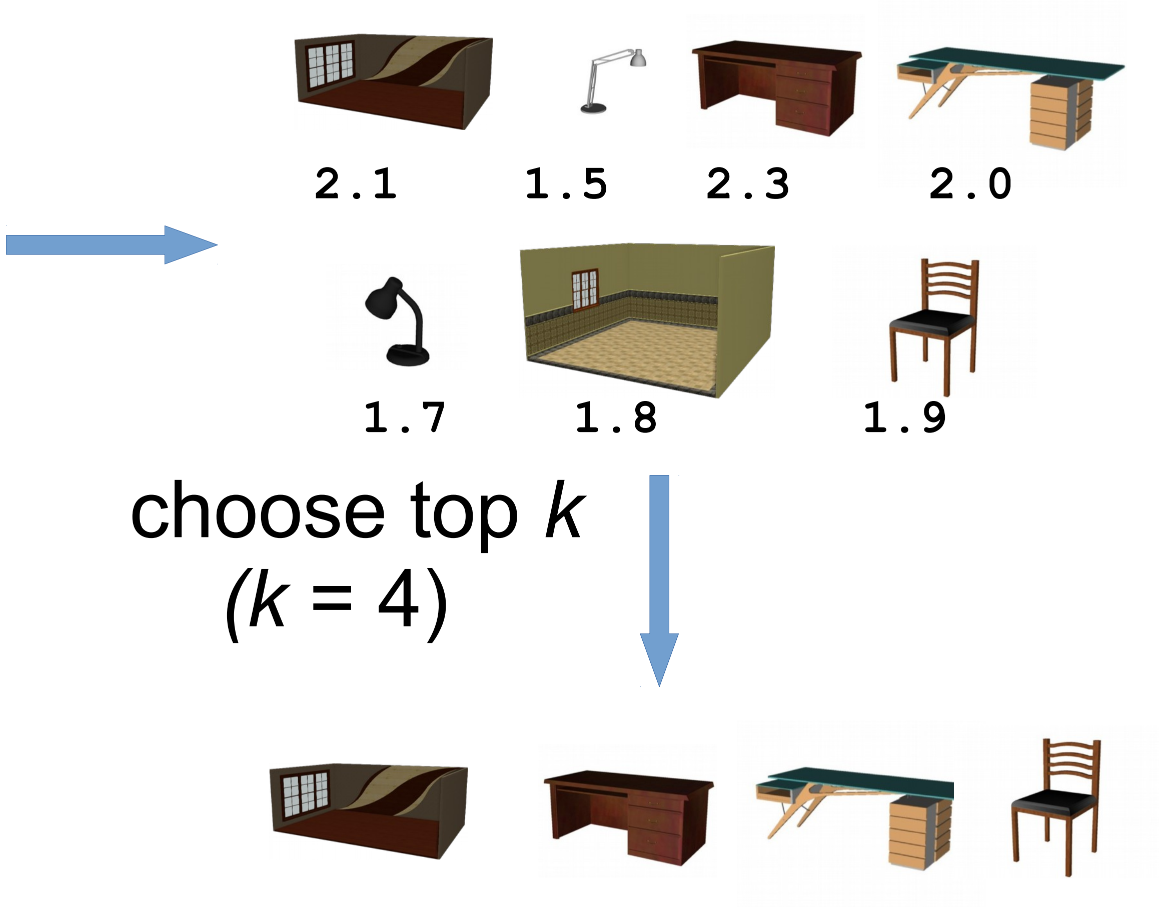


group by object

sum weights

Baseline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.

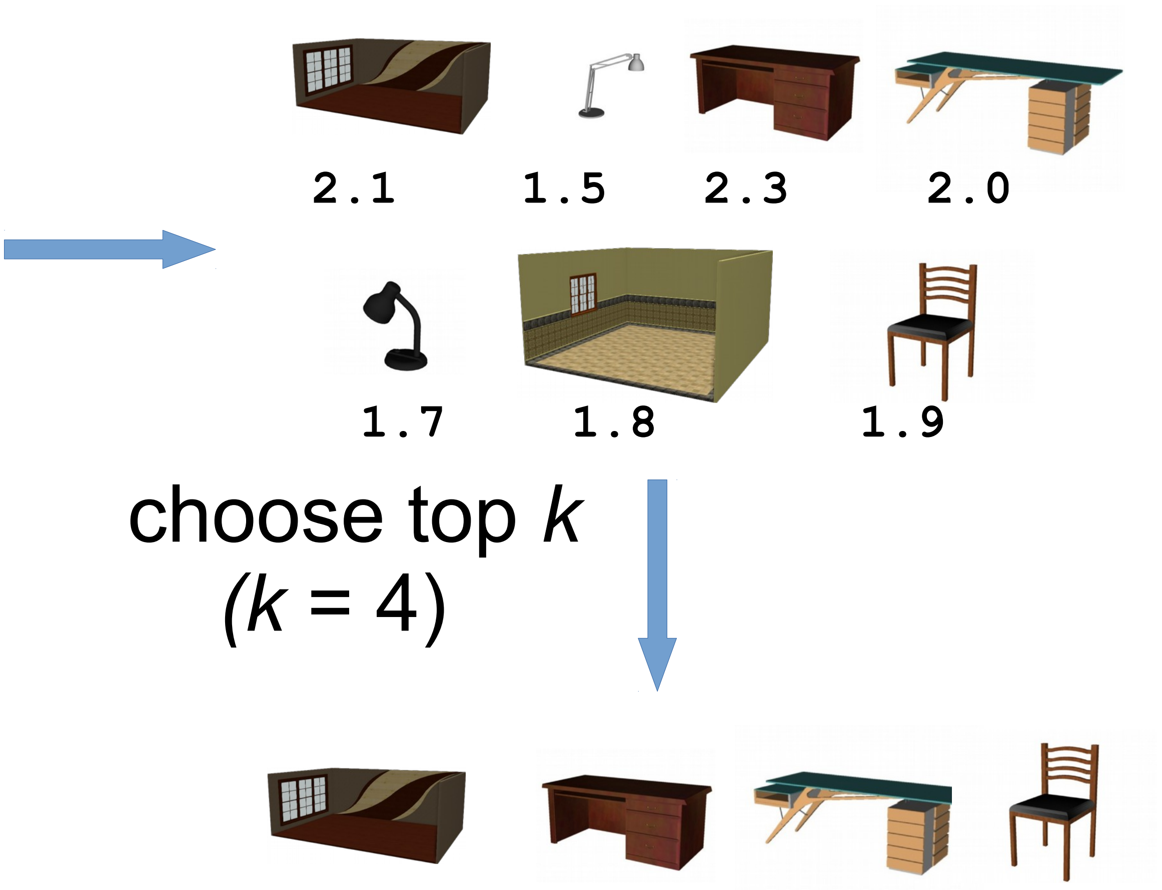


choose top *k*
(*k* = 4)

K = 4, average number of objects in human-constructed scenes

Baseline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.

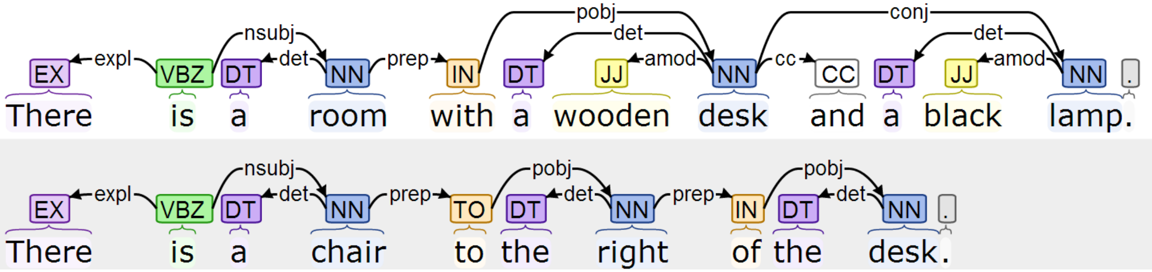


choose top *k*
(*k* = 4)

No relationship enforced between objects!
Combine with rule-based parser?

Rule-based parsing

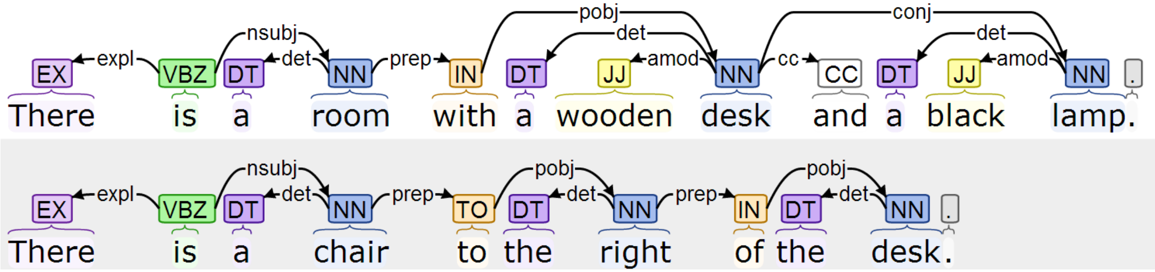
There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.


pobj
det
conj
det
nsubj
det
expl
prep
amod
cc
amod
EX VBZ DT NN IN DT JJ NN CC DT JJ NN .
There is a room with a wooden desk and a black lamp.
nsubj
pobj
pobj
expl
det
prep
det
prep
det
EX VBZ DT NN TO DT NN IN DT NN .
There is a chair to the right of the desk.

(Chang et al., 2014)

Rule-based parsing

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.



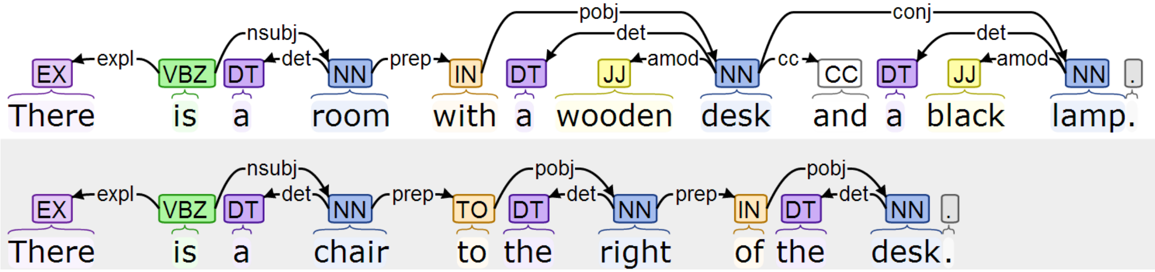
EX VBZ DT NN IN DT JJ NN CC DT JJ NN .
There is a room with a wooden desk and a black lamp.
expl
nsubj
det
prep
pobj
det
amod
cc
conj
det
amod
EX VBZ DT NN TO DT NN IN DT NN .
There is a chair to the right of the desk.
expl
nsubj
det
prep
pobj
det
prep
pobj
det

- Identify **object categories** using **noun phrases**

(Chang et al., 2014)

Rule-based parsing

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.

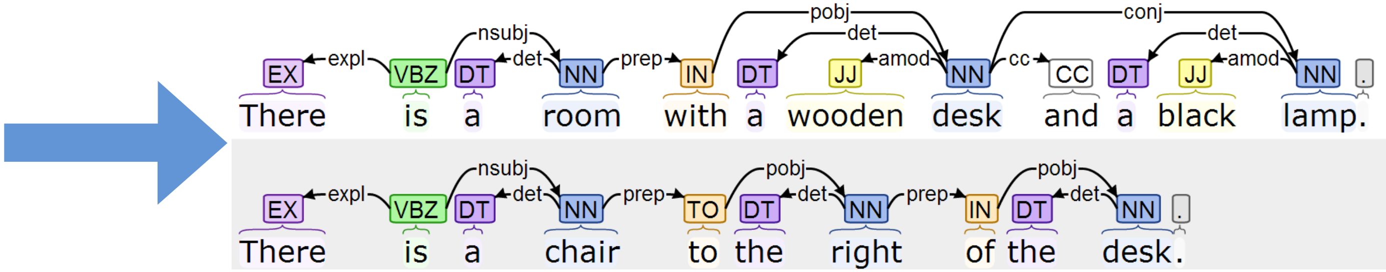

EX VBZ DT NN IN DT JJ NN CC DT JJ NN .
There is a room with a wooden desk and a black lamp.
expl
nsubj
det
prep
pobj
det
amod
cc
conj
det
amod
EX VBZ DT NN TO DT NN IN DT NN .
There is a chair to the right of the desk.
expl
nsubj
det
prep
pobj
det
prep
pobj
det

- Identify object categories using noun phrases
- Identify **attributes** and **keywords** using **modifiers and dependency patterns**

(Chang et al., 2014)

Rule-based parsing

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.

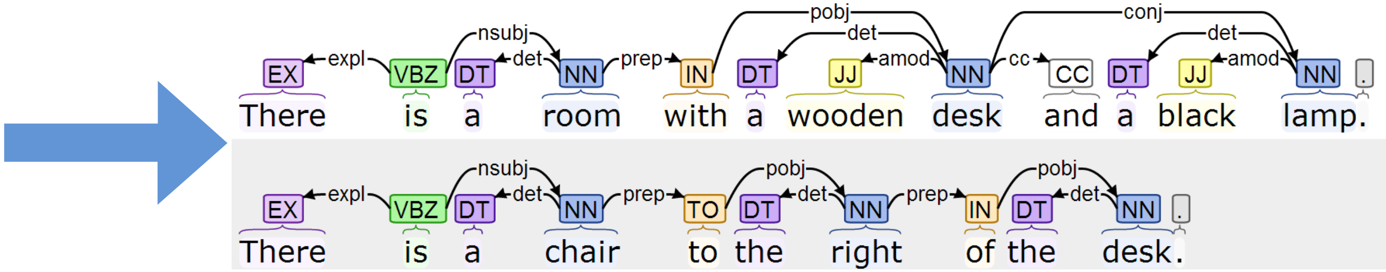


- Identify object categories using noun phrases
- Identify attributes and keywords using modifiers and dependency patterns
- Identify **spatial relations** using **dependency patterns**

(Chang et al., 2014)

Rule-based parsing

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.



There is a room with a wooden desk and a black lamp.
EX VBZ DT NN IN DT JJ NN CC DT JJ NN .
expl nsubj det prep pobj det amod cc conj det amod
There is a chair to the right of the desk.
EX VBZ DT NN TO DT NN IN DT NN .
expl nsubj det prep pobj det prep pobj det

- Identify object categories using noun phrases
- Identify attributes and keywords using modifiers and dependency patterns
- Identify spatial relations using dependency patterns
- Look up objects from DB using **categories** and **keywords**

(Chang et al., 2014)

Parsing + learned lexical grounding

there is a room with
a wooden desk and
a **black lamp**

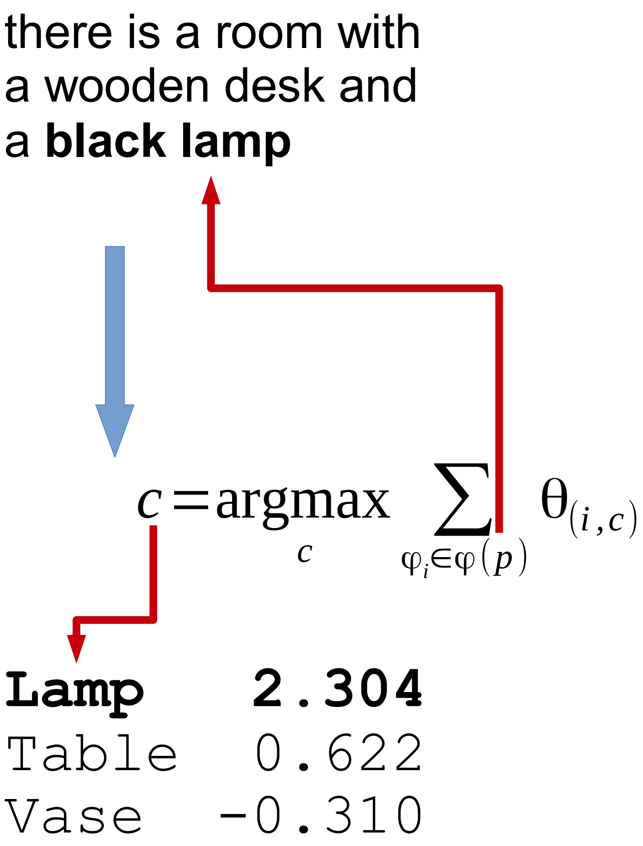
Parsing + learned lexical grounding

there is a room with
a wooden desk and
a **black lamp**

$$c = \underset{c}{\operatorname{argmax}} \sum_{\varphi_i \in \varphi(p)} \theta_{(i,c)}$$

`Lamp`
`Table`
`Vase`

Parsing + learned lexical grounding



there is a room with
a wooden desk and
a **black lamp**

$$c = \operatorname*{argmax}_{c} \sum_{\varphi_i \in \varphi(p)} \theta_{(i,c)}$$

Lamp	**2.304**
Table	0.622
Vase	-0.310

Parsing + learned lexical grounding

there is a room with
a wooden desk and
a **black lamp**

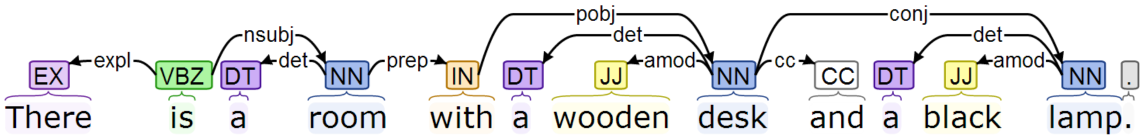
$$c = \operatorname{argmax}_{c} \sum_{\varphi_i \in \varphi(p)} \theta_{(i,c)}$$

$$m = \operatorname{argmax}_{m \in c} \left(\lambda_d \sum_{\varphi_i \in \varphi(d)} \theta_{(i,m)} + \lambda_x \sum_{\varphi_i \in \varphi(x)} \theta_{(i,m)} \right)$$

Lamp	**2.304**
Table	0.622
Vase	-0.310

Parsing + learned lexical grounding

there is a room with a wooden desk and a **black lamp**



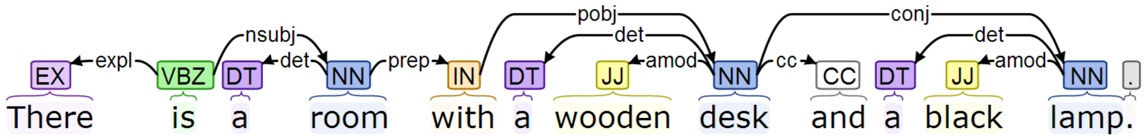
$$c=\underset{c}{\operatorname{argmax}}\sum_{\varphi_i\in\varphi(p)}\theta_{(i,c)}$$

$$m=\underset{m\in c}{\operatorname{argmax}}\left(\lambda_d\sum_{\varphi_i\in\varphi(d)}\theta_{(i,m)}+\lambda_x\sum_{\varphi_i\in\varphi(x)}\theta_{(i,m)}\right)$$

Lamp	**2.304**
Table	0.622
Vase	-0.310

Parsing + learned lexical grounding

there is a room with a wooden desk and a **black lamp**



$$c = \underset{c}{\operatorname{argmax}} \sum_{\varphi_i \in \varphi(p)} \theta_{(i,c)}$$

$$m = \underset{m \in c}{\operatorname{argmax}} \left(\lambda_d \sum_{\varphi_i \in \varphi(d)} \theta_{(i,m)} + \lambda_x \sum_{\varphi_i \in \varphi(x)} \theta_{(i,m)} \right)$$

Lamp	**2.304**
Table	0.622
Vase	-0.310

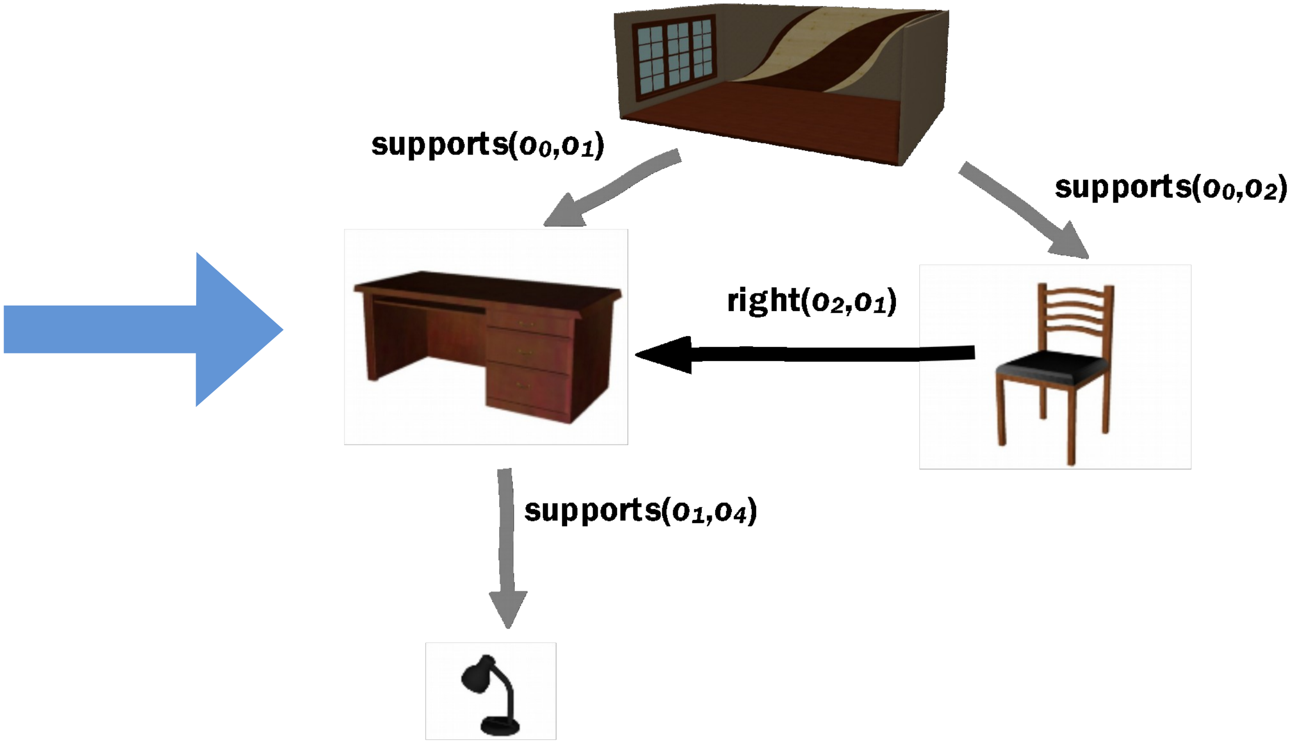
0.302

0.460

-0.021

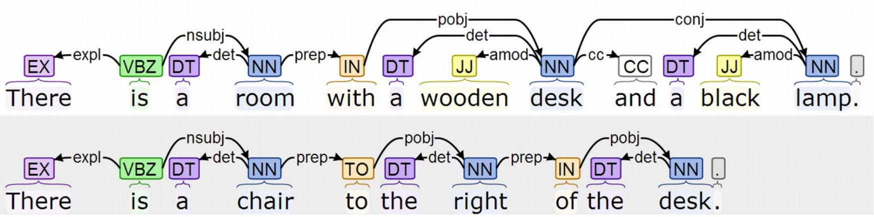
Parsing + learned lexical grounding

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.



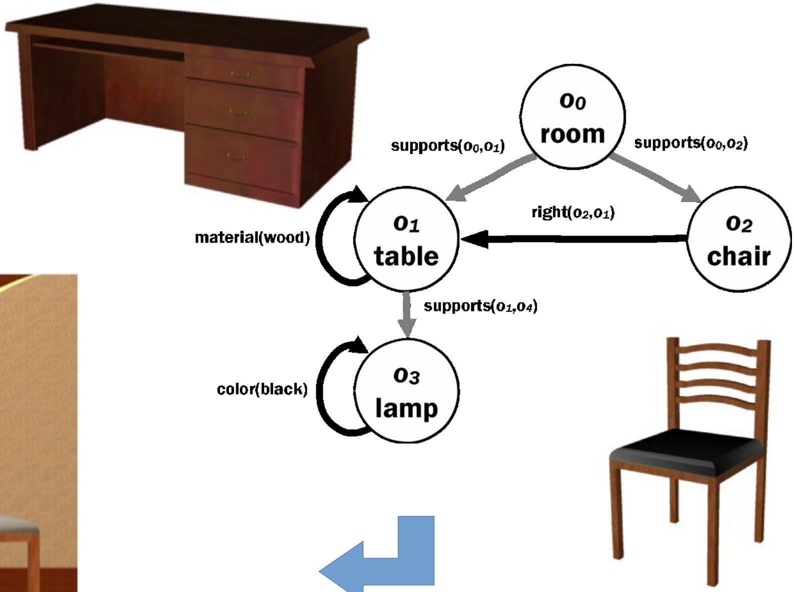
Scene generation pipeline

There is a room with a wooden desk and a black lamp. There is a chair to the right of the desk.

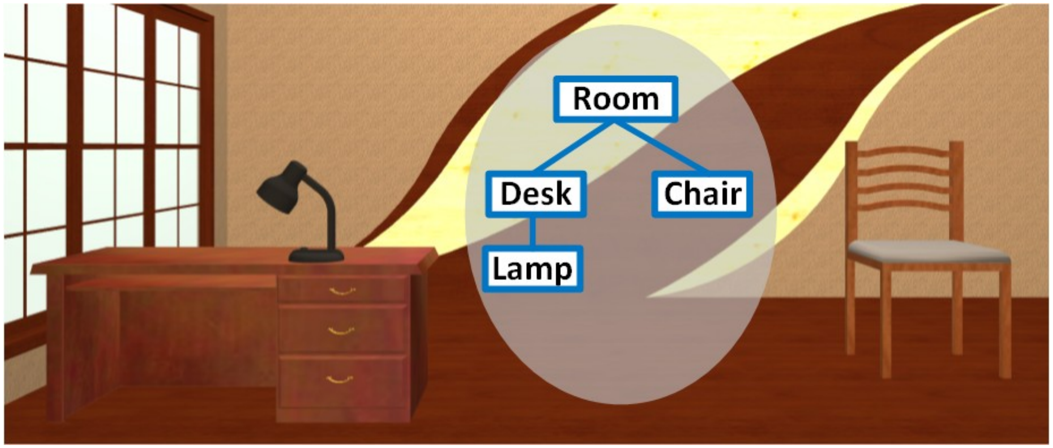

EX VBZ DT NN IN DT JJ NN CC DT JJ NN .
There is a room with a wooden desk and a black lamp.
expl nsubj det prep pobj det amod cc conj det amod
EX VBZ DT NN TO DT NN IN DT NN .
There is a chair to the right of the desk.
expl nsubj det prep pobj det prep pobj det

parsing

object selection



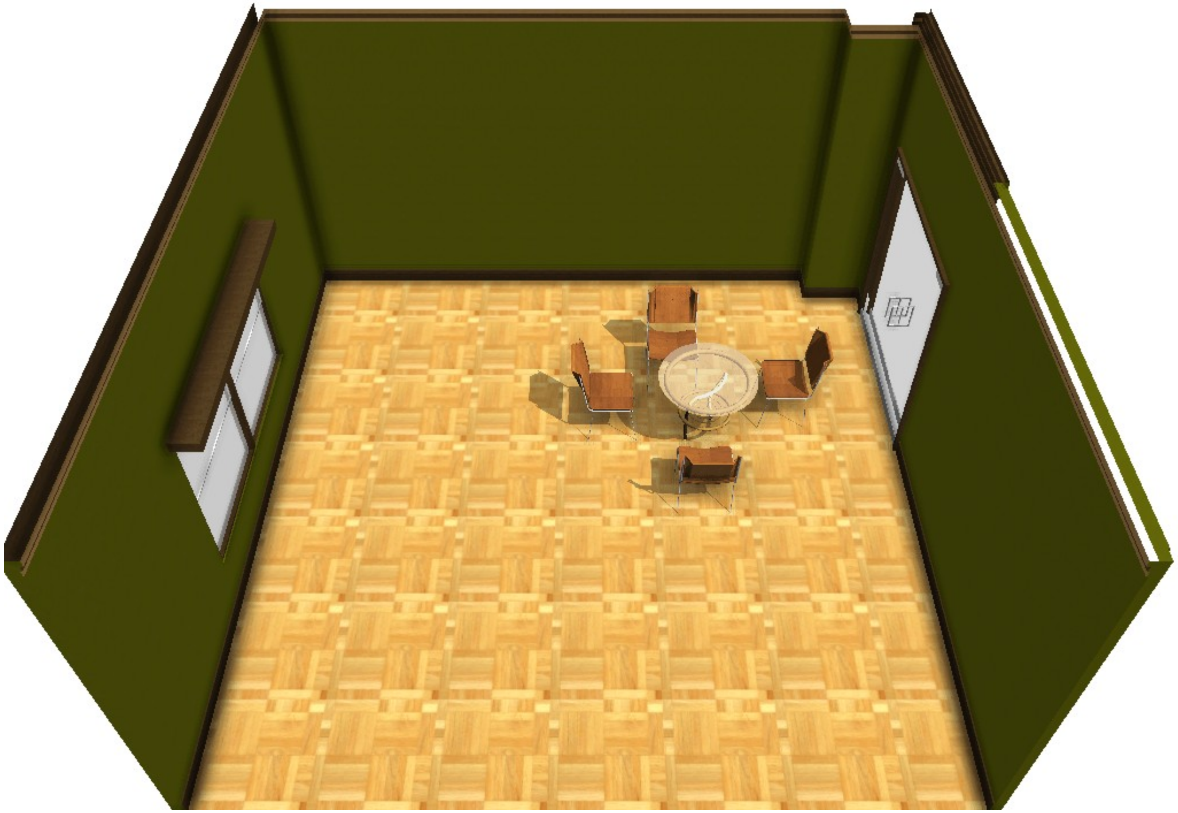
layout



(Chang et al., 2014)

Generated scene examples

A round table is in the center of the room with four chairs around the table. There is a double window facing west. A door is on the east side of the room.



Outline

- Introduction and prior work
- Dataset
- Lexical learning
- Generation with lexical grounding
- Evaluation
- Challenges and conclusion

Evaluation

- Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

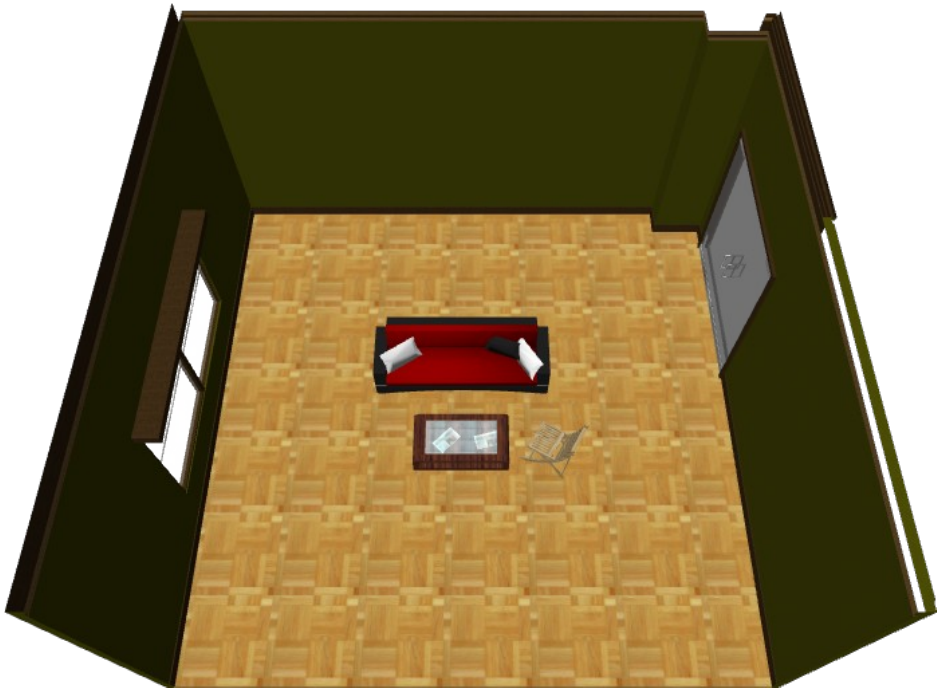
Evaluation

- Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

- Compare scenes generated with four methods against human-built scenes

Evaluation

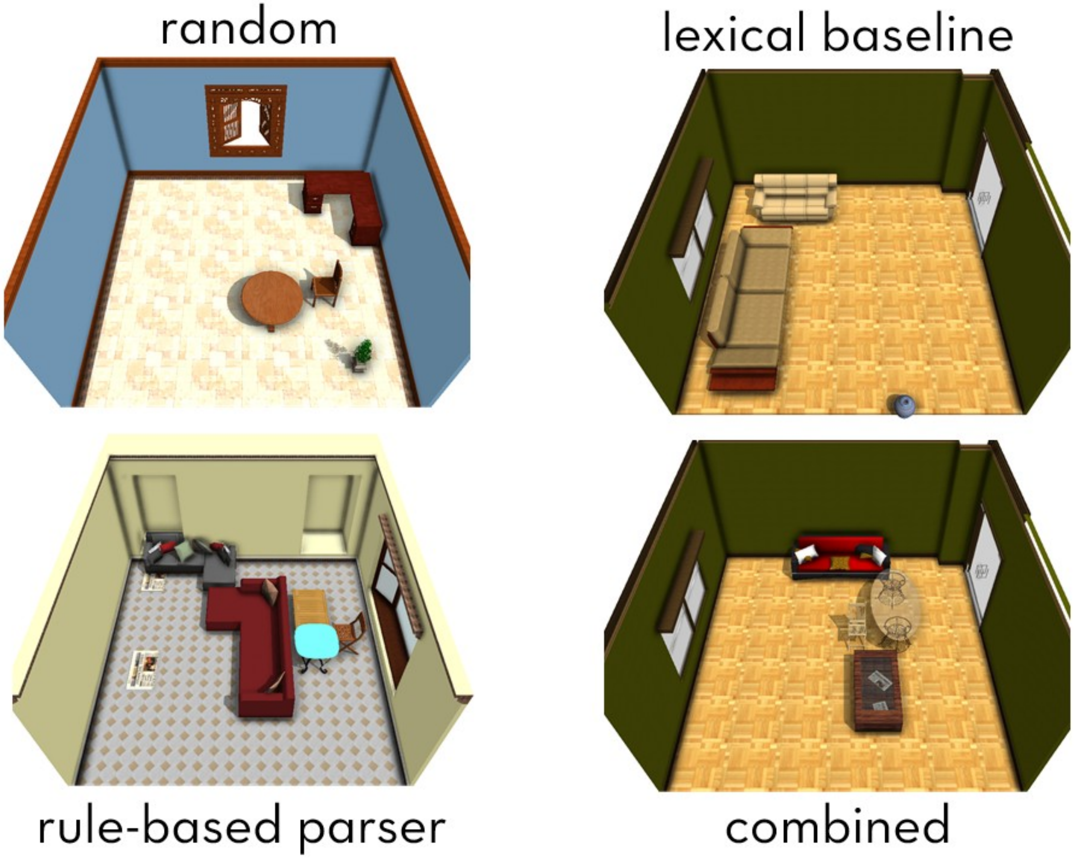
In between the doors and the window, there is a black couch with red cushions, two white pillows, and one black pillow. In front of the couch, there is a wooden coffee table with a glass top and two newspapers. Next to the table, facing the couch, is a wooden folding chair.



human-built

Evaluation

In between the doors and the window, there is a black couch with red cushions, two white pillows, and one black pillow. In front of the couch, there is a wooden coffee table with a glass top and two newspapers. Next to the table, facing the couch, is a wooden folding chair.



random

lexical baseline

rule-based parser

combined

Evaluation

- Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

- Compare scenes generated with 4 methods (*random, lexical baseline, rule-based-parser, combined*) against *human-built* scenes

Evaluation

- Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

- Compare scenes generated with 4 methods (*random, lexical baseline, rule-based-parser, combined*) against *human-built* scenes

- Two sets of scene descriptions
  Seed: seed sentences
  Mturk: descriptions provided by turkers

Dataset

Seed

There is a bed and there is a chair next to the bed.

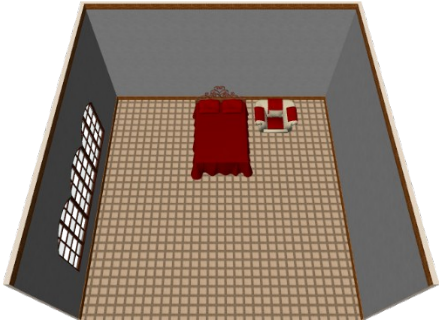
Dataset

Seed

There is a bed and there is a chair next to the bed.

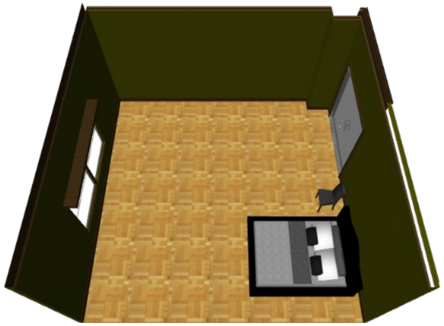
Simple, no modifiers

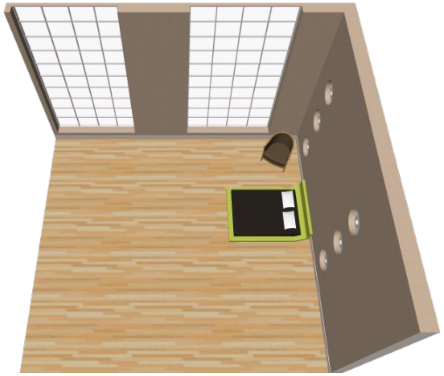
Dataset



Seed

There is a
bed and
there is a
chair next
to the bed.



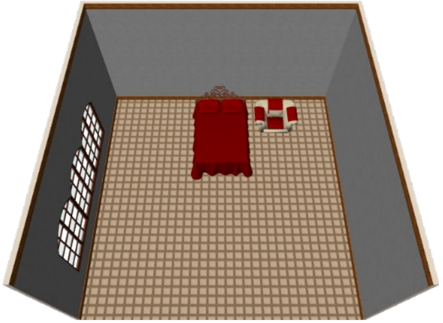


Dataset

Seed

There is a bed and there is a chair next to the bed.

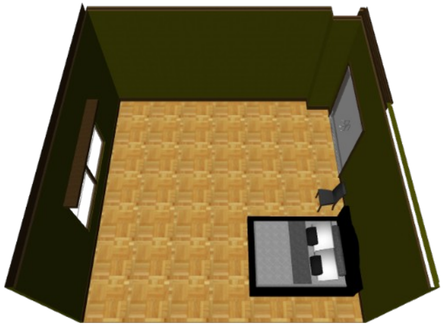
Mturk



The room has three windows on one wall. There is a red bed in the back of the room. Along side the bed is a side chair that is red and white.

This room has a bed with red bedding against the wall. Next to the bed is a chair.

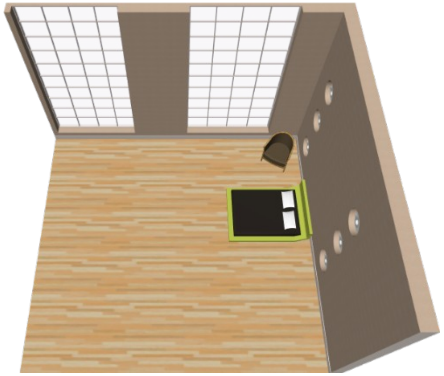
there is a antique looking bed with red covers and pillows in a room. next to it is a recliner chair with red padding. also there are windows.



there is a bed with five pillows on it, and next to it is a chair

There is a bed in the room with two pillows and a small chair near to the right side of it.

There is a large grey bed in the bottom right corner of the room. Above the bed is a small black chair.



Floor to ceiling windows on back wall. Green bed with two pillows and black blanket. Lights recessed into right side wall. Light wood flooring. A chair is in the upper right hand corner

There is a bed on the side of the room. There is a chair in the corner, next to the windows.

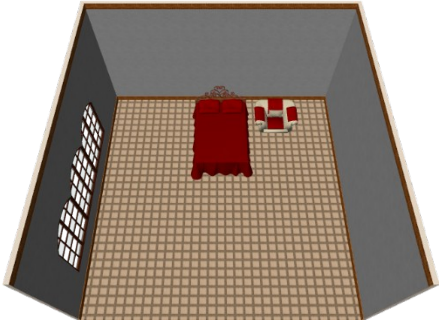
I see a bed and a chair.

Dataset

Seed

There is a bed and there is a chair next to the bed.

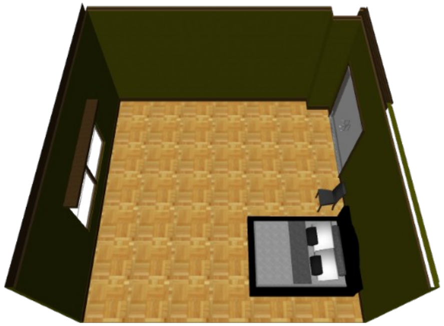
Mturk



The room has three windows on one wall. There is a red bed in the back of the room. Along side the bed is a side chair that is red and white.

This room has a bed with red bedding against the wall. Next to the bed is a chair.

there is a antique looking bed with red covers and pillows in a room. next to it is a recliner chair with red padding. also there are windows.

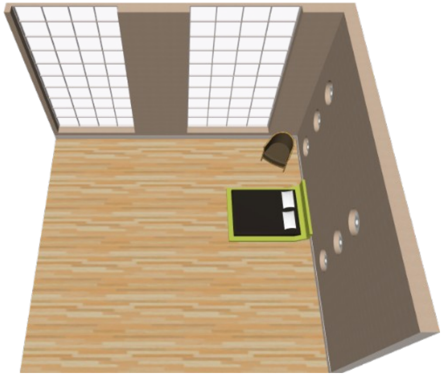


ther hair

The and

The of th

More complex, varied language



Floor to ceiling windows on back wall. Green bed with two pillows and black blanket. Lights recessed into right side wall. Light wood flooring. A chair is in the upper right hand corner

There is a bed on the side of the room. There is a chair in the corner, next to the windows.

I see a bed and a chair.

Evaluation Results

Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

Method
Random
Lexical baseline
Rule-based parser
Combined
Human-built

168 participants, average 4.2 ratings per scene-description pair

Evaluation Results

Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

Method	Seed
Random	2.03
Lexical baseline	3.51
Rule-based parser	
Combined	
Human-built	

168 participants, average 4.2 ratings per scene-description pair

Evaluation Results

Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

Method	Seed
Random	2.03
Lexical baseline	3.51
Rule-based parser	
Combined	
Human-built	

168 participants, average 4.2 ratings per scene-description pair

Evaluation Results

Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

Method	Seed
Random	2.03
Lexical baseline	3.51
Rule-based parser	**5.44**
Combined	
Human-built	6.06

168 participants, average 4.2 ratings per scene-description pair

Evaluation Results

Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

Method	Seed	Mturk
Random	2.03	1.68
Lexical baseline	3.51	2.61
Rule-based parser	**5.44**	→ **3.15**
Combined		
Human-built	6.06	5.87

168 participants, average 4.2 ratings per scene-description pair

Evaluation Results

Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

Method	Seed	Mturk
Random	2.03	1.68
Lexical baseline	3.51	2.61
Rule-based parser	**5.44**	**3.15**
Combined		
Human-built	6.06	5.87

168 participants, average 4.2 ratings per scene-description pair

Evaluation Results

Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

Method	Seed	Mturk
Random	2.03	1.68
Lexical baseline	3.51	2.61
Rule-based parser	**5.44**	3.15
Combined	5.23	**3.73**
Human-built	6.06	5.87

168 participants, average 4.2 ratings per scene-description pair

Evaluation Results

Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

Method	Seed	Mturk
Random	2.03	1.68
Lexical baseline	3.51	2.61
Rule-based parser	**5.44**	3.15
Combined	5.23	**3.73**
Human-built	6.06	5.87

168 participants, average 4.2 ratings per scene-description pair

Outline

- Introduction and prior work
- Dataset
- Lexical learning
- Generation with lexical grounding
- Evaluation
- Challenges and conclusion

Evaluation Results

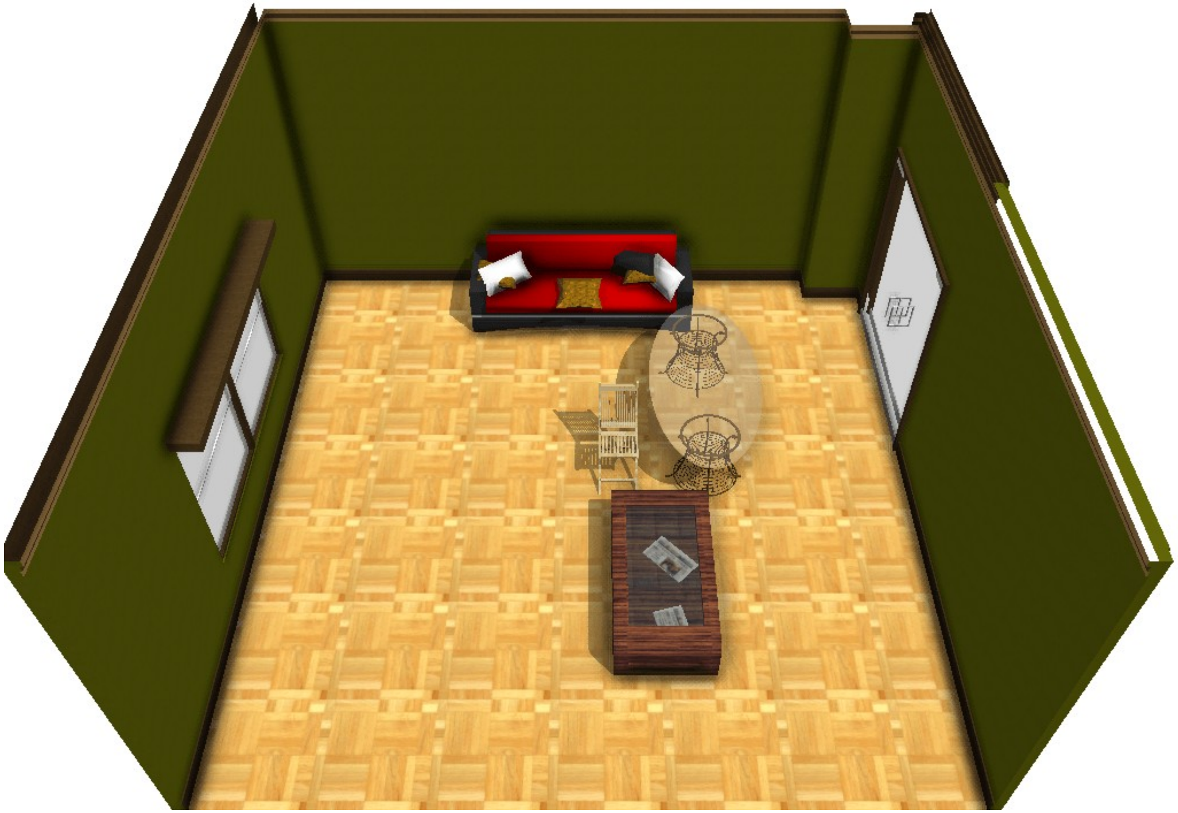
Turkers rated fidelity of generated scenes on a scale of 1 (poor) to 7 (good)

Method	Seed	Mturk
Random	2.03	1.68
Lexical baseline	3.51	2.61
Rule-based parser	**5.44**	3.15
Combined	5.23	**3.73**
Human-built	6.06	5.87

168 participants, average 4.2 ratings per scene-description pair

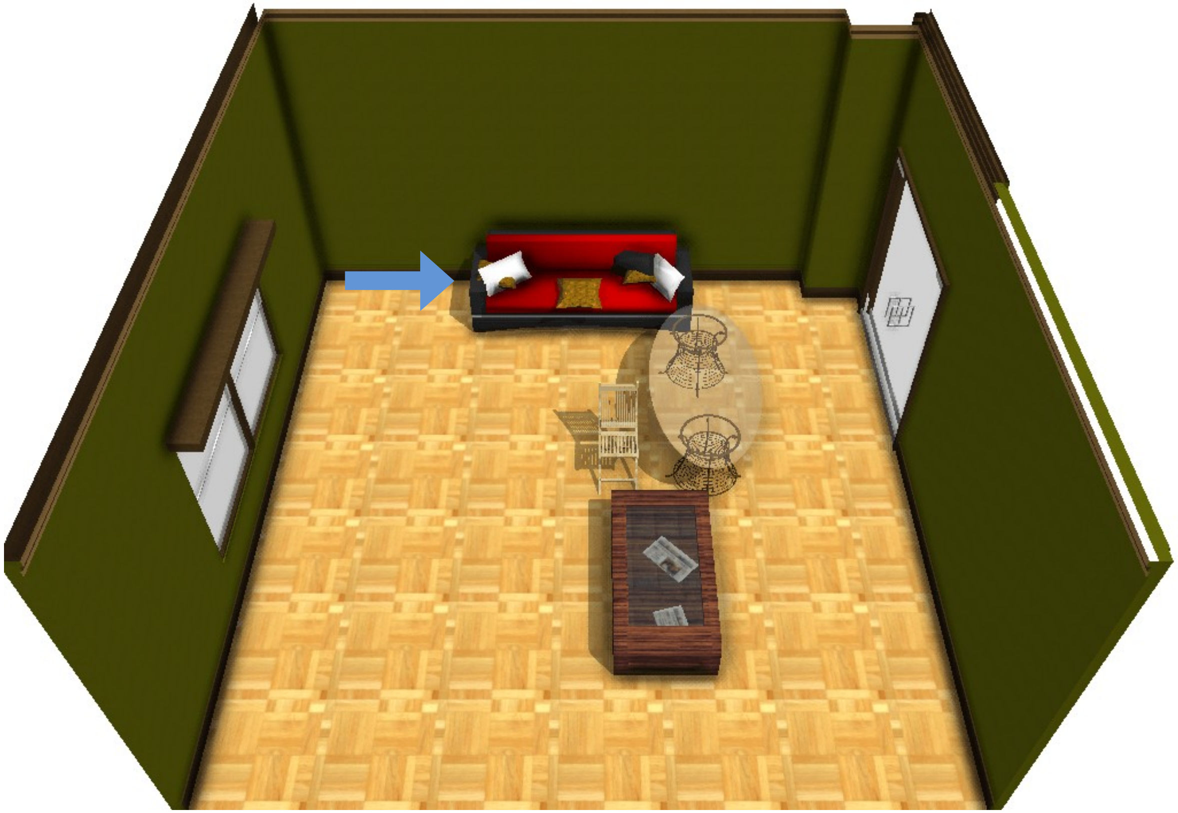
Generated scene examples

In between the doors and the window, there is a black couch with red cushions, two white pillows, and one black pillow. In front of the couch, there is a wooden coffee table with a glass top and two newspapers. Next to the table, facing the couch, is a wooden folding chair.



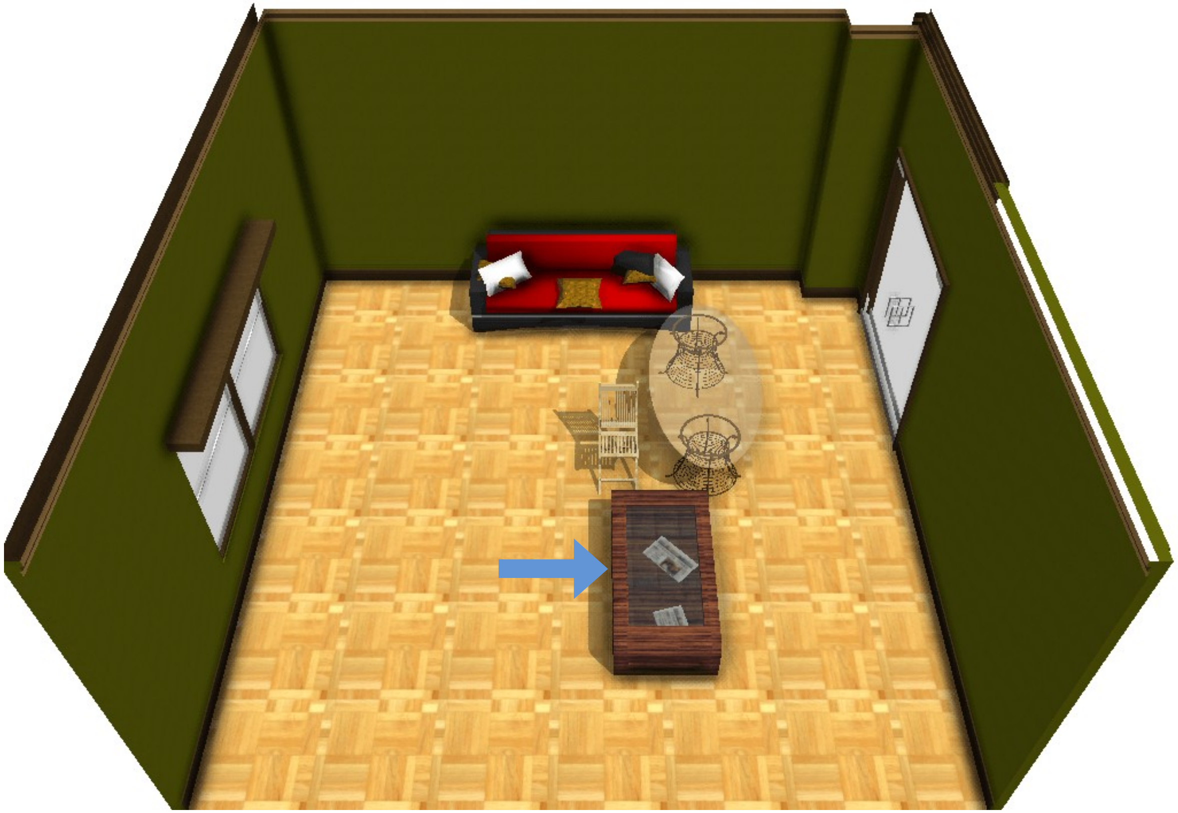
Generated scene examples

In between the doors and the window, there is a black couch with red cushions, two white pillows, and one black pillow. In front of the couch, there is a wooden coffee table with a glass top and two newspapers. Next to the table, facing the couch, is a wooden folding chair.



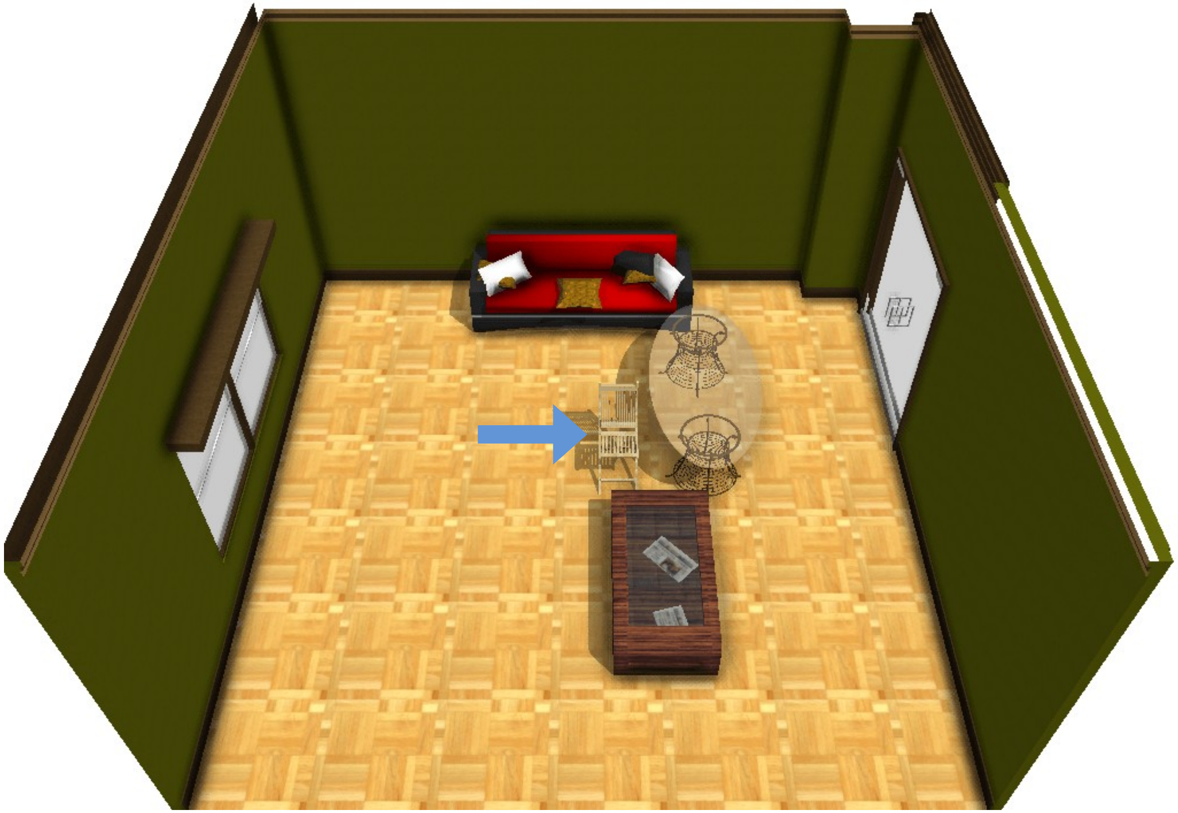
Generated scene examples

In between the doors and the window, there is a black couch with red cushions, two white pillows, and one black pillow. In front of the couch, there is a wooden coffee table with a glass top and two newspapers. Next to the table, facing the couch, is a wooden folding chair.



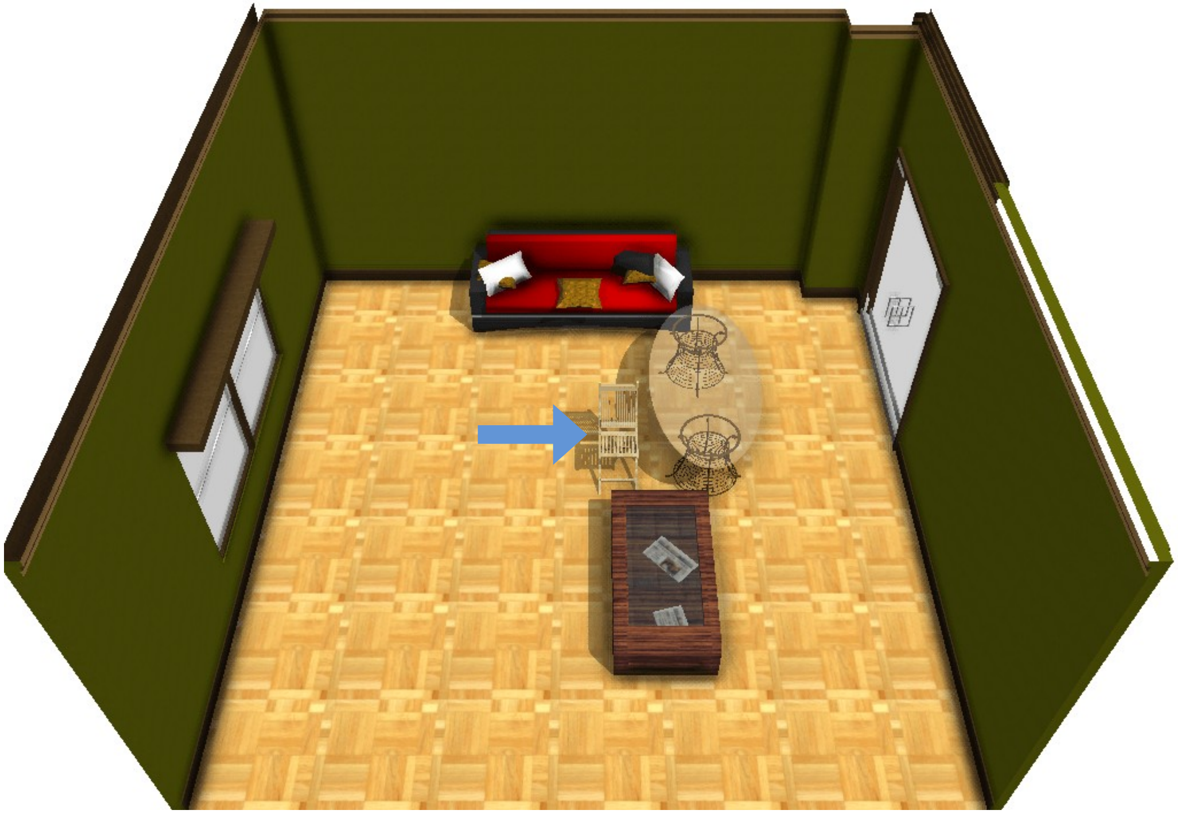
Generated scene examples

In between the doors and the window, there is a black couch with red cushions, two white pillows, and one black pillow. In front of the couch, there is a wooden coffee table with a glass top and two newspapers. Next to the table, facing the couch, is a wooden folding chair.



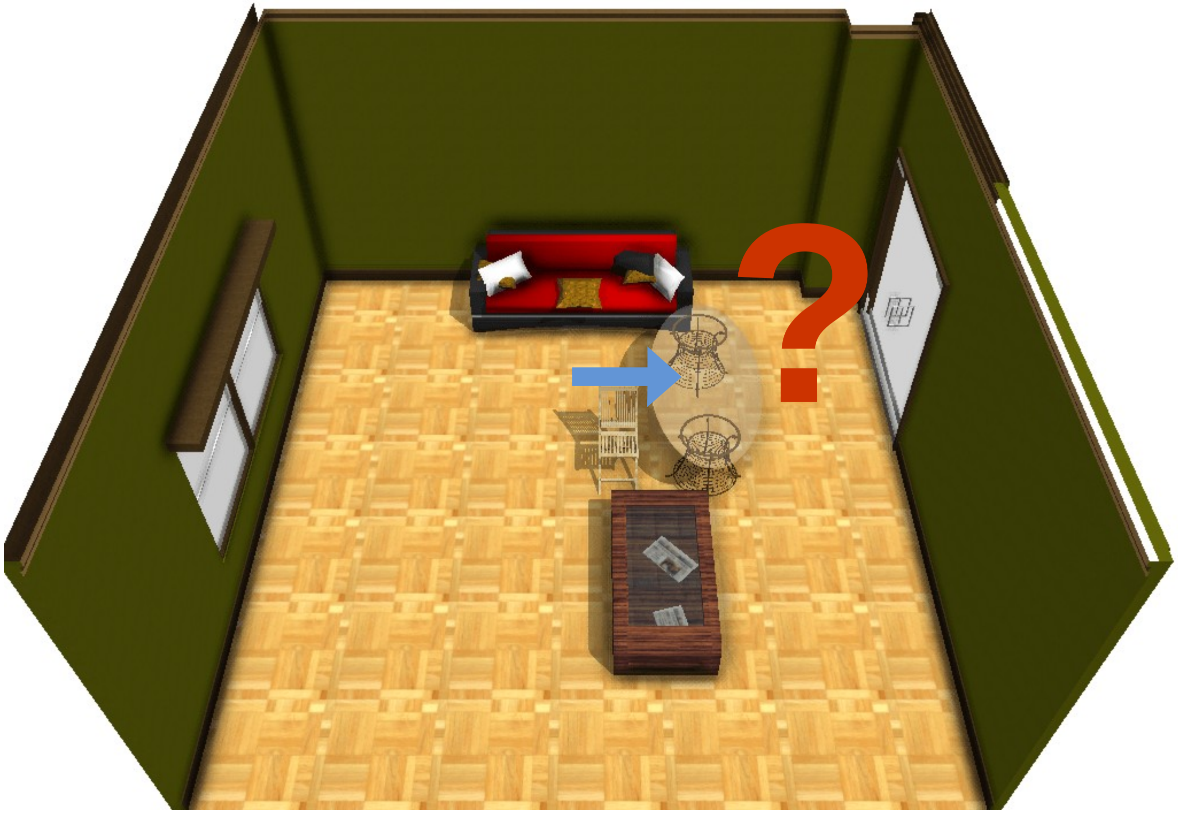
Generated scene examples

In between the doors and the window, there is a black couch with red cushions, two white pillows, and one black pillow. In front of the couch, there is a wooden coffee table with a glass top and two newspapers. Next to the table, facing the couch, is a wooden folding chair.



Generated scene examples

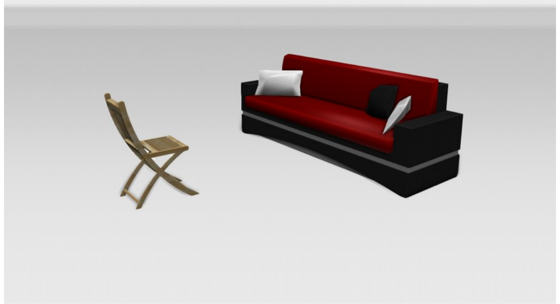
In between the doors and the window, there is a black couch with red cushions, two white pillows, and one black pillow. In front of the couch, there is a wooden coffee table with a glass top and two newspapers. Next to the table, facing the couch, is a wooden folding chair.



Remaining Challenges

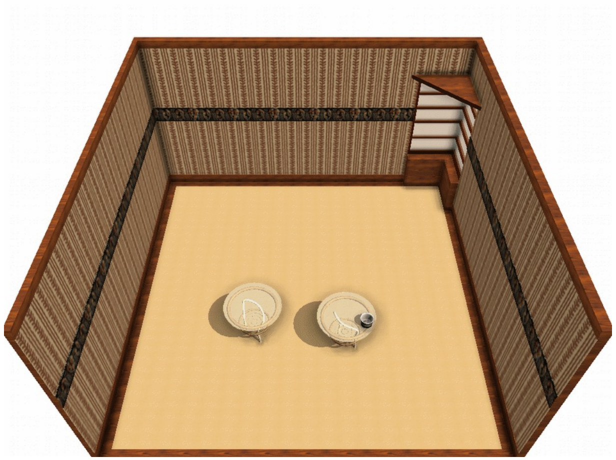
- Grounding of spatial relations

facing the couch



- Coreference

There in the middle is a ***table****.*
On the ***table*** *is a cup.*



Summary

- Learning of lexical grounding to handle linguistic variation in scene description

red cup
round yellow table

Summary

- Learning of lexical grounding to handle linguistic variation in scene description

- Combined rule-based parser and learned lexical groundings for scene generation

Summary

- Learning of lexical grounding to handle linguistic variation in scene description
- Combined rule-based parser and learned lexical groundings for scene generation
- Evaluation demonstrating improved text to scene generation

Thank you!

Dataset is publicly available
http://nlp.stanford.edu/data/text2scene.shtml