Neuro-symbolic Models for Understanding Complex Questions

Jonathan Berant

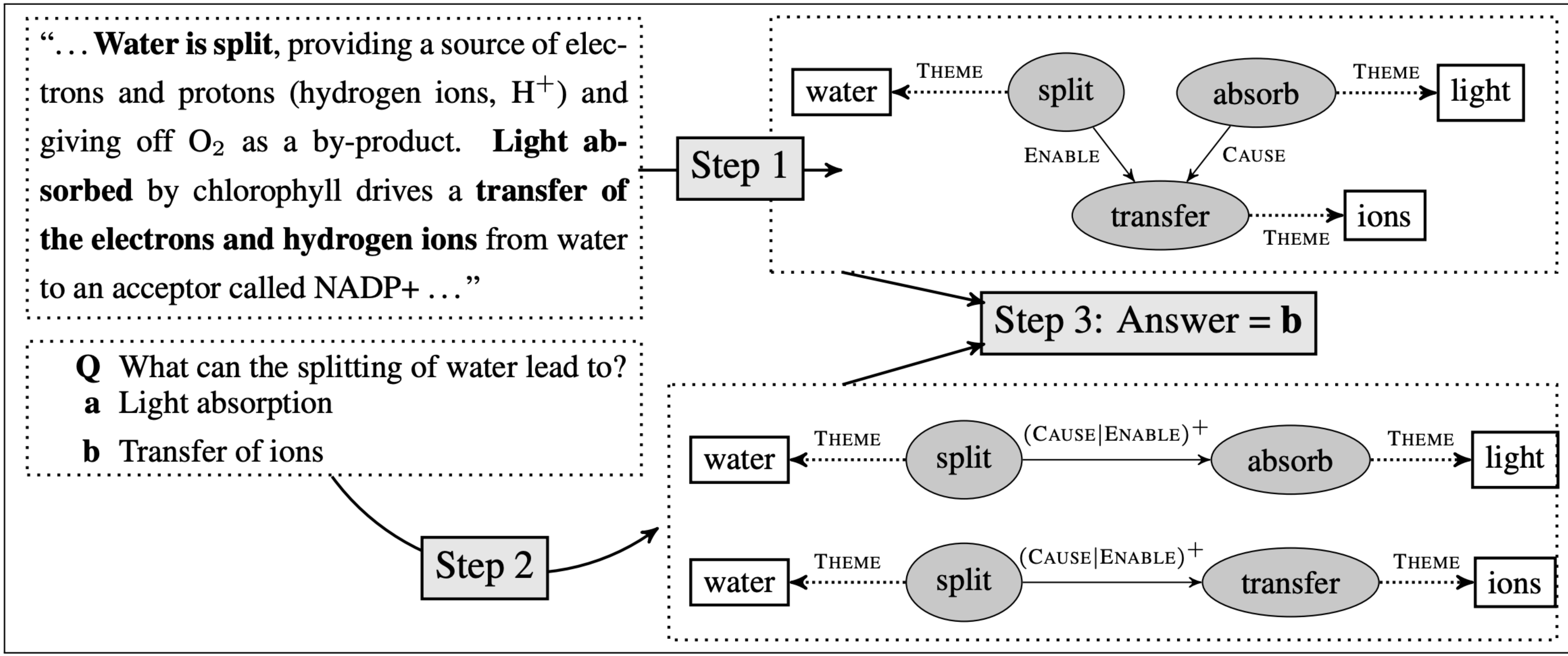
Aug 21, 2021

NSNLI Workshop

TEL AVIV UNIVERSITY אוניברסיטת תל אביב

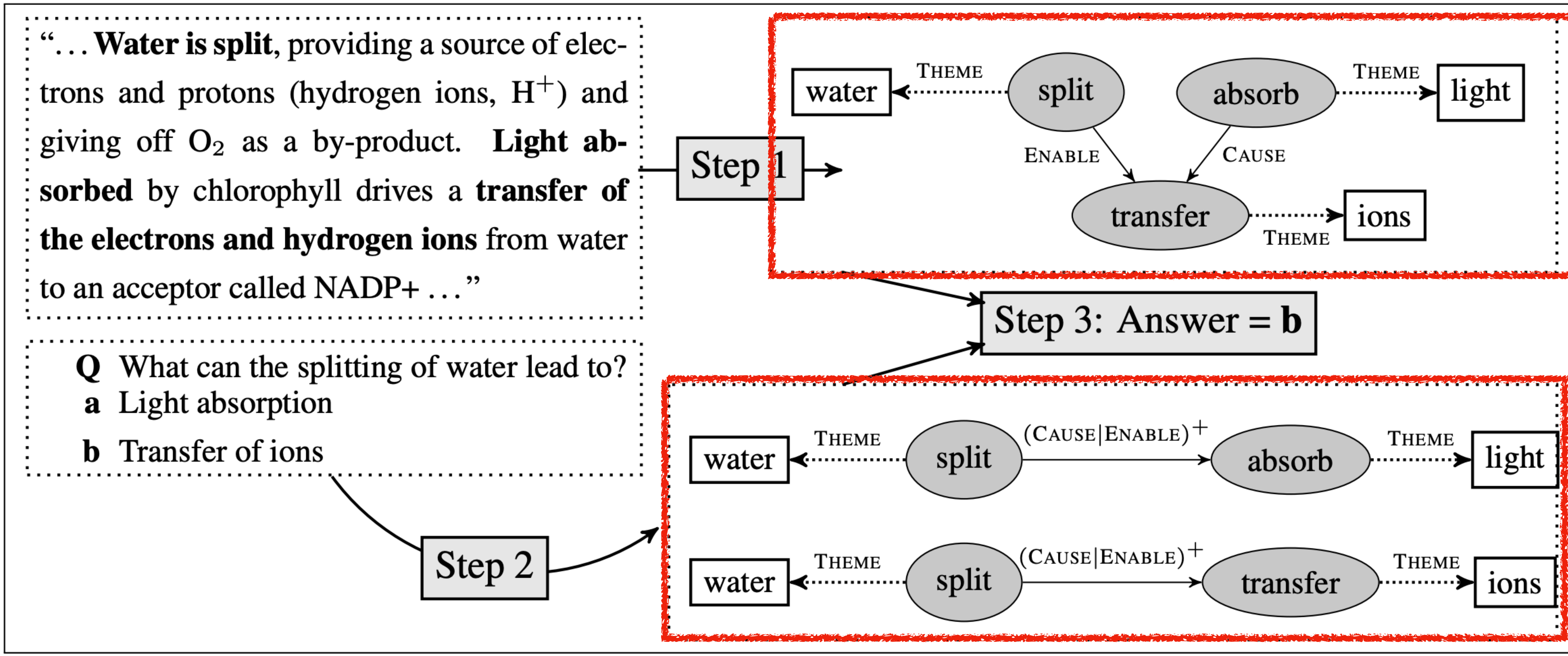
TAU NLP

NLP circa 2014: symbolic



[Berant et al., 2014]

NLP circa 2014: symbolic



[Berant et al., 2014]

NLP in 2021: not very symbolic

Paragraph A, Return to Olympus:
[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

Paragraph B, Mother Love Bone:
[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success.* [8] The album was finally released a few months later.

Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?

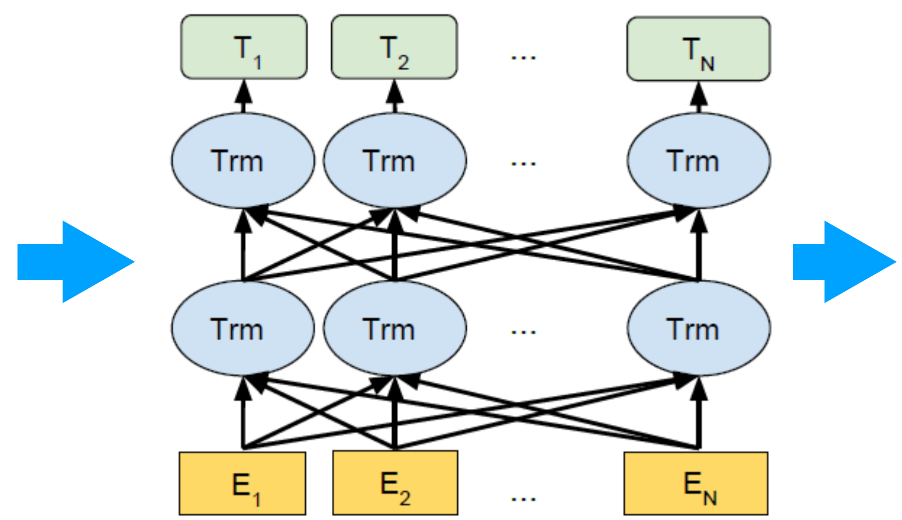
[Yang et al., 2018]

NLP in 2021: not very symbolic

Paragraph A, Return to Olympus:
[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

Paragraph B, Mother Love Bone:
[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success.* [8] The album was finally released a few months later.

Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?



`malfunkshun`

[Yang et al., 2018]

Symbolic vs. Neural Approaches in NLP

- Fully neural approaches have become the de-facto standard:
  - Why?
    - **It works**: training end-to-end differentiable networks with backpropagation (especially given pre-training)
    - **Expressive:** interactions between inputs are learned and not pre-specified
    - **Simple**: No need to design meaning representations

Symbolic vs. Neural Approaches in NLP

- Fully neural approaches have become the de-facto standard:
    - Why?
        - **It works**: training end-to-end differentiable networks with backpropagation (especially given pre-training)
        - **Expressive:** interactions between inputs are learned and not pre-specified
        - **Simple**: No need to design meaning representations

Given enough (x,y) pairs we can use supervised learning on the training set and obtain good i.i.d generalization

Symbolic vs. Neural Approaches in NLP

Given enough (x,y) pairs we can use supervised learning on the training set and obtain good i.i.d generalization

- So what doesn't work?
  - Out-of-distribution generalization: adversarial examples, domain generalization, etc.
  - Few-shot
  - Interpretability
  - …

Plan: discuss in the context of answering complex questions

- Symbolic structures for **compositional generalization**
- Symbolic structures for **evaluating model robustness** through automatic example generation

Generalization

A natural out-of-distribution setup: compositional generalization

winged insect

+

giraffe

=

winged giraffe

A natural out-of-distribution setup: compositional generalization

winged insect

+

giraffe

=

winged giraffe

A natural out-of-distribution setup: compositional generalization

winged insect

+

giraffe

=

winged giraffe

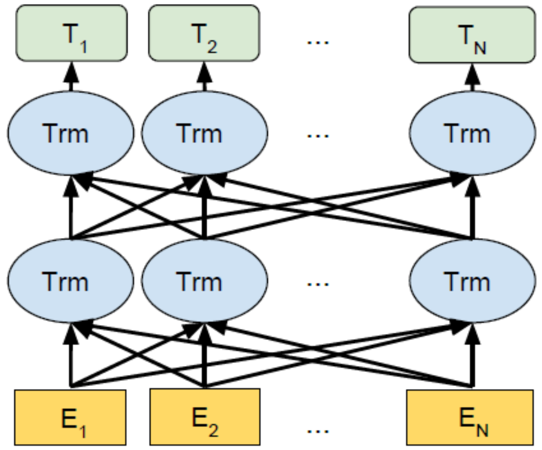
- **Advantages:**
  - Well-defined: all atoms and operations at test time should appear at training time
  - Humans can do it (Fodor and Pylyshyn, 1988)

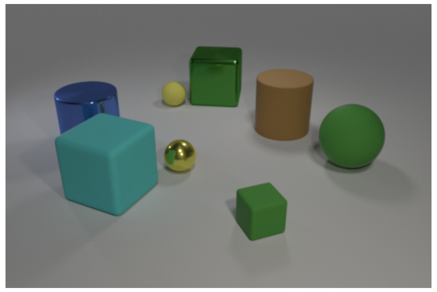
A natural out-of-distribution setup: compositional generalization

- **Advantages:**
  - Well-defined: all atoms and operations at test time should appear at training time
  - Humans can do it (Fodor and Pylyshyn, 1988)

Standard architectures fail at compositional generalization

Improving Text-to-SQL Evaluation Methodology	Finegan-Dollak et al., 2018
Measuring Compositional Generalization	Keysers et al., 2020
SQOOP/SCAN	Lake and Baroni, 2018 / Bahdanau et al., 2019
CLEVR/CLOSURE	Johnson et al., 2017 / Bahdanau et al., 2019





What is the shape of the large thing that is on the right side of the metallic cube?

Standard architectures fail at compositional generalization

Improving Text-to-SQL Evaluation Methodology Finegan-Dollak et al., 2018

Measurin

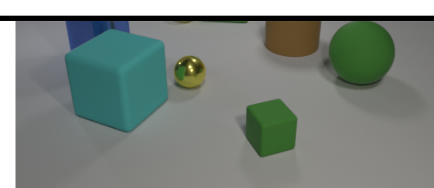
SQOOP/S 2019

CLEVR/CL 019

Measuring Compositional Generalization in X

Homer Simpson
University of Springfield
742 evergreen terrace
`homers@springfield.edu`

- People can compositionally generalize (Fodor and Pylyshyn, 1988).
- we create a bechmark to test whether models can compositionally generalize in X.
- We find out current models do not compositionally generalize in X.



large thing that is on the right side of the metallic cube?

Standard architectures fail at compositional generalization

Improving Text-to-SQL Evaluation Methodology Finegan-Dollak et al., 2018

Measurin

SQOOP/S , 2019

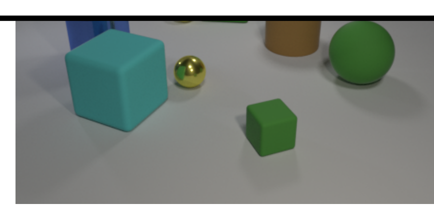
CLEVR/Cl 019

Measuring Compositional Generalization in X

Homer Simpson
University of Springfield
742 evergreen terrace
homers@springfield.edu

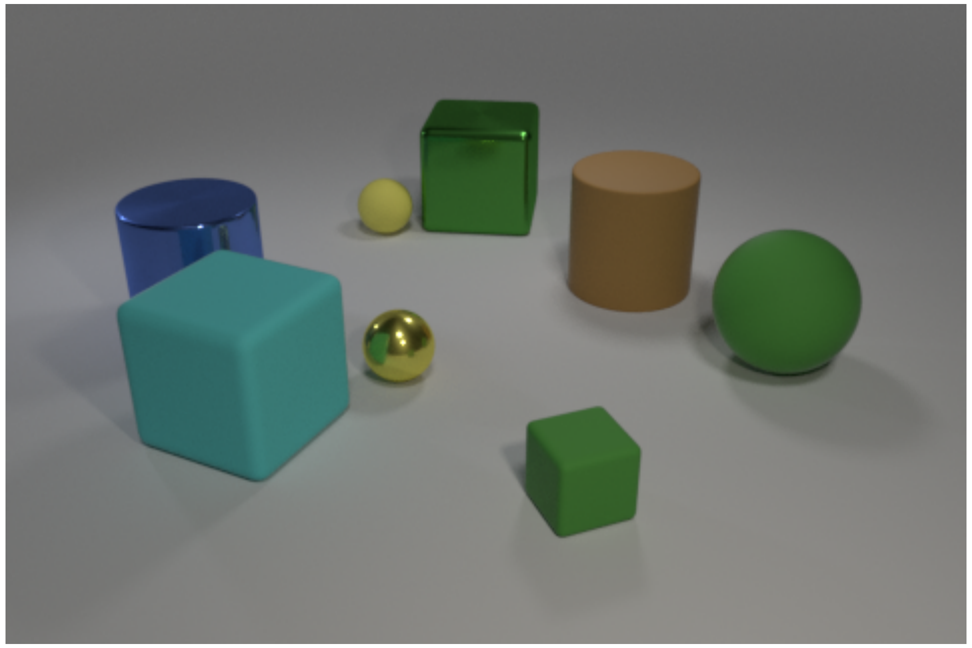
Can we improve compositional generalization with a tree-based model?

- People … and Pylyshyn, 1988).
- we create a bechmark to test whether models can compositionally generalize in X.
- We find out current models do not compositionally generalize in X.

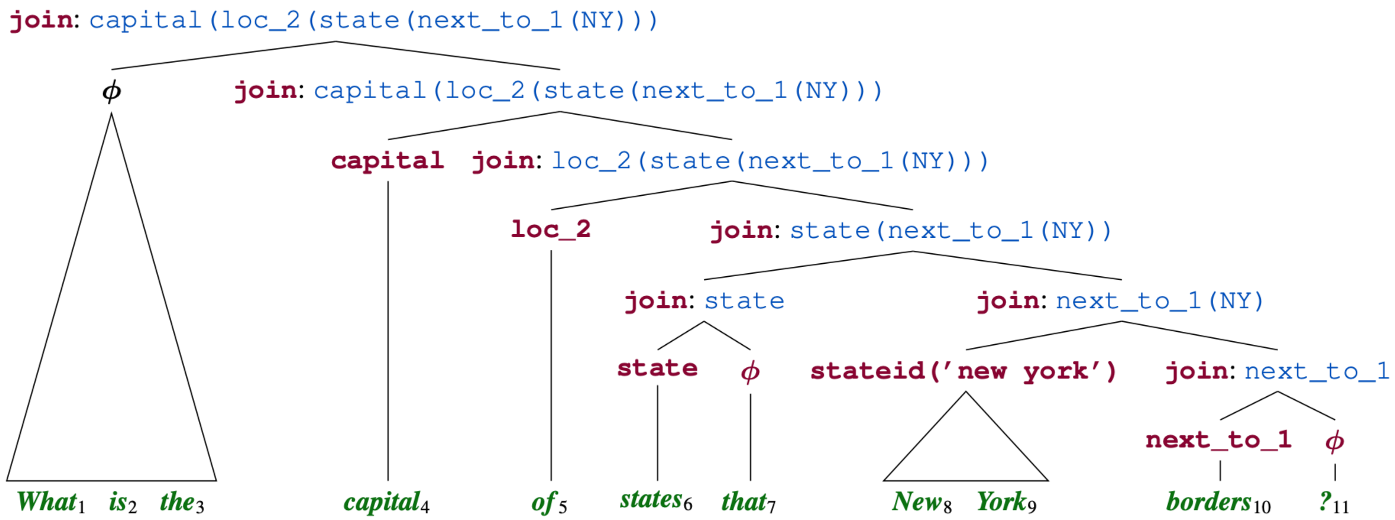


large thing that is on the right side of the metallic cube?

tl;dr: yes



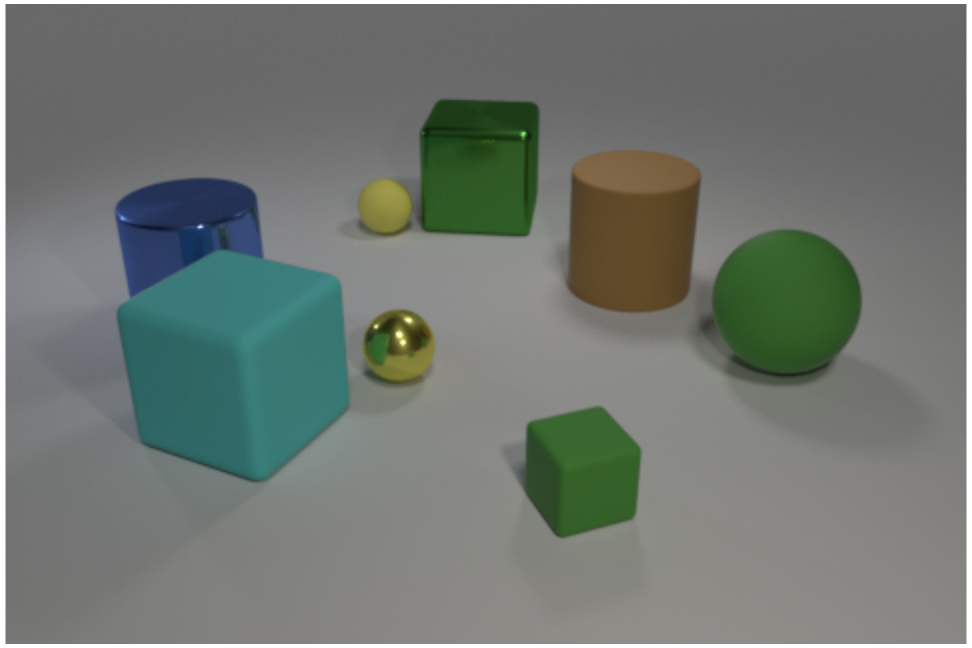
What is the shape of the large thing that is on the right side of the metallic cube and left of the green sphere? [Bogin et al., 2021]



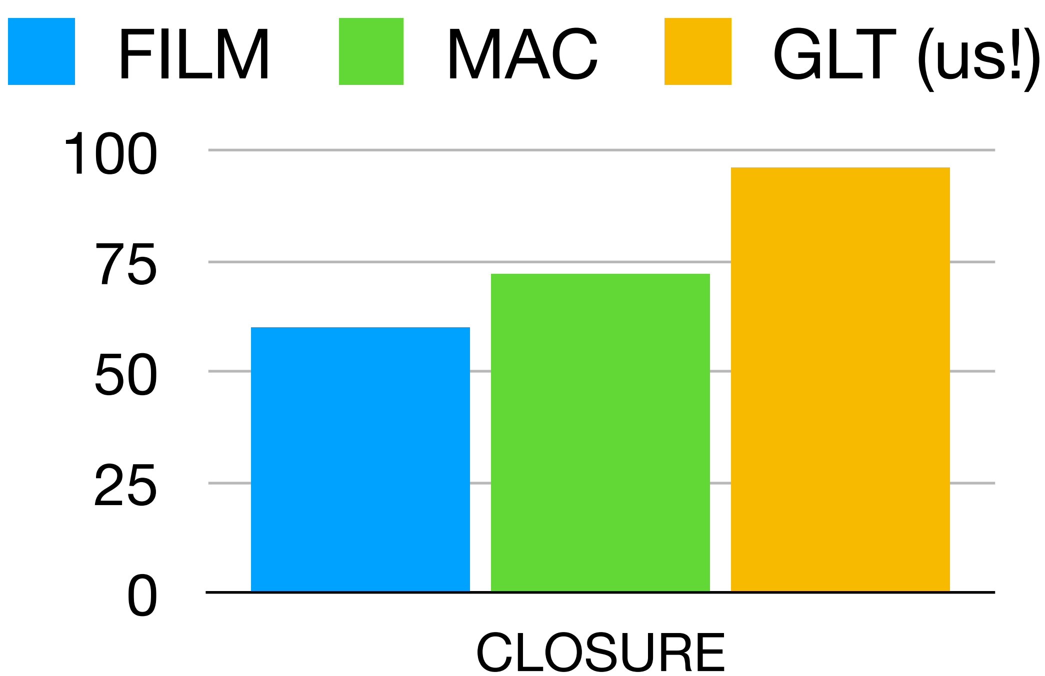
What is the capital of states that New York borders?

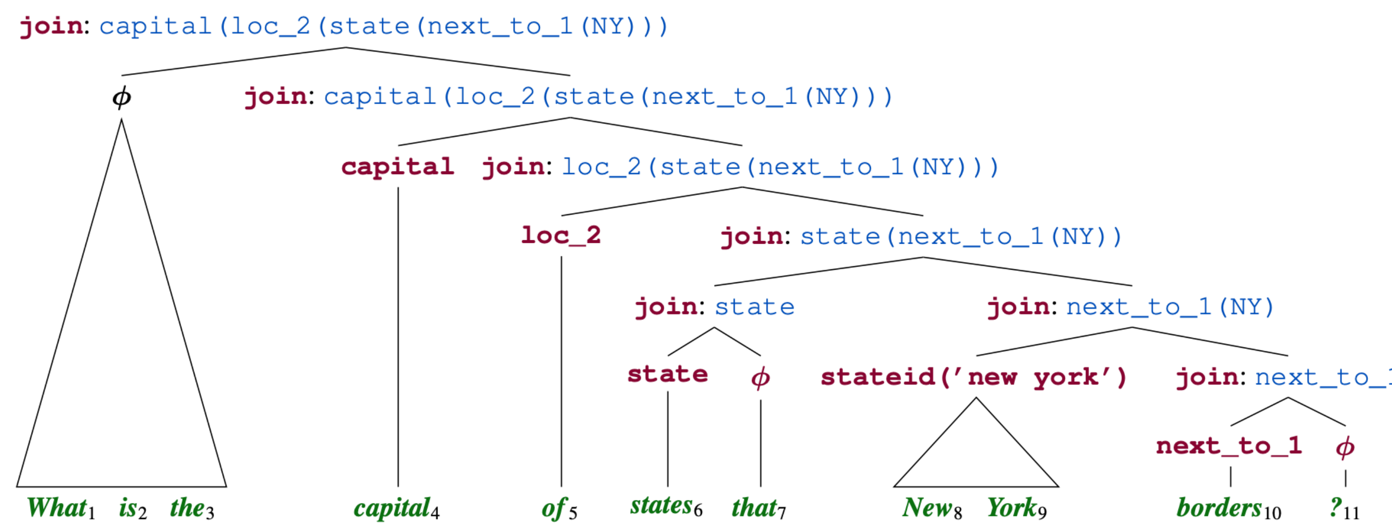
[Herzig and Berant, 2021]

tl;dr: yes



What is the shape of the large thing that is on the right side of the metallic cube and left of the green sphere? [Bogin et al., 2021]

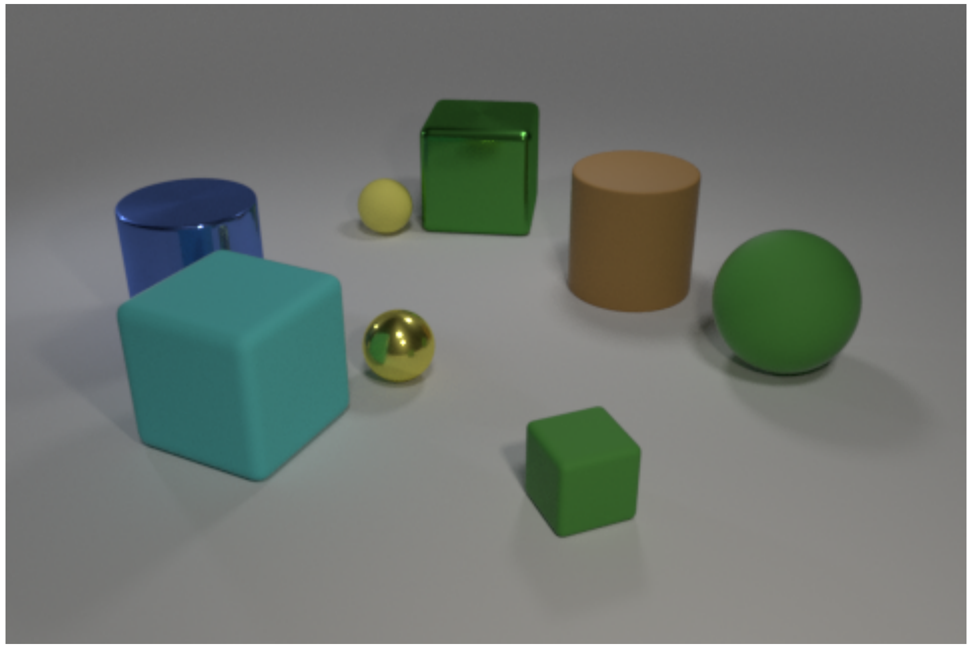




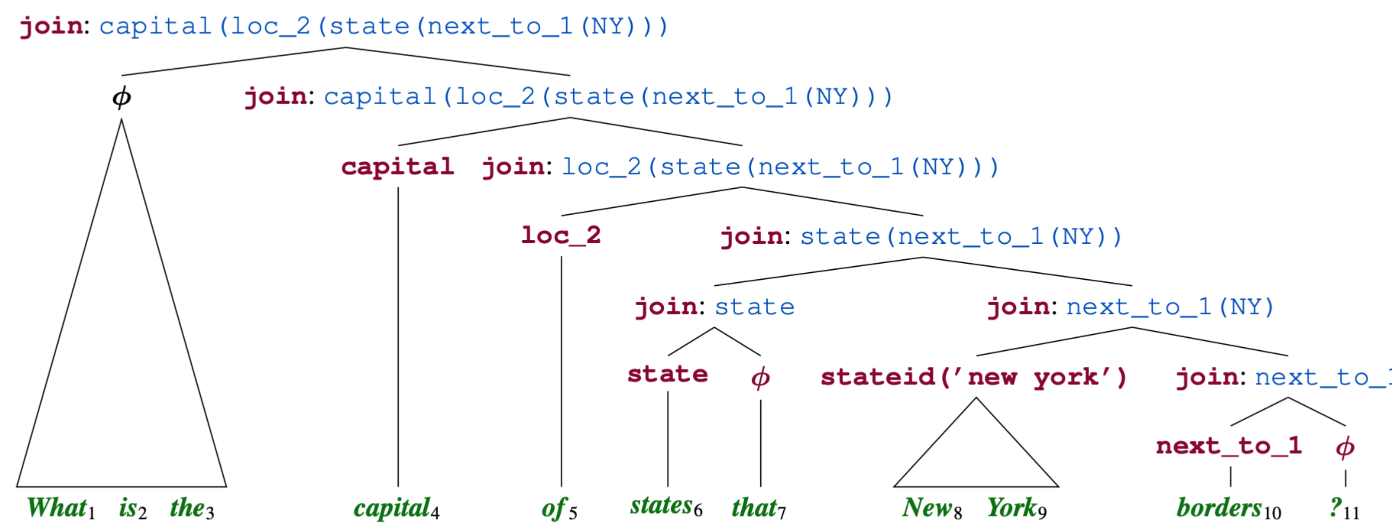
What is the capital of states that New York borders?

[Herzig and Berant, 2021]

tl;dr: yes

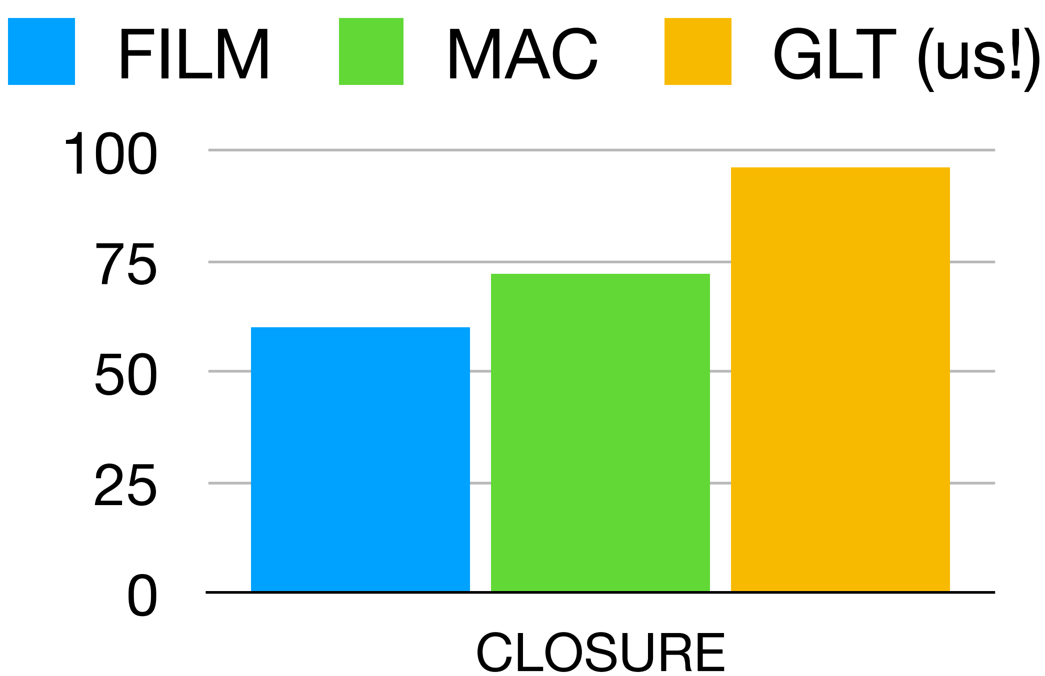


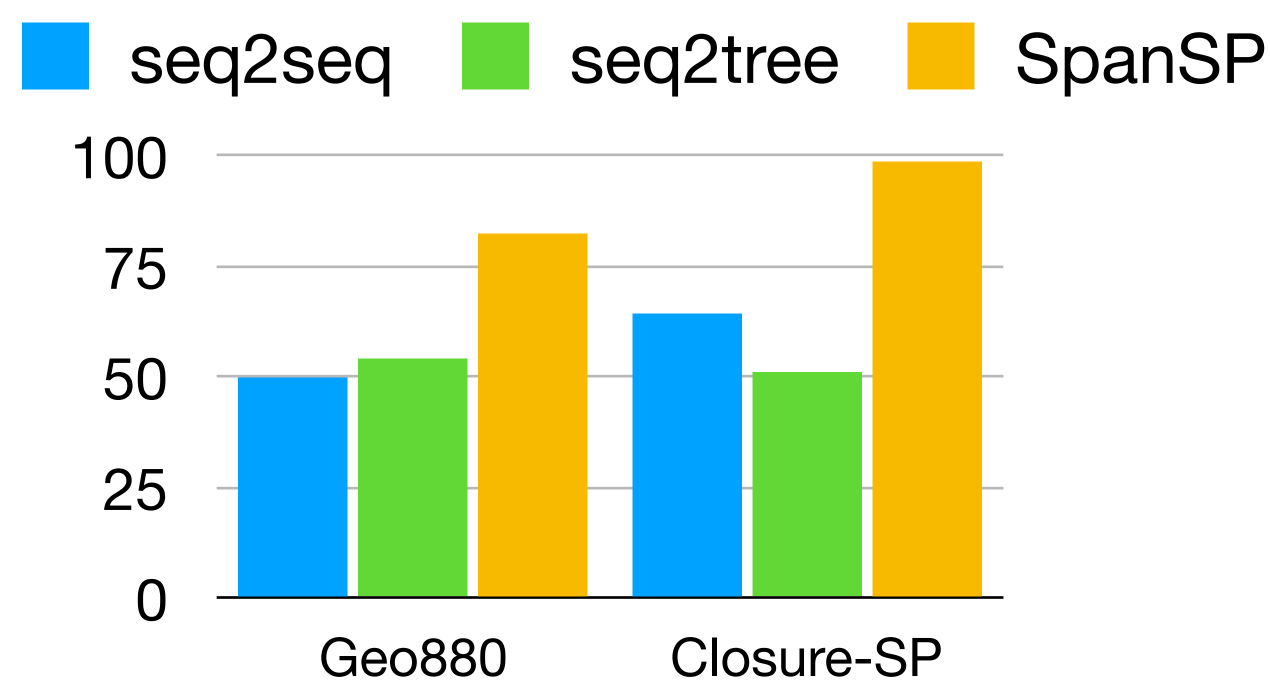
What is the shape of the large thing that is on the right side of the metallic cube and left of the green sphere? [Bogin et al., 2021]



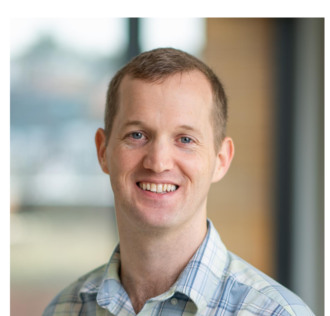
What is the capital of states that New York borders?

[Herzig and Berant, 2021]

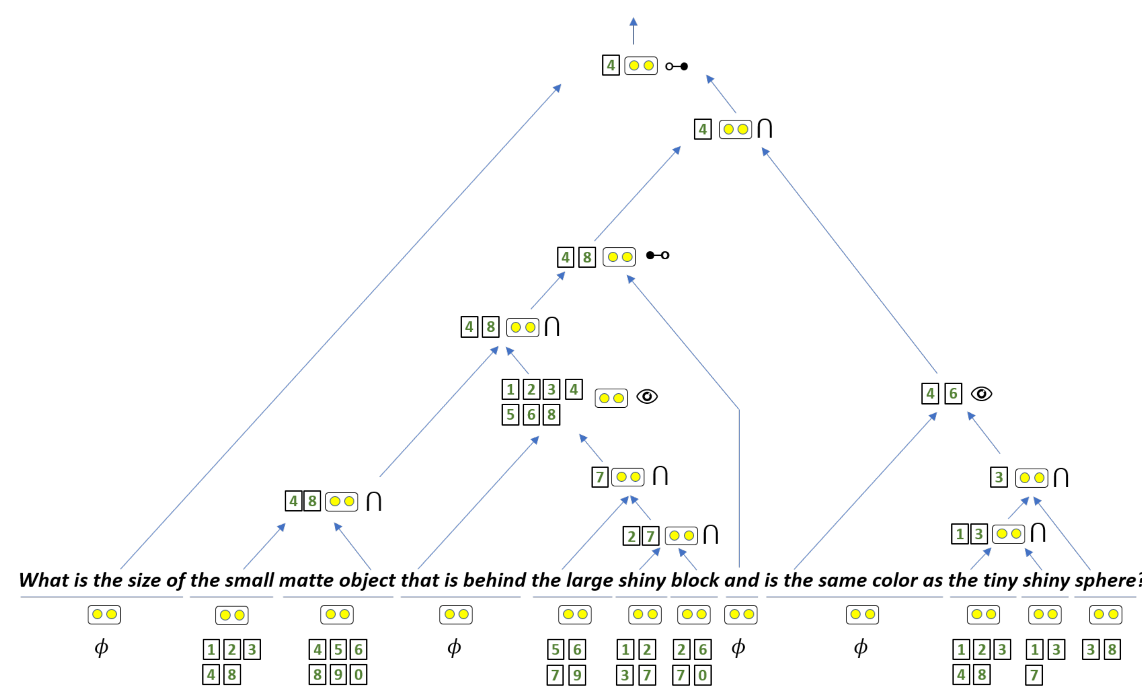




Latent compositional representations improve systematic generalization in grounded question answering

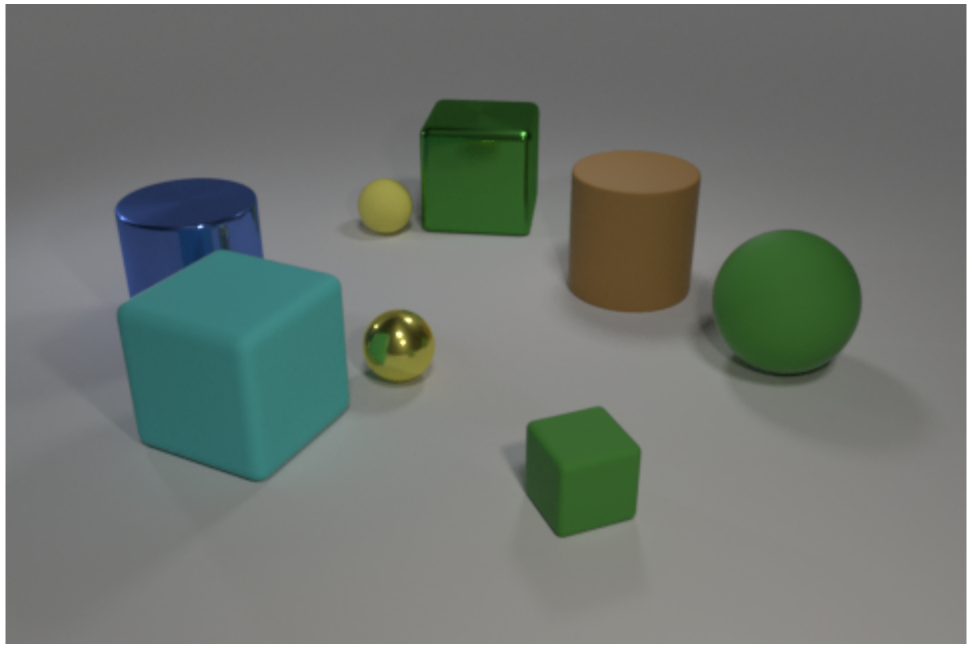
Ben Bogin Sanjay Subramanian Matt Gardner



Setup: visual question answering

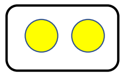
What is the shape of the large thing that is on the right side of the metallic cube and left of the green sphere?

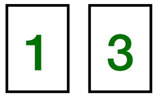
`cylinder`



Only source of the supervision is the final answer

Compositional model (CKY)

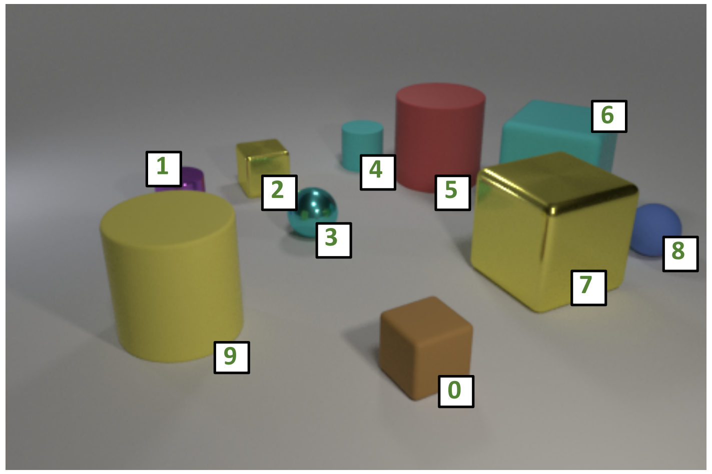




- Compute span **representations** and **denotations** recursively

- **End-to-end differentiable:** learn from downstream supervision

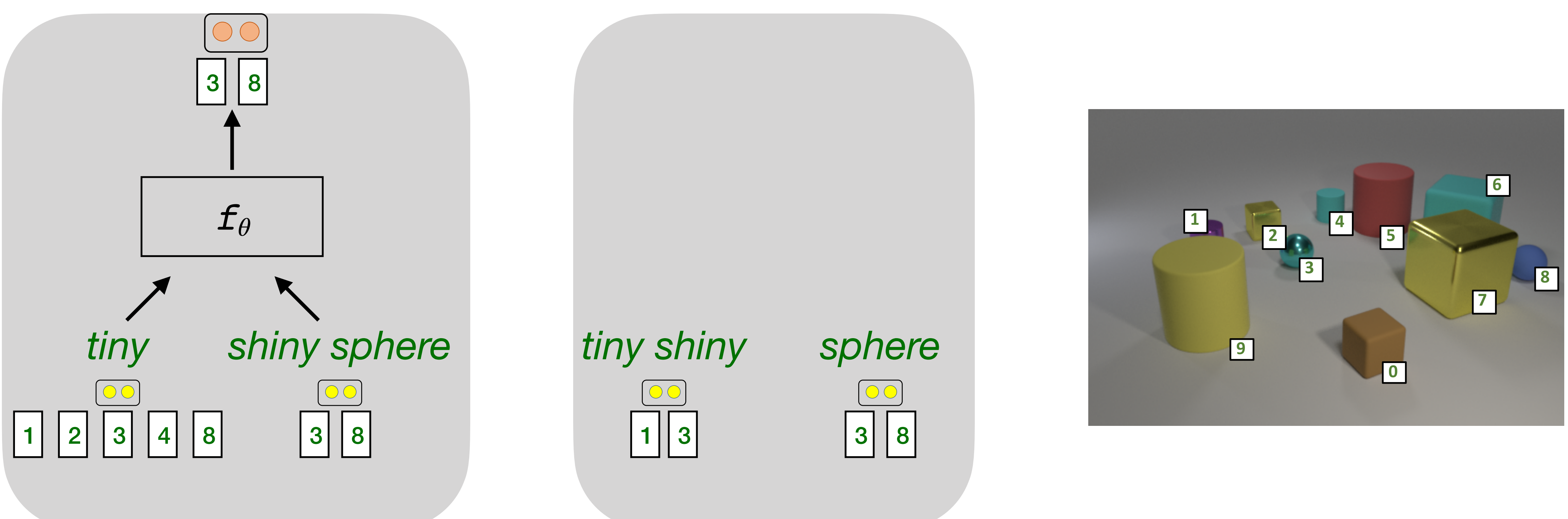
tiny shiny sphere



Compositional model (CKY)

- Compute span **representations** and **denotations** recursively
- **End-to-end differentiable:** learn from downstream supervision

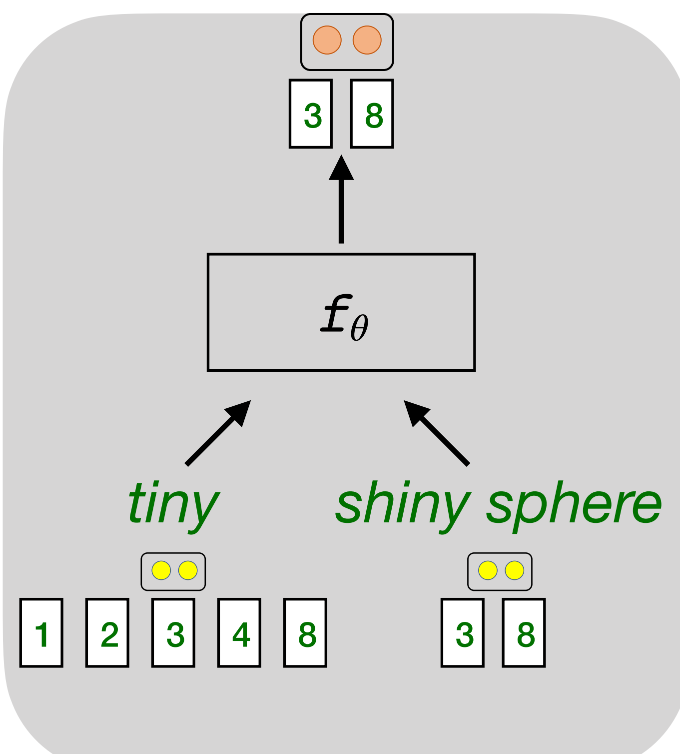
tiny shiny sphere

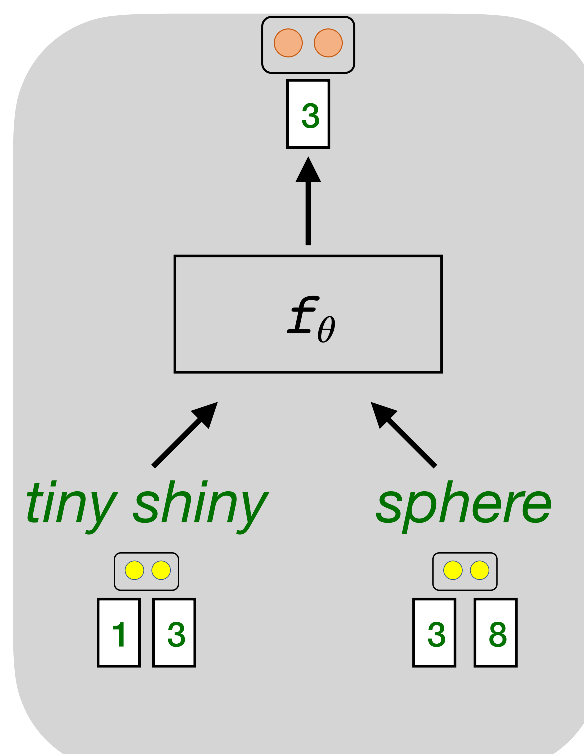


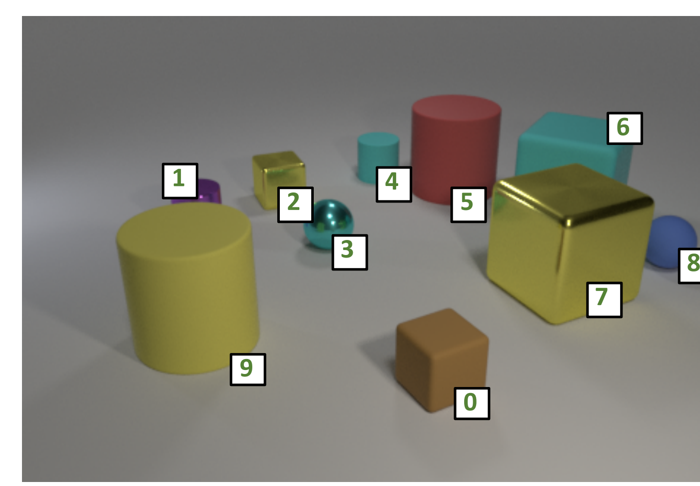
Compositional model (CKY)

- Compute span **representations** and **denotations** recursively
- **End-to-end differentiable:** learn from downstream supervision

tiny shiny sphere



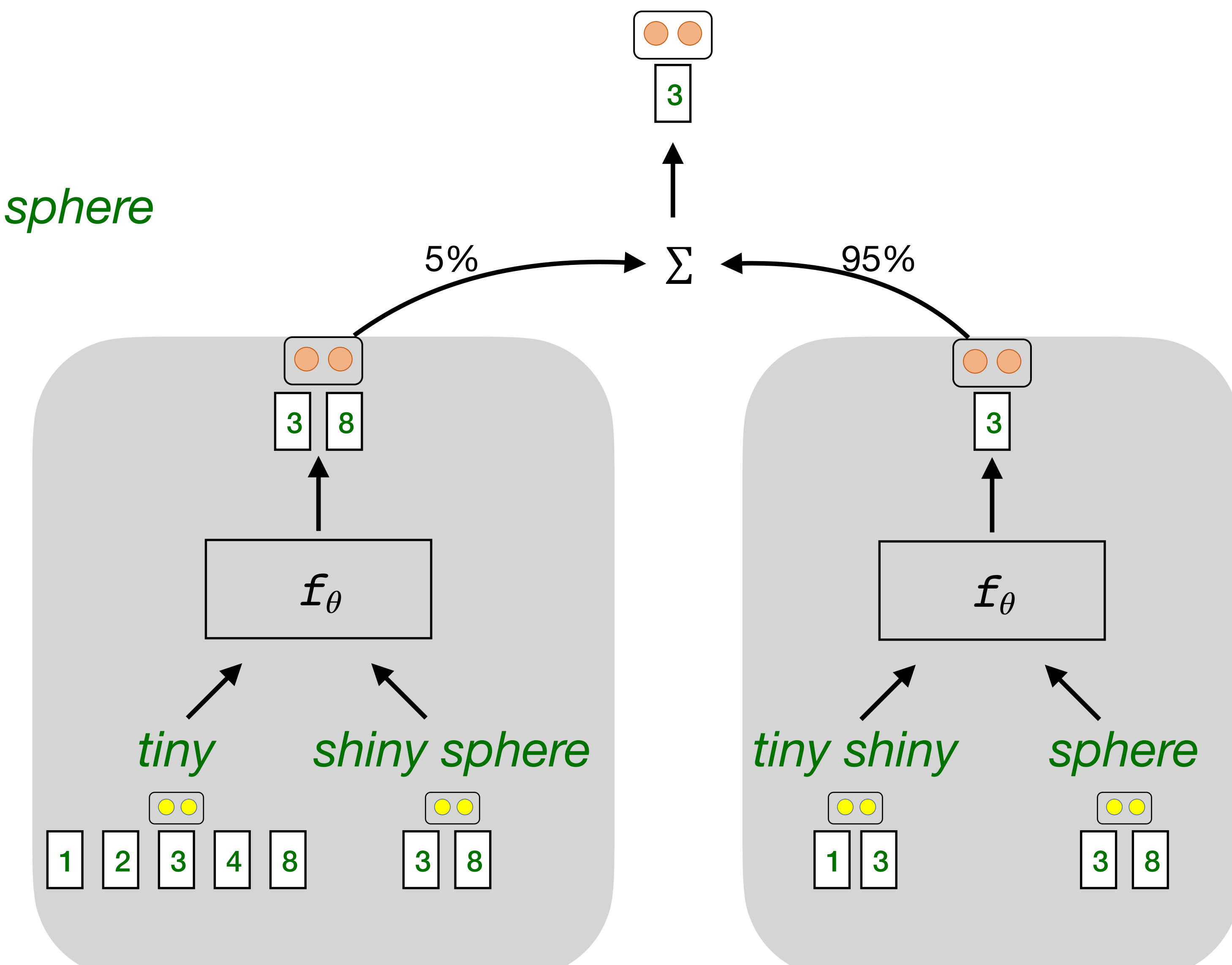


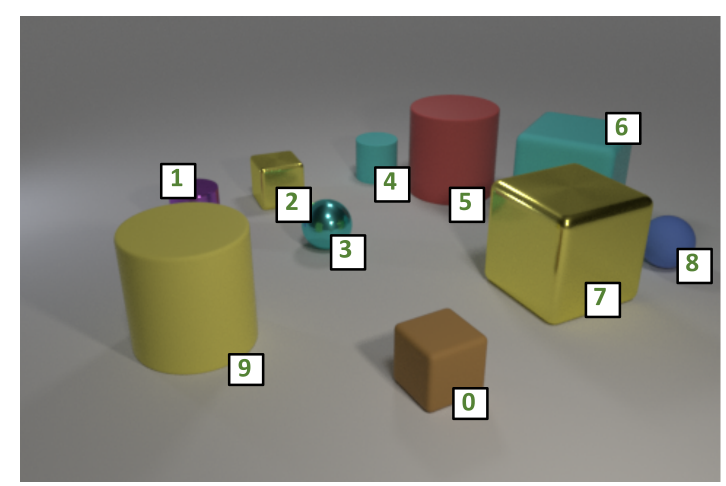


Compositional model (CKY)

- Compute span **representations** and **denotations** recursively
- **End-to-end differentiable:** learn from downstream supervision

tiny shiny sphere





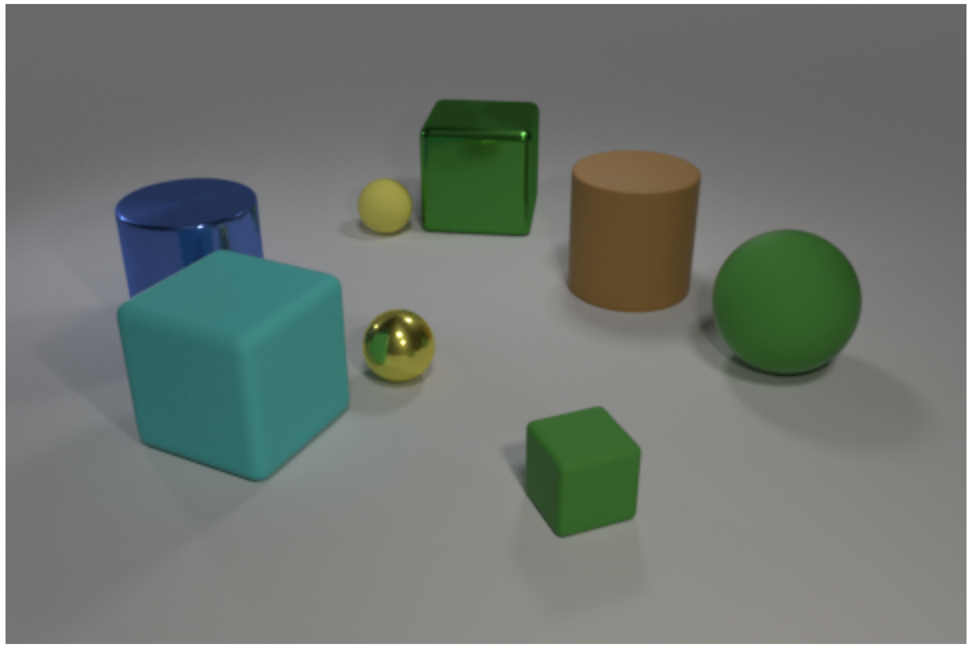
Compositional model

- Compute span **representations** and **denotations** recursively (symbolic)
- **End-to-end differentiable:** learn good representations by training from downstream supervision only (neural)

Compositional model

- Compute span **representations** and **denotations** recursively (symbolic)
- **End-to-end differentiable:** learn good representations by training from downstream supervision only (neural)

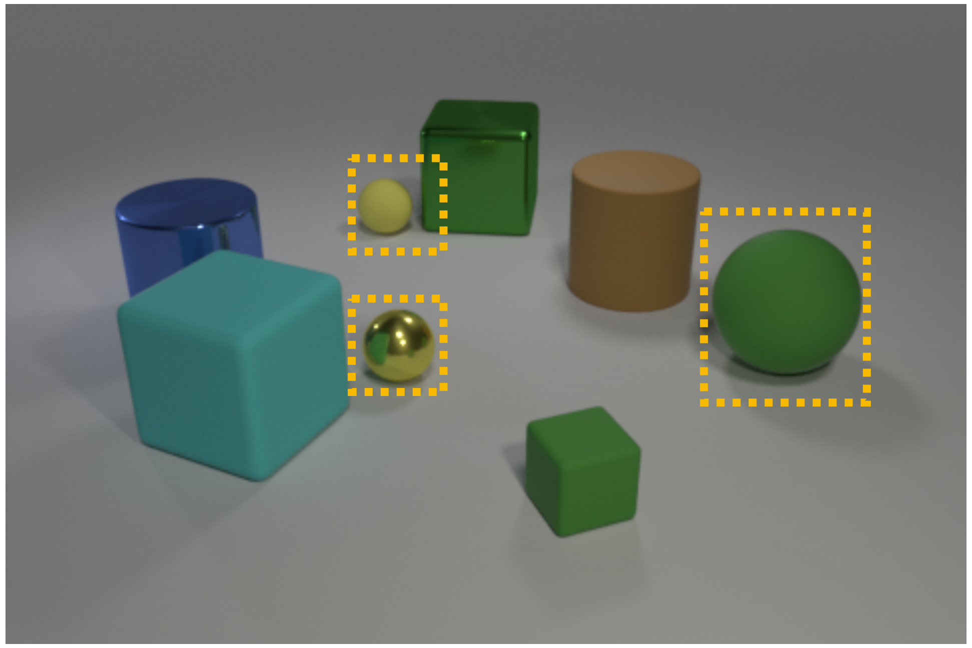
What is the shape of the large thing that is on the right side of the metallic cube and left of the green sphere?



Compositional model

- Compute span **representations** and **denotations** recursively (symbolic)
- **End-to-end differentiable:** learn good representations by training from downstream supervision only (neural)

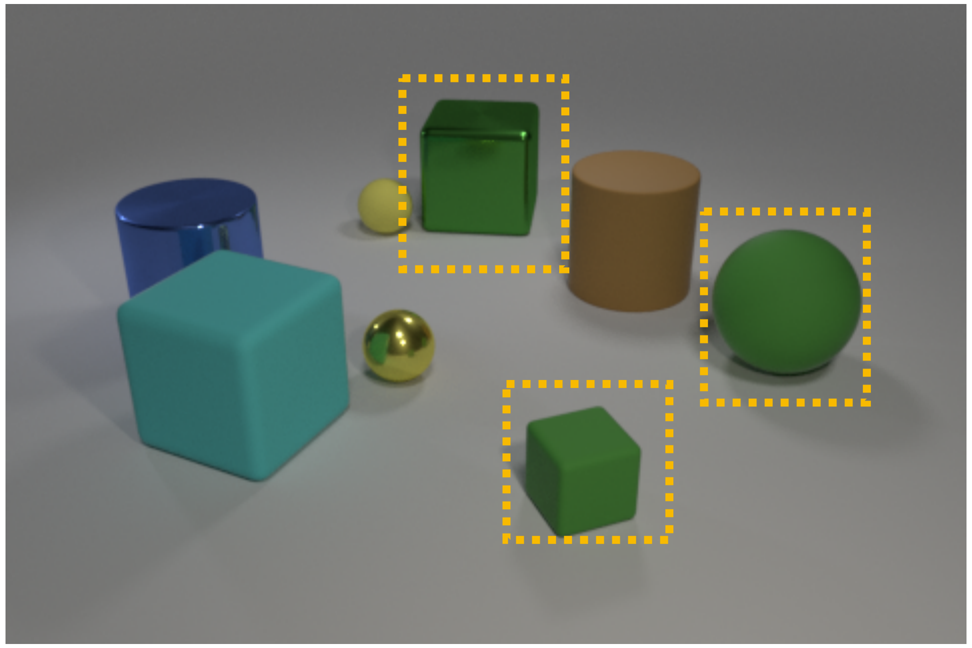
*What is the shape of the large thing that is on the right side of the metallic cube and left of the green **sphere**?*



Compositional model

- Compute span **representations** and **denotations** recursively (symbolic)
- **End-to-end differentiable:** learn good representations by training from downstream supervision only (neural)

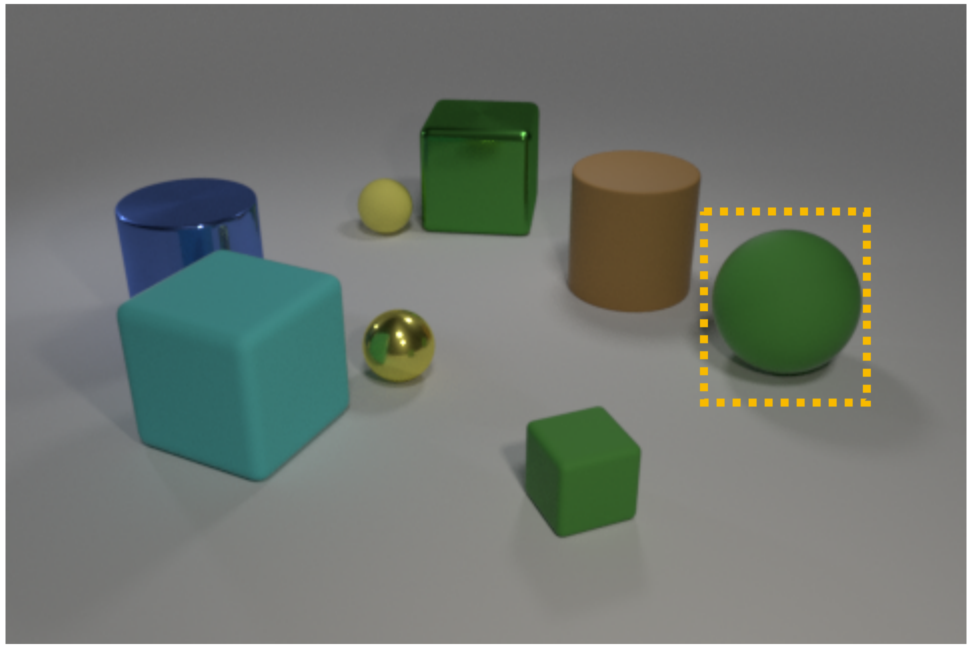
*What is the shape of the large thing that is on the right side of the metallic cube and left of the **green** sphere?*



Desired model properties

- Compute span **representations** and **denotations** recursively (symbolic)
- **End-to-end differentiable:** learn good representations by training from downstream supervision only (neural)

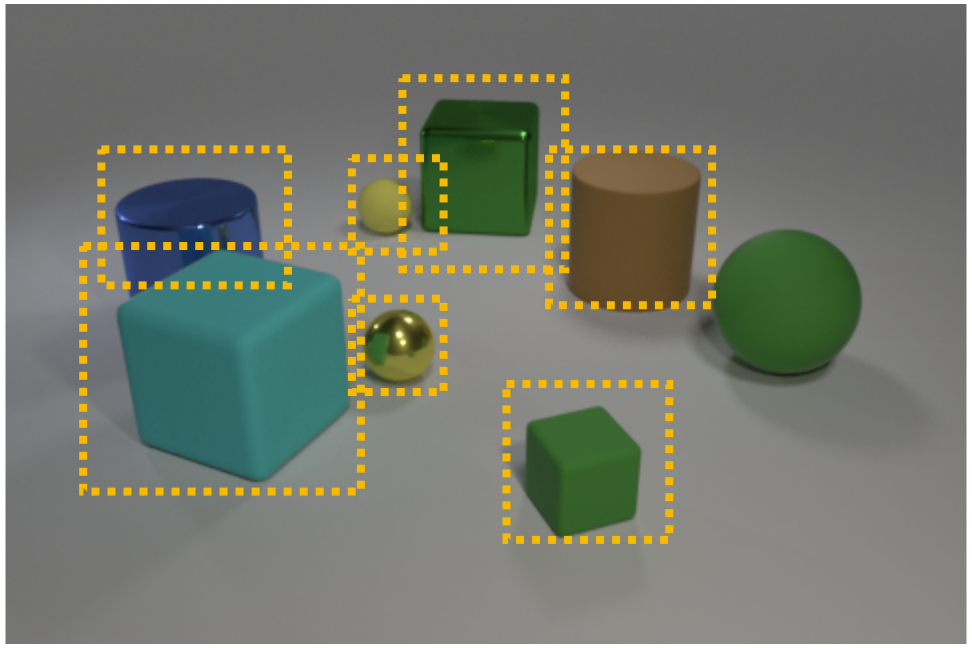
What is the shape of the large thing that is on the right side of the metallic cube and left of the ***green sphere****?*



Compositional model

- Compute span **representations** and **denotations** recursively (symbolic)
- **End-to-end differentiable:** learn good representations by training from downstream supervision only (neural)

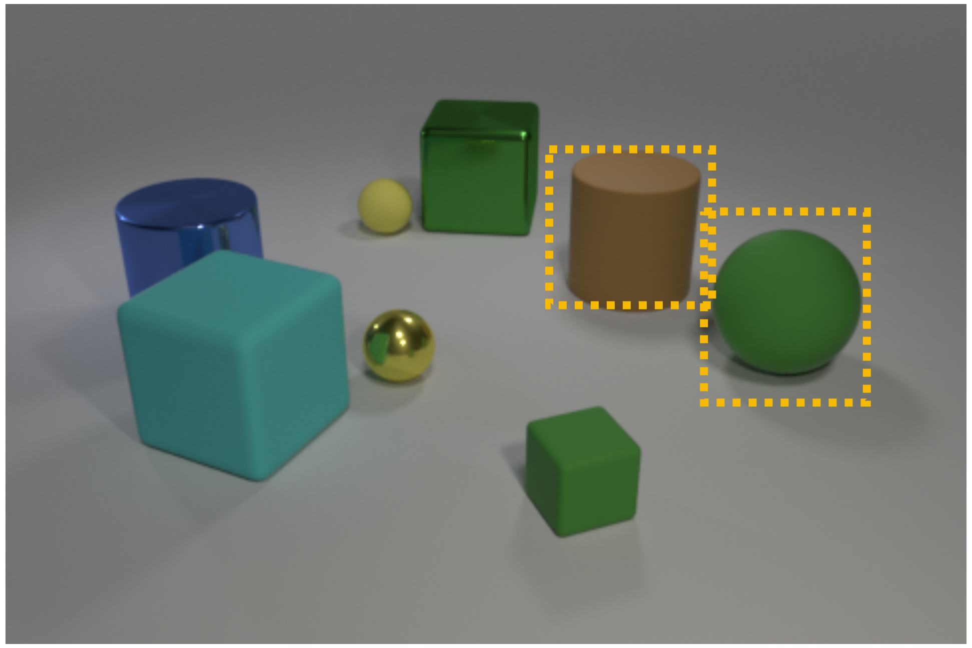
*What is the shape of the large thing that is on the right side of the metallic cube and **left of the green sphere**?*



Compositional model

- Compute span **representations** and **denotations** recursively (symbolic)
- **End-to-end differentiable:** learn good representations by training from downstream supervision only (neural)

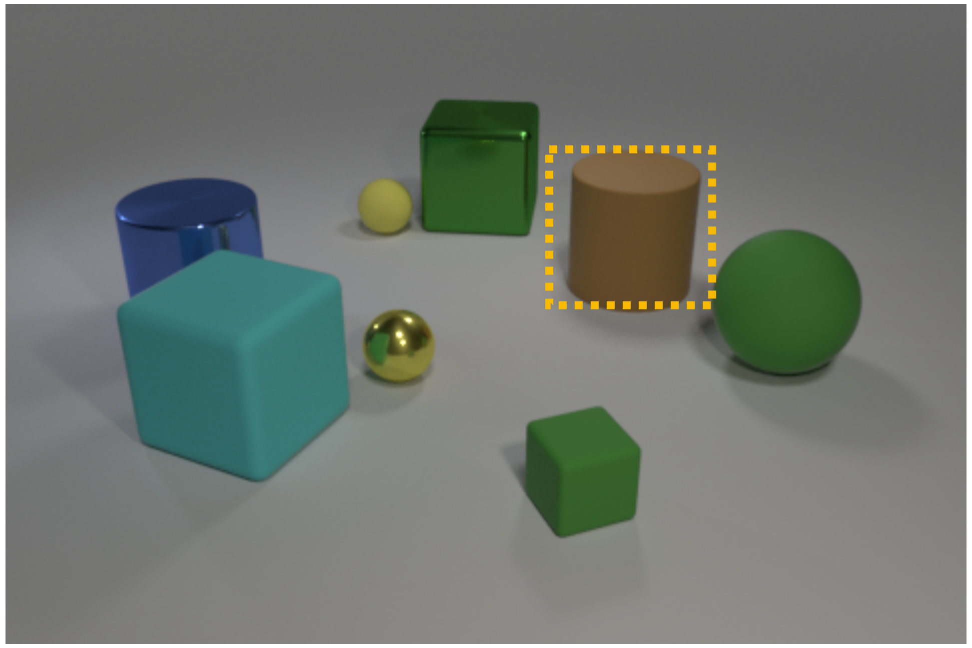
What is the shape of the ***large thing that is on the right side of the metallic cube*** *and left of the green sphere****?***



Compositional model

- Compute span **representations** and **denotations** recursively (symbolic)
- **End-to-end differentiable:** learn good representations by training from downstream supervision only (neural)

*What is the shape of the **large thing that is on the right side of the metallic cube and left of the green sphere**?*

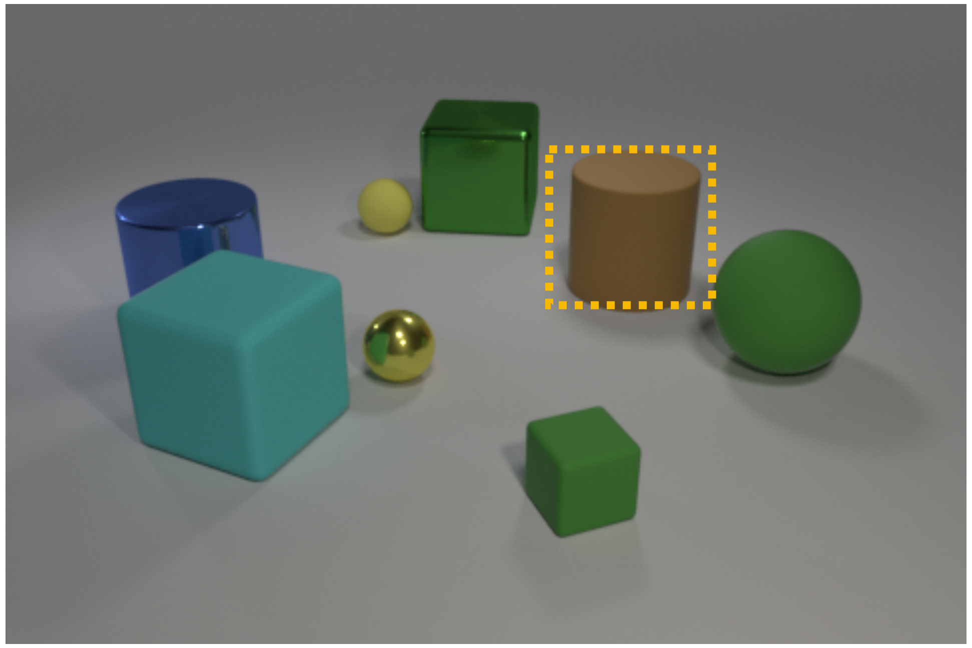


Compositional model

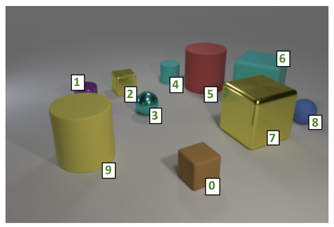
- Compute span **representations** and **denotations** recursively (symbolic)
- **End-to-end differentiable:** learn good representations by training from downstream supervision only (neural)

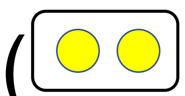
What is the shape of the large thing that is on the right side of the metallic cube and left of the green sphere?

`cylinder`



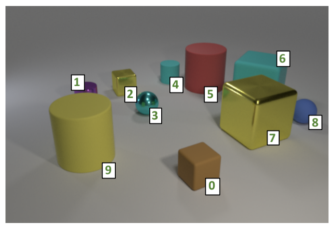
Grounded Latent Trees (GLT):



1. Compute a representation () and denotation ($\boxed{4}$) for spans of length 1, then length 2, etc. (CKY)

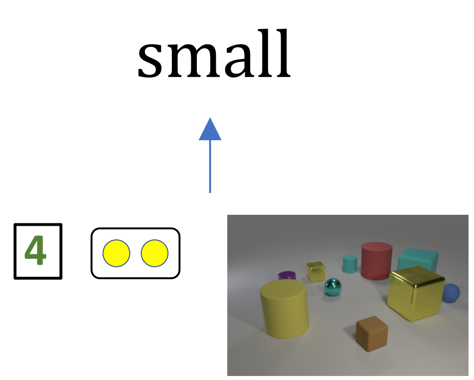
2. Take the representation and denotation of the entire sentence ("the root") and predict the answer

Grounded Latent Trees (GLT):

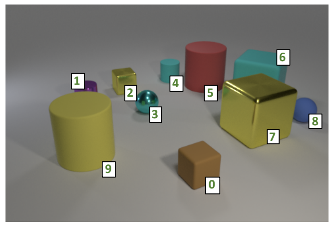


1. Compute a representation ( ) and denotation ($\boxed{4}$) for spans of length 1, then length 2, etc. (CKY)

2. Take the representation and denotation of the entire sentence ("the root") and predict the answer



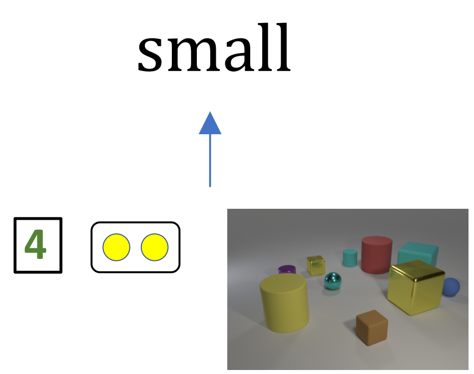
Grounded Latent Trees (GLT):



1. Compute a representation () and denotation ($\boxed{4}$) for spans of length 1, then length 2, etc. (CKY)

Information flow is more restricted than a transformer

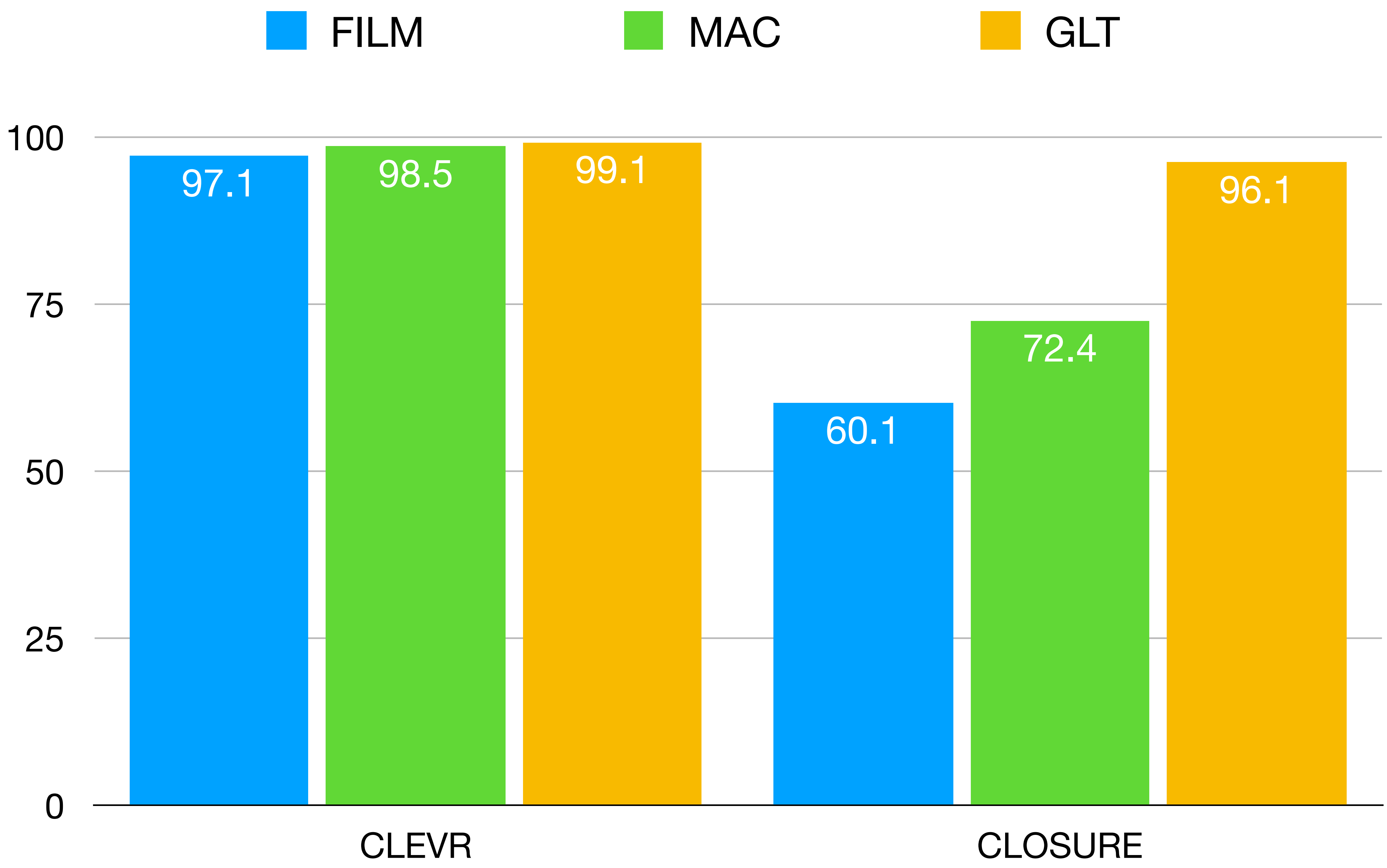
2. Take the representation and denotation of the entire sentence ("the root") and predict the answer



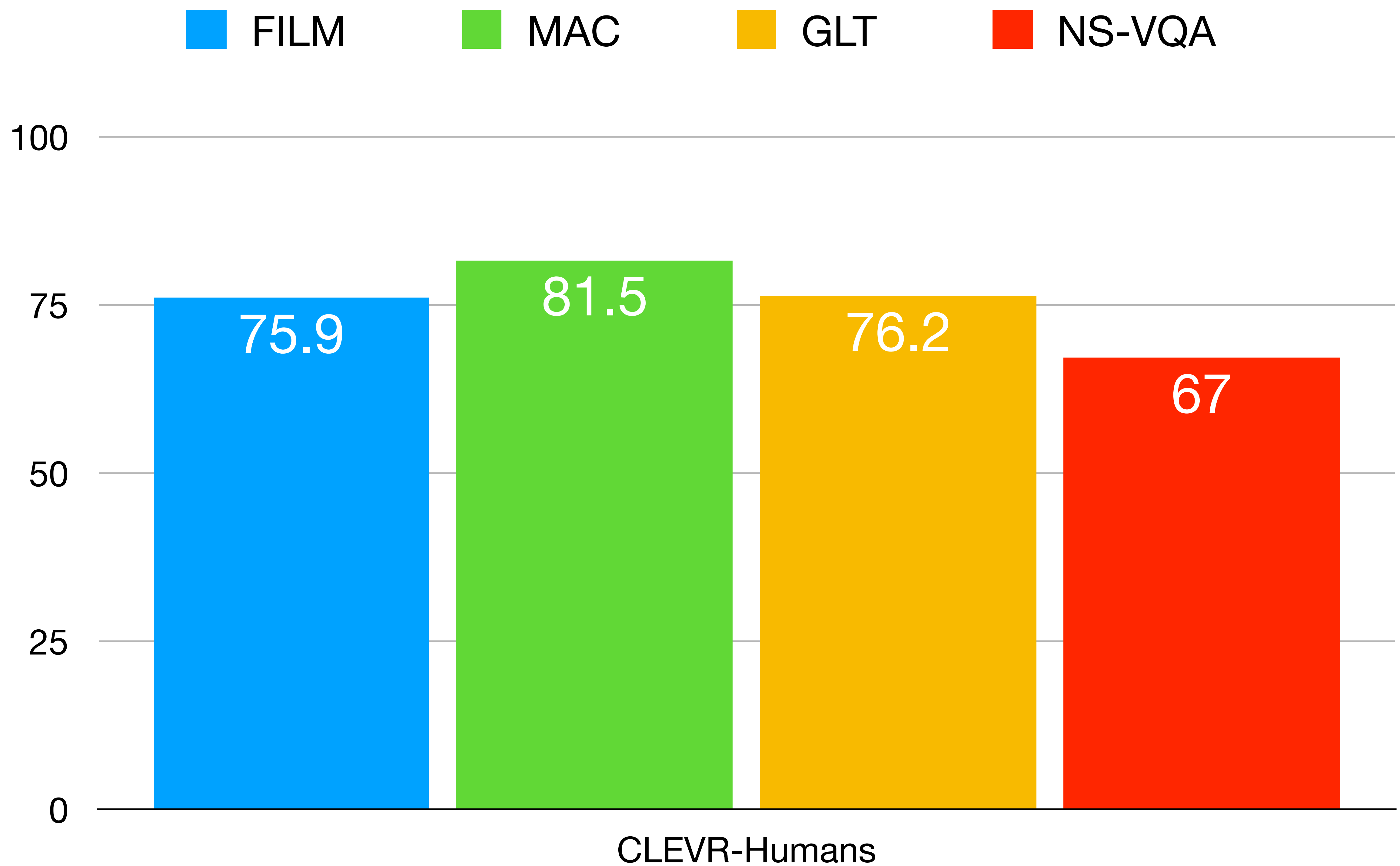
Evaluation: CLEVR, CLOSURE, CLEVR-Humans

- CLEVR: synthetic questions over synthetic images
- CLOSURE: synthetic questions over new compositions
- CLEVR-Humans: Human-authored questions over CLEVR images

Evaluation: CLEVR, CLOSURE

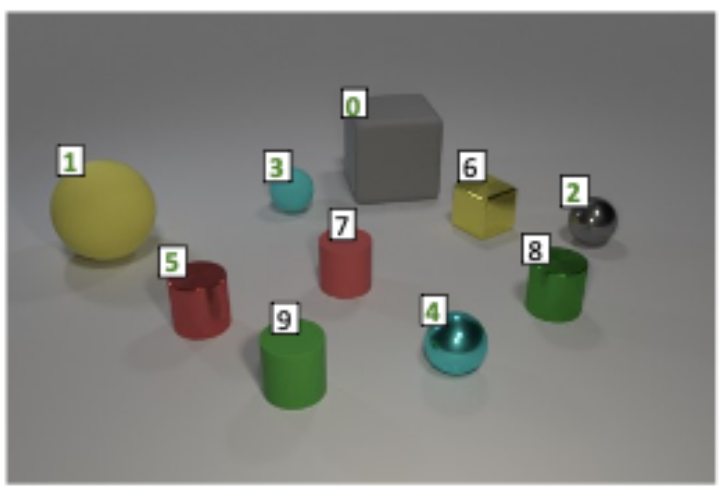


Evaluation: CLEVR-Humans



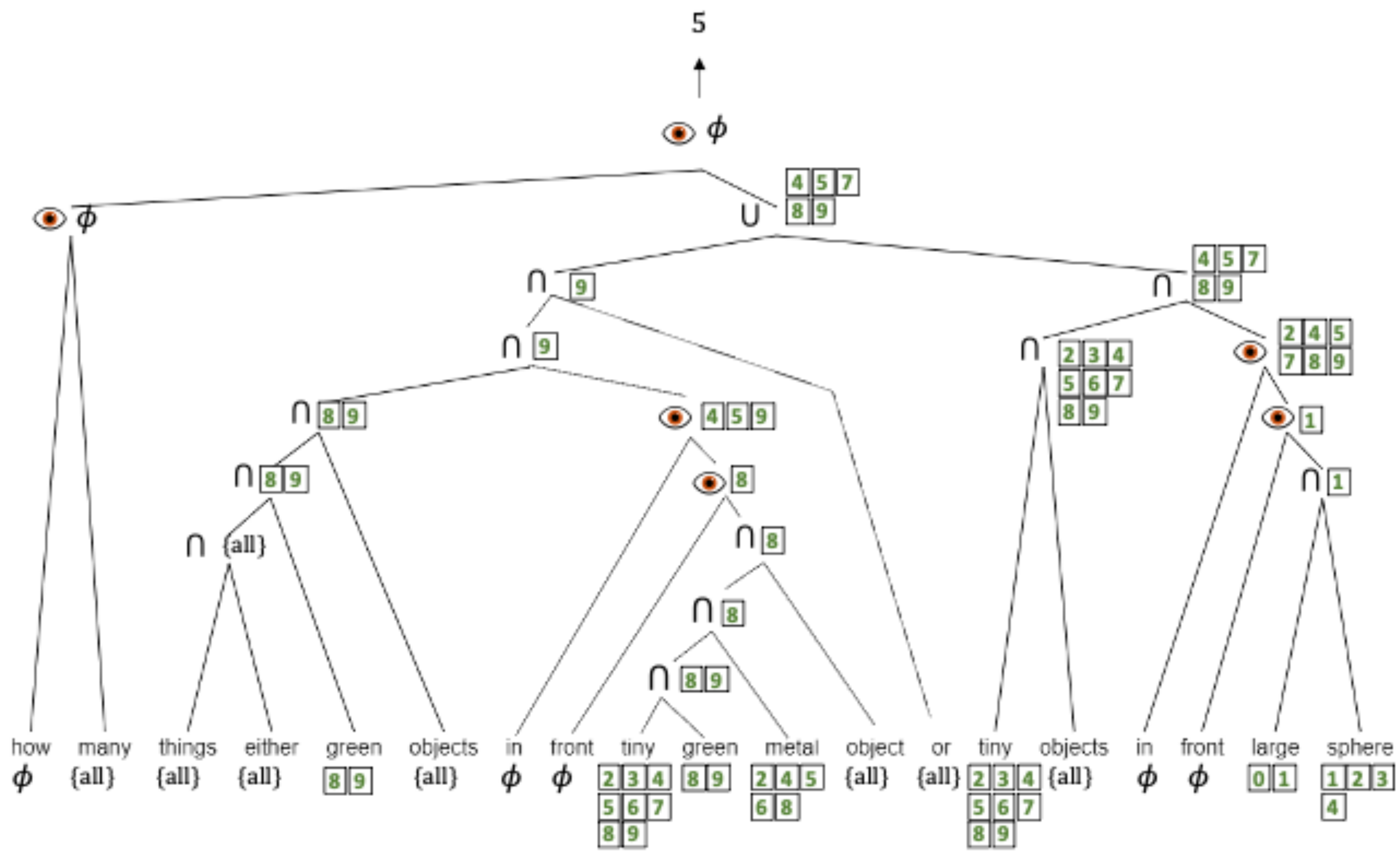
Interpretability

- SKIP (L) SKIP (R)
- ⋃ UNION
- ⋂ INTERSECTION
- VISUAL



How many things are either green objects in front of the tiny green metal object or tiny objects in front of the large sphere? (5)

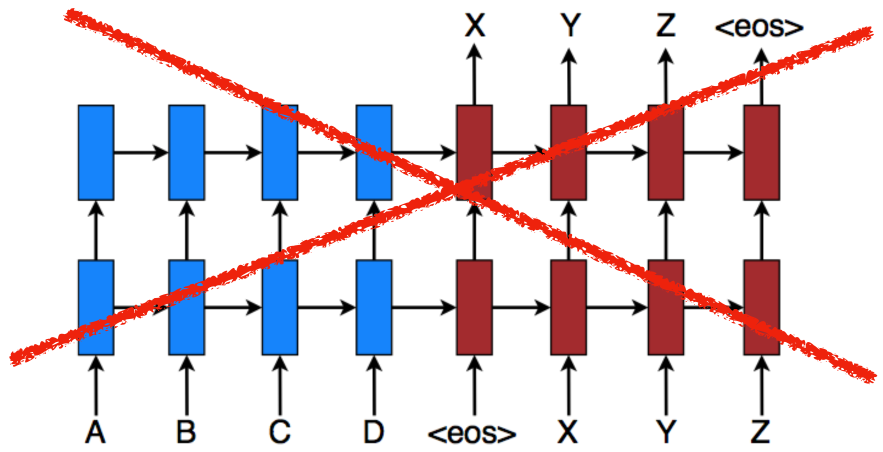
{all} = 0 1 2 3 4 5 6 7 8 9

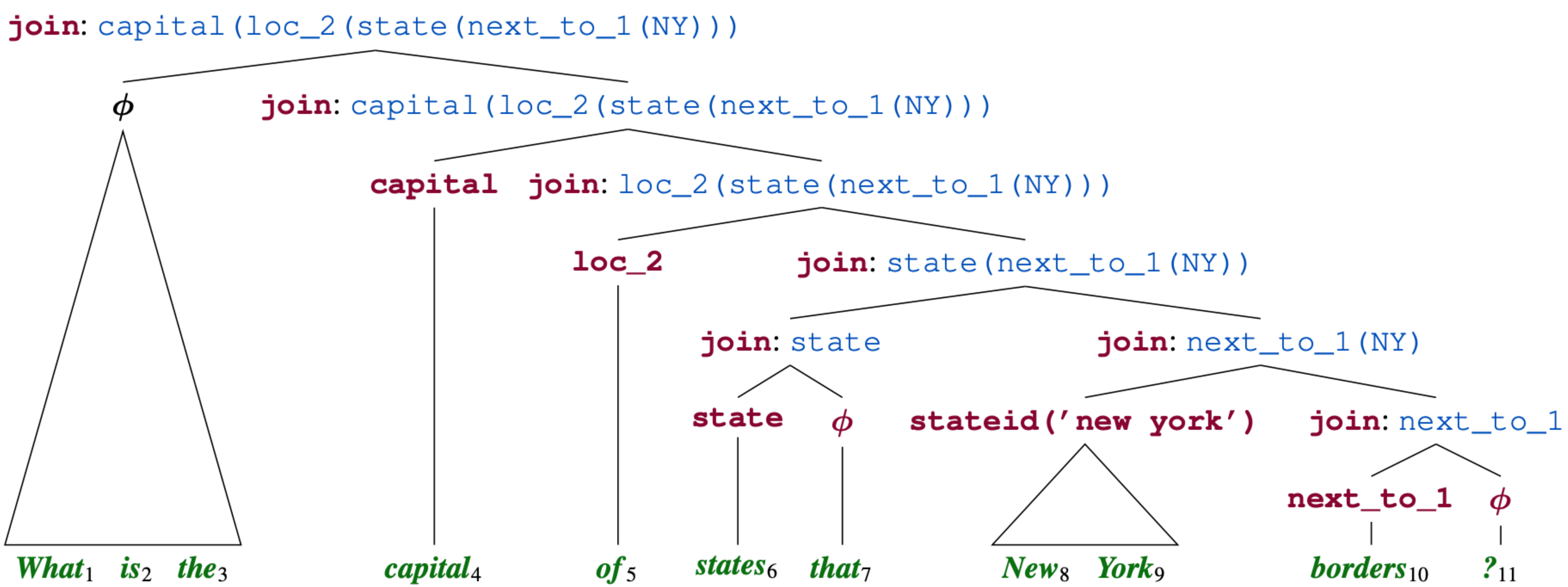


Span-based Semantic Parsing for Compositional Generalization

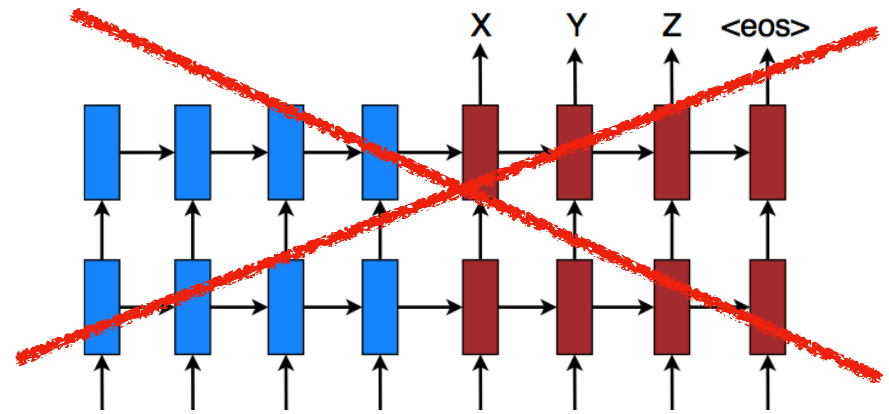
Jonathan Herzig

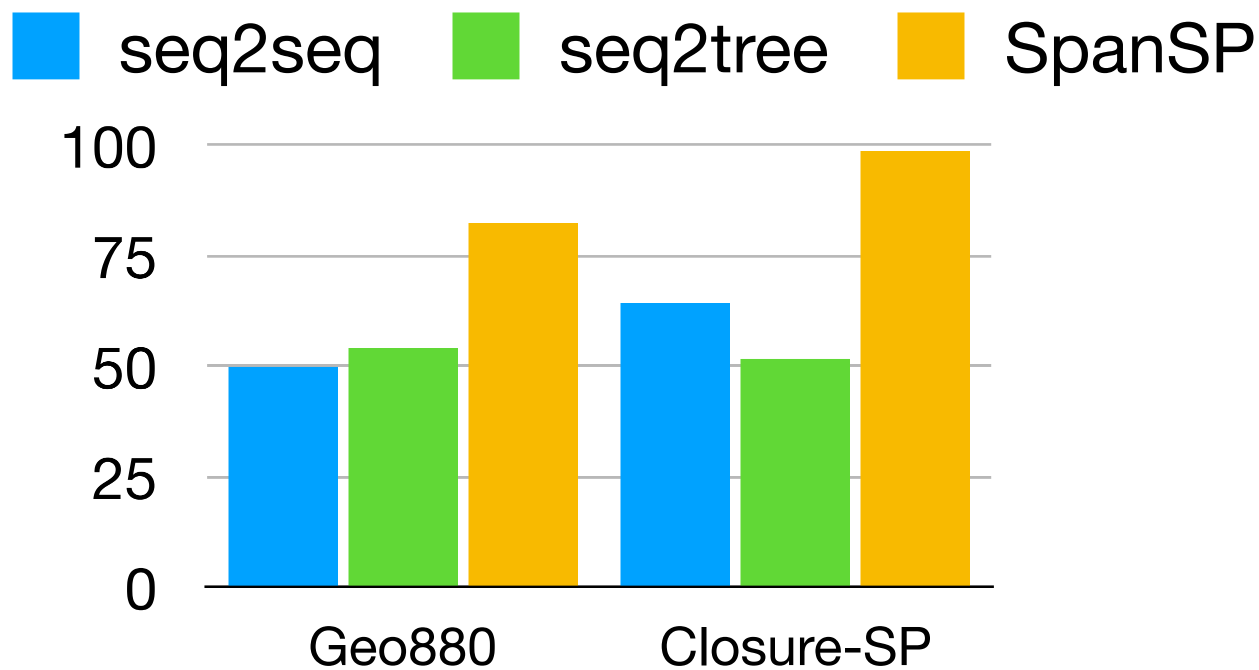
Span-based semantic parsing

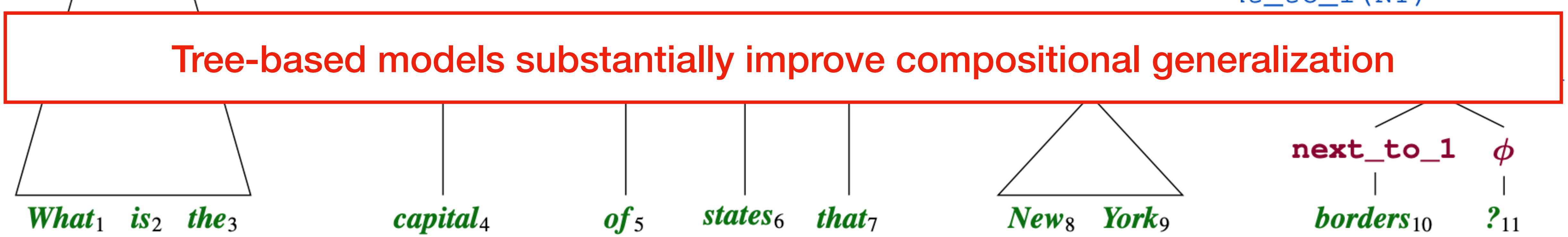




Span-based semantic parsing



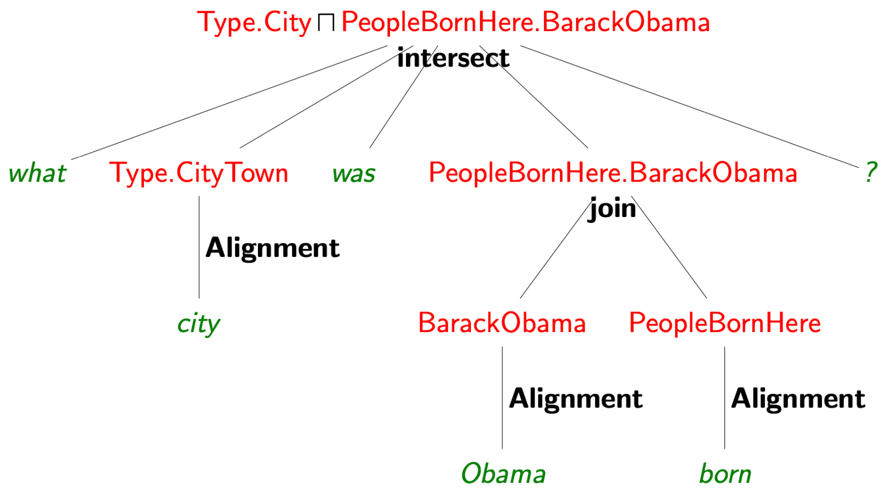




Tree-based models substantially improve compositional generalization

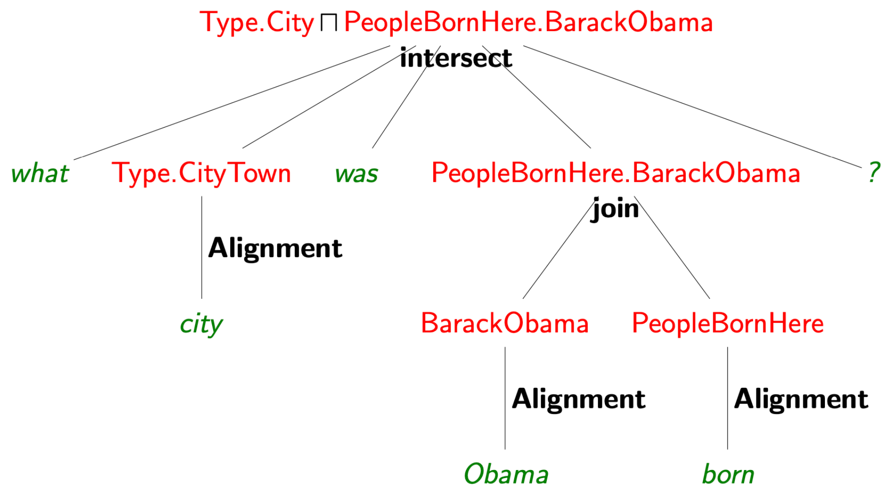
Summary

early 2016

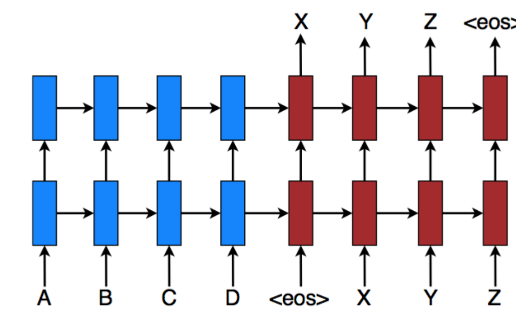


Summary

early 2016



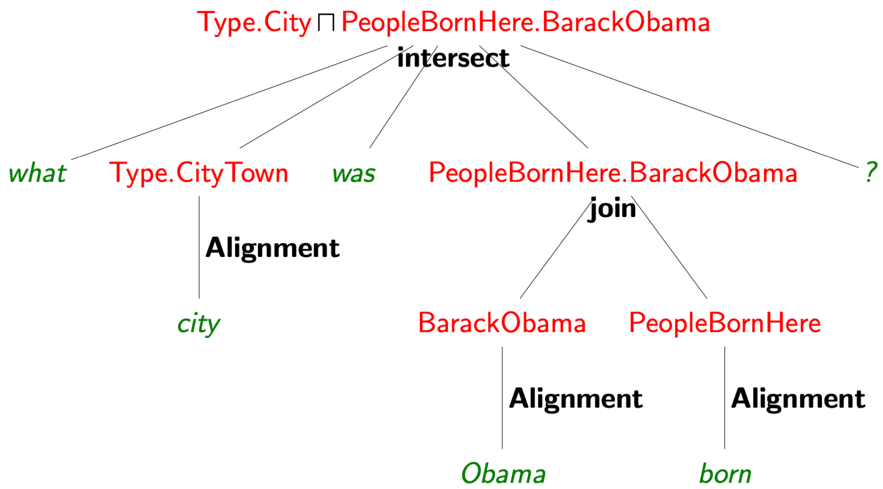
late 2016



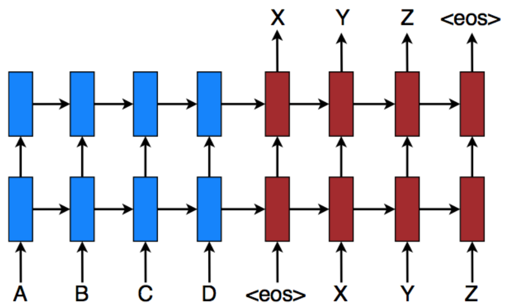
Simple, general, flexible! 😃

Summary

early 2016



late 2016

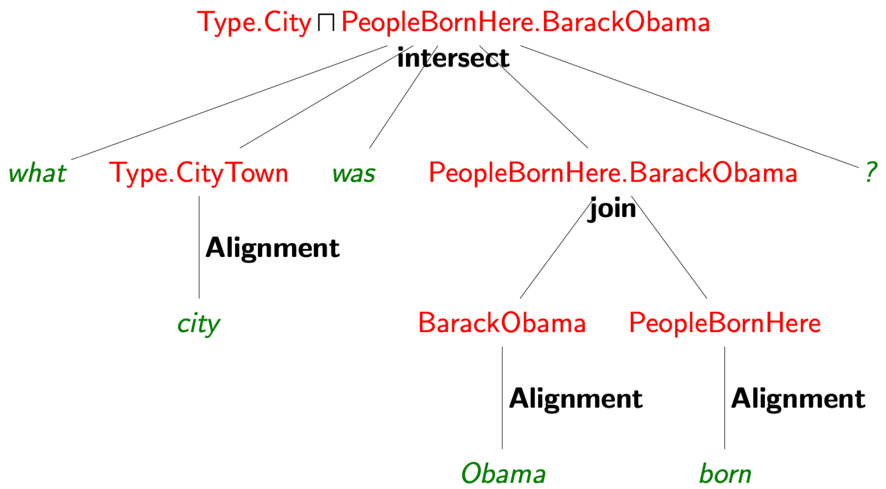


Simple, general, flexible! 😃

But it does not compositionally generalize

Summary

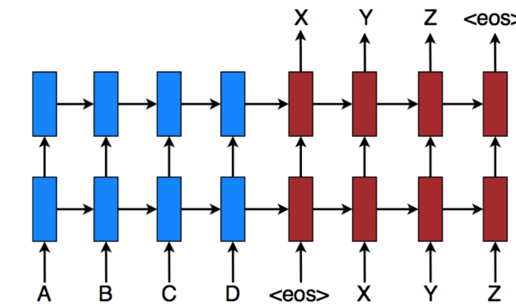
early 2016



What's here?

?

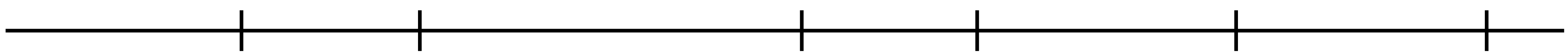
late 2016



Simple, general, flexible!

But it does not compositionally generalize

The method spectrum of compositional generalization



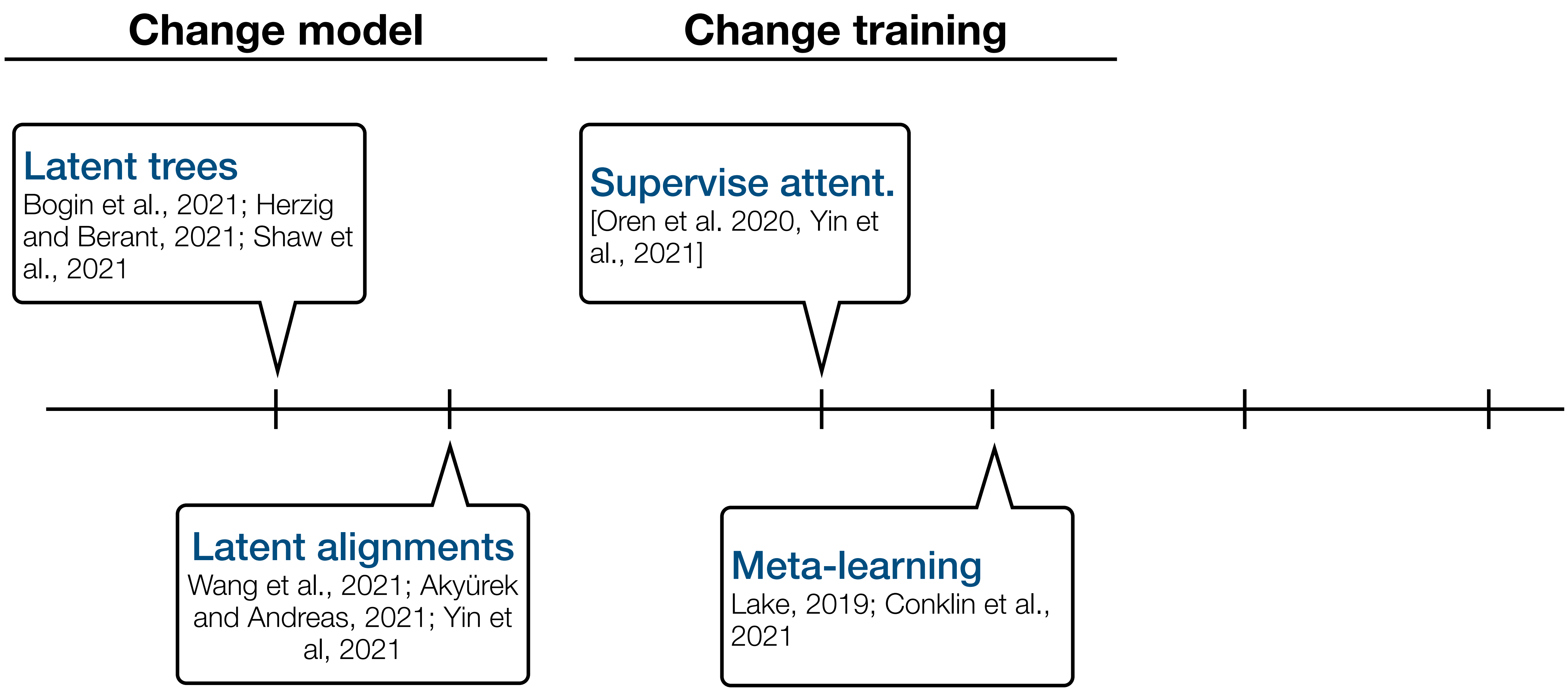
The method spectrum of compositional generalization

Change model

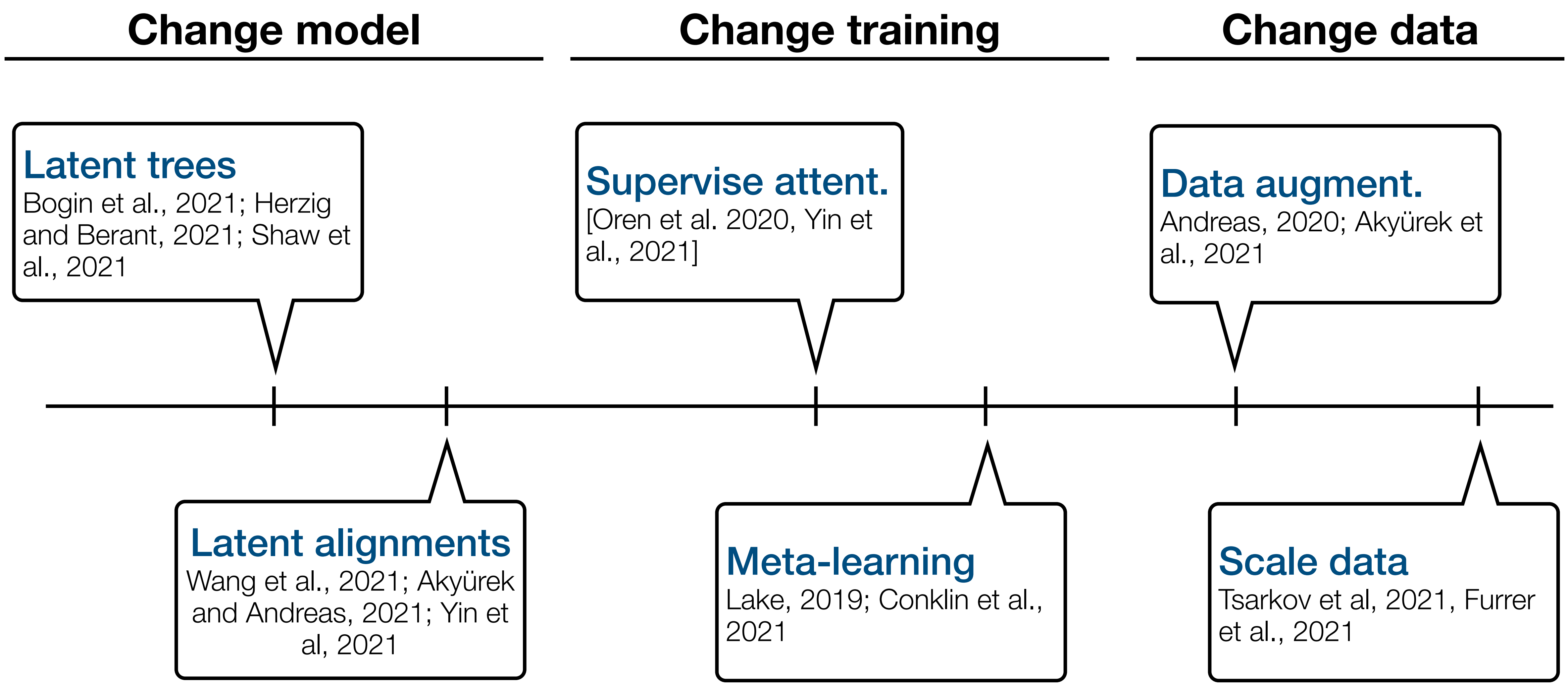
Latent trees
Bogin et al., 2021; Herzig and Berant, 2021; Shaw et al., 2021

Latent alignments
Wang et al., 2021; Akyürek and Andreas, 2021; Yin et al, 2021

The method spectrum of compositional generalization



The method spectrum of compositional generalization



The method spectrum of compositional generalization

Change model

Latent trees
Bogin et al., 2021; Herzig and Berant, 2021; Shaw et al., 2021

Latent alignments
Wang et al., 2021; Akyürek and Andreas, 2021; Yin et al, 2021

Change training

Supervise attent.
[Oren et al. 2020, Yin et al., 2021]

Meta-learning
Lake, 2019; Conklin et al., 2021

Change data

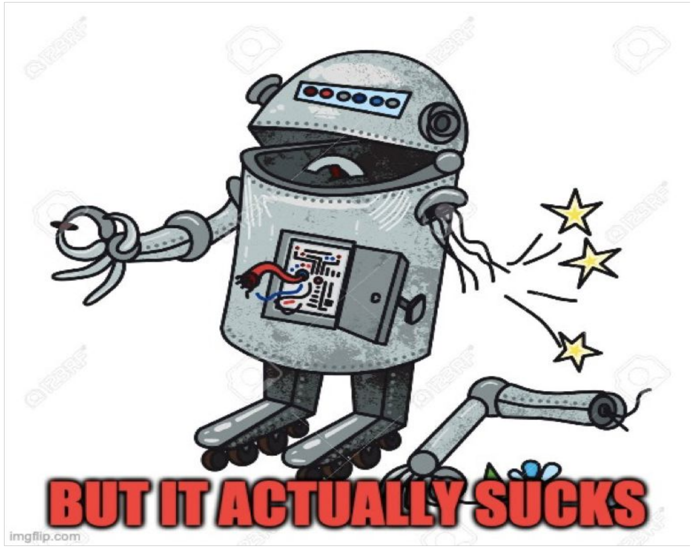
Data augment.
Andreas, 2020; Akyürek et al., 2021

Scale data
Tsarkov et al, 2021, Furrer et al., 2021

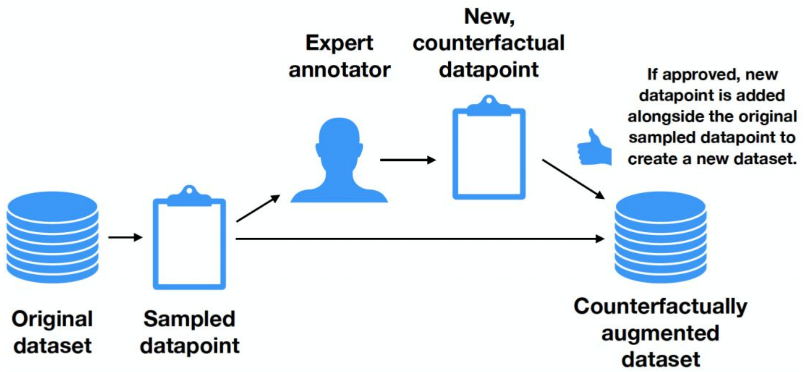
Currently there is still trade-off between performance and inductive bias, but…

Evaluating robustness by controlled generation from symbolic representations

Evaluation crisis in NLP



Perturbations for evaluations

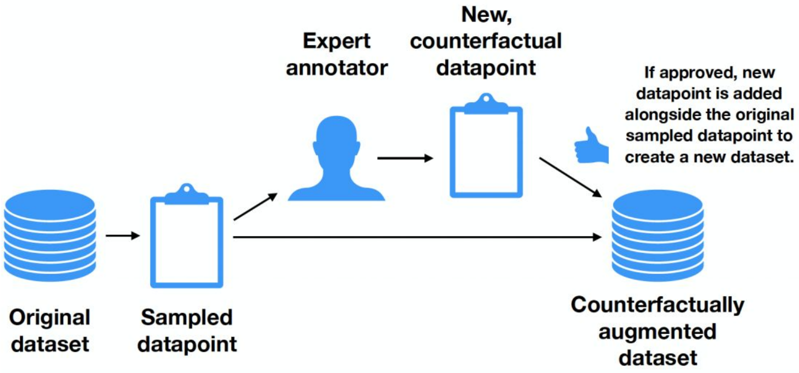


From the first moment I felt **bored** 😔

From the first moment I felt **absorbed** 😃

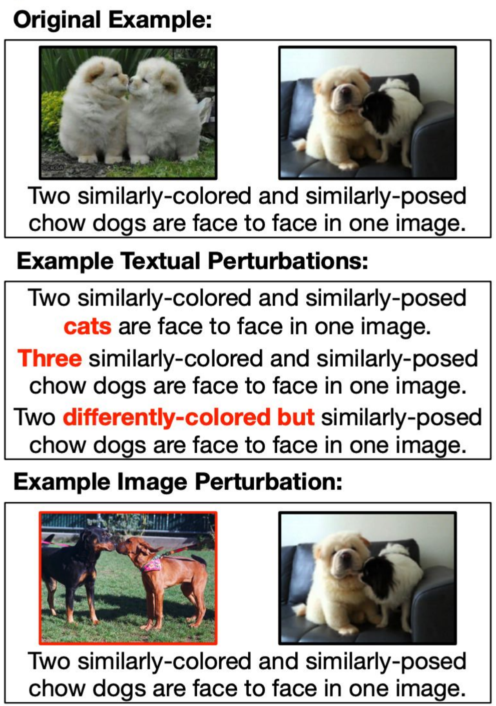
[Kaushik, Hovy, and Lipton, 2020; Gardner et al, 2020]

Perturbations for evaluations



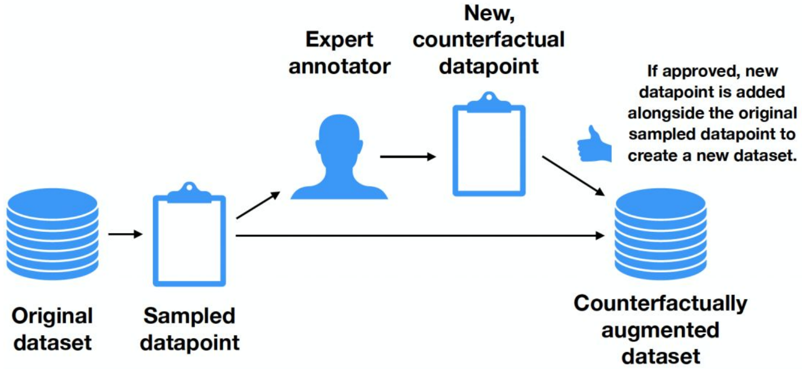
From the first moment I felt **bored**

From the first moment I felt **absorbed**



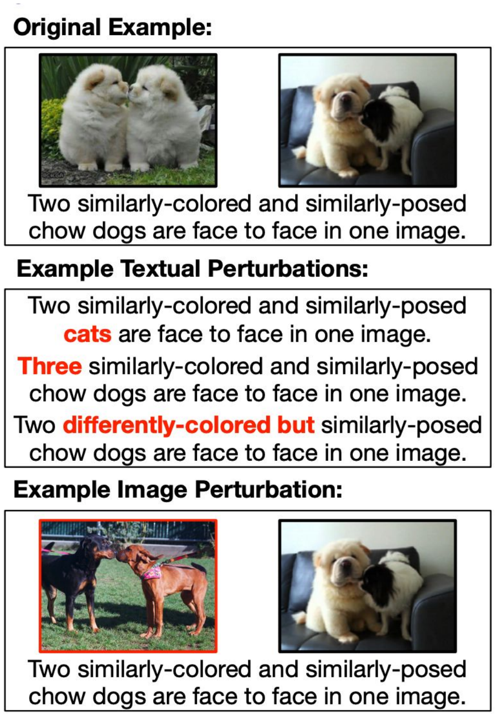
[Kaushik, Hovy, and Lipton, 2020; Gardner et al, 2020]

Perturbations for evaluations



From the first moment I felt **bored**

From the first moment I felt **absorbed**



[Kaushik, Hovy, and Lipton, 2020; Gardner et al, 2020]

Break, Perturb, Build: Automatic Perturbation of Reasoning Paths through Question Decomposition

Mor Geva

Tomer Wolfson

Getting **control** with structured meaning representations

Which gallery was founded later,
Hughes/Donahue or Art Euphoric?

Getting **control** with structured meaning representations

Which gallery was founded later,
Hughes/Donahue or Art Euphoric?

QDMR

1. When was the Hughes/Donahue gallery founded?
2. When was the Art Euphoric gallery founded?
3. Which is larger of #1 and #2?

Wolfson et al, 2020

Getting **control** with structured meaning representations

Which gallery was founded later, Hughes/Donahue or Art Euphoric?

QDMR

1. When was the Hughes/Donahue gallery founded?
2. When was the Art Euphoric gallery founded?
3. Which is larger of #1 and #2?

easy to manipulate!

Wolfson et al, 2020

Getting **control** with structured meaning representations

Which gallery was founded later, Hughes/Donahue or Art Euphoric?

↓

QDMR

1. When was the Hughes/Donahue gallery founded?
2. ~~When was the Art Euphoric gallery founded?~~
3. ~~Which is larger of #1 and #2?~~

↓

When was the Hughes/Donahue gallery founded?

Wolfson et al, 2020

Getting **control** with structured meaning representations

Which gallery was founded later, Hughes/Donahue or Art Euphoric?

QDMR

1. When was the Hughes/Donahue gallery founded?
2. When was the Art Euphoric gallery founded?
3. Which is ~~larger~~ **smaller** of #1 and #2?

Which gallery was founded first, Hughes/Donahue or Art Euphoric?

Wolfson et al, 2020

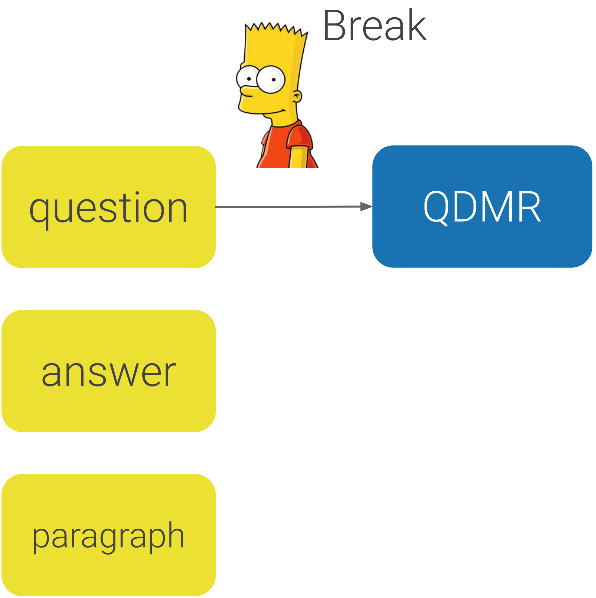
Example generation pipeline (reading comprehension)

question

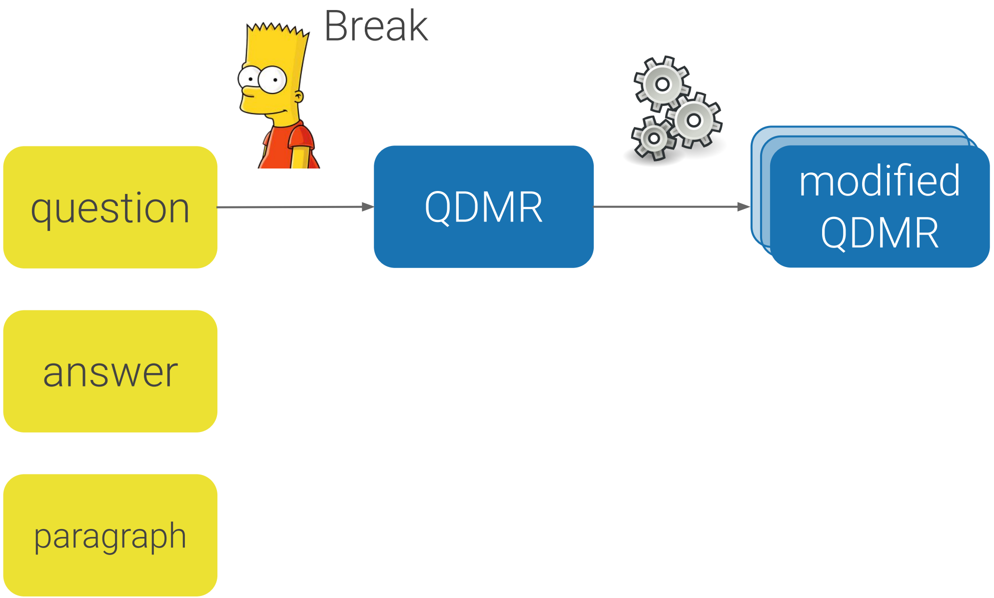
answer

paragraph

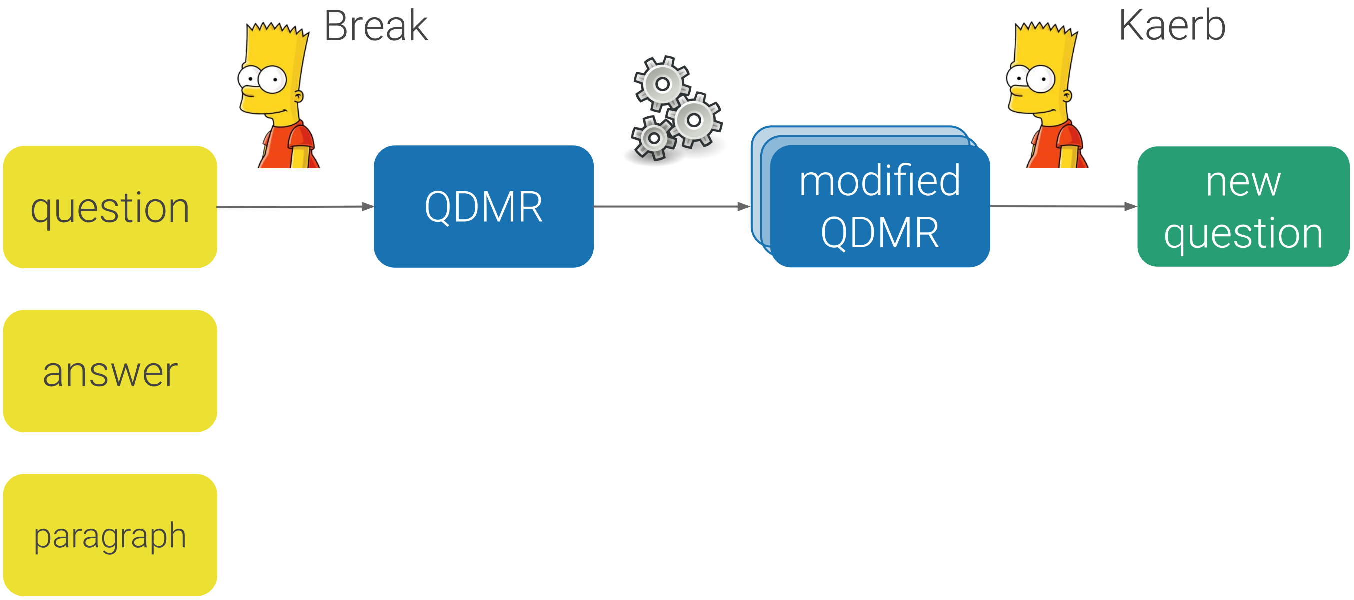
Example generation pipeline (reading comprehension)



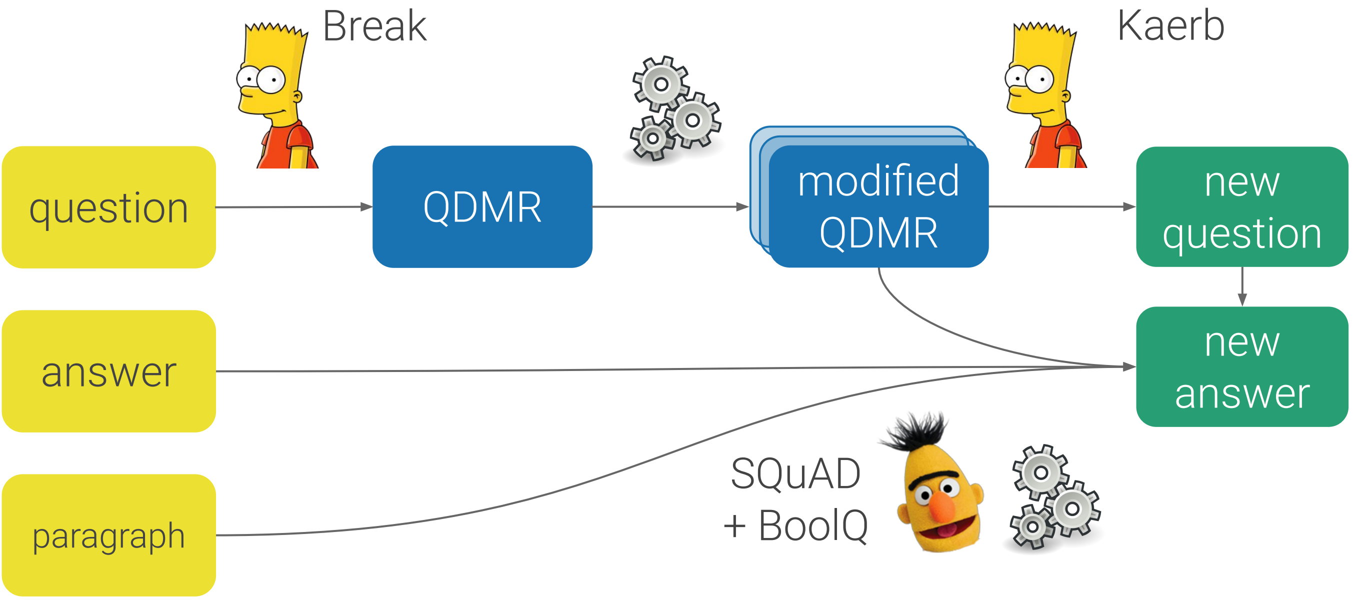
Example generation pipeline (reading comprehension)



Example generation pipeline (reading comprehension)



Example generation pipeline (reading comprehension)



Generated example from HotpotQA

question + answer

How many novels are there in the series of novels of which Shadows in Flight is the tenth novel?

fifteen

QDMR

1. series of novels of which Shadows in Flight is the tenth novel
2. novels in #1
3. number of #2

modified QDMR

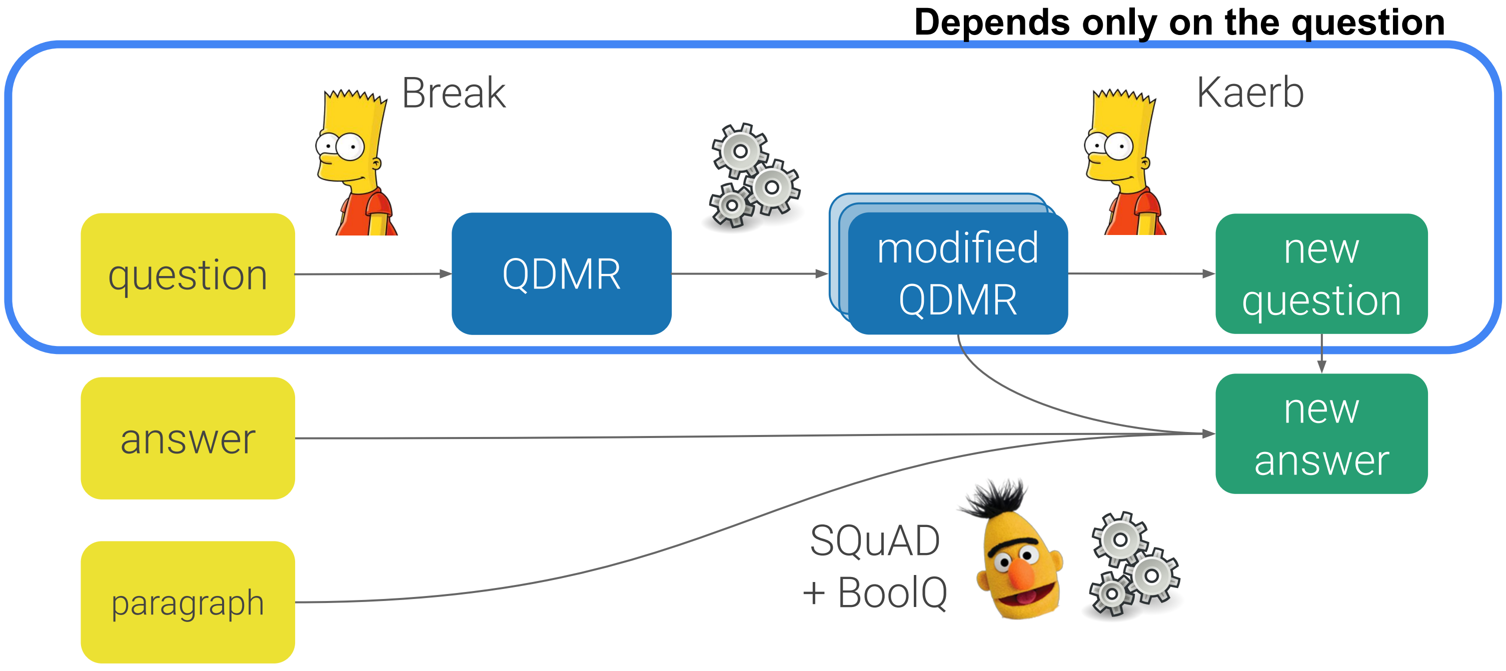
1. series of novels of which Shadows in Flight is the tenth novel
2. novels in #1
3. number of #2
4. **if #3 is equal to 23**

question + answer

If Shadows in Flight is the tenth novel in a series of 23 novels?

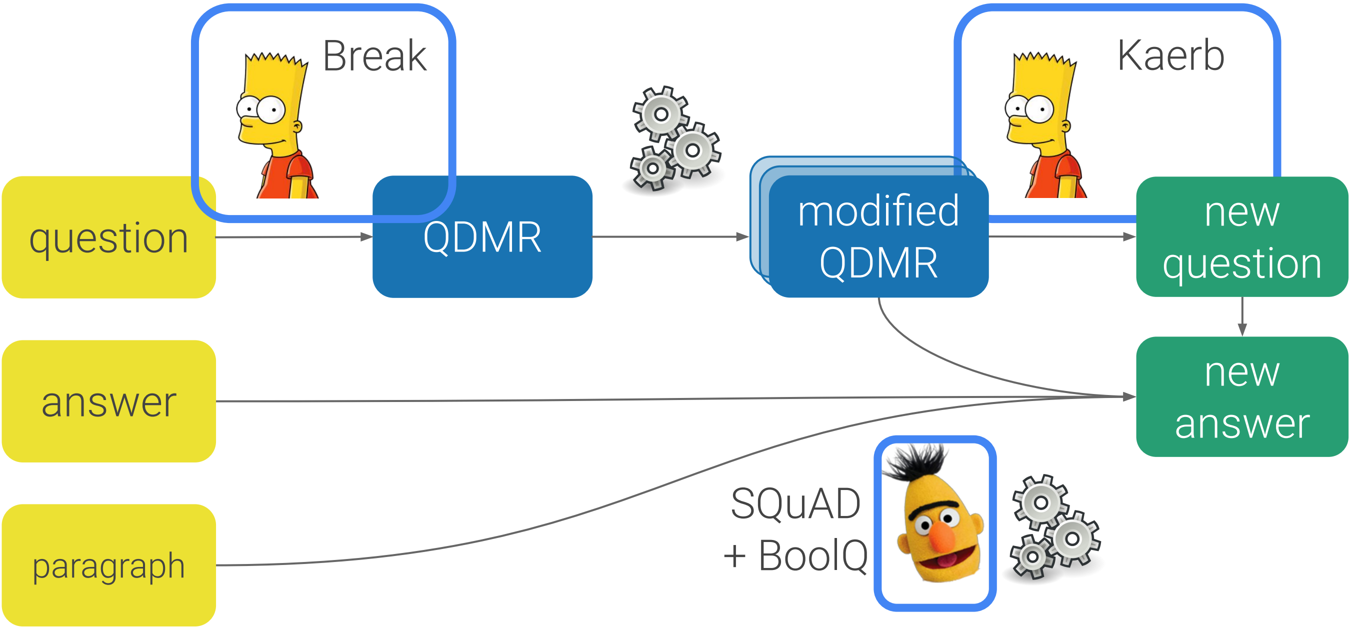
no

Observations



Applicable to other modalities (video, image, table)

Observations



QDMR parsers, question generators, and reading comprehension work!

The generated examples cover most of the original datasets

DROP	HotpotQA	IIRC
64%	70%	45%

The vast majority of generated examples is valid

200-500 examples per transformation, each was validated by 3 crowdworkers

85%

DROP

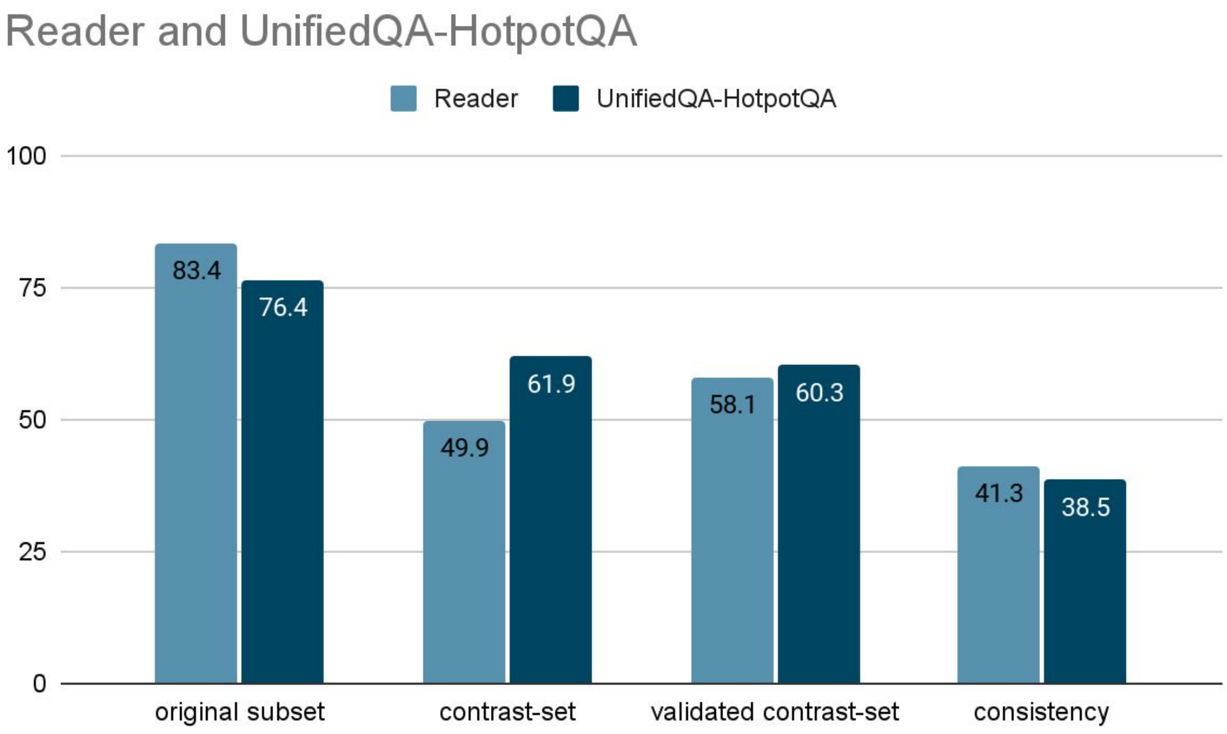
89%

HotpotQA

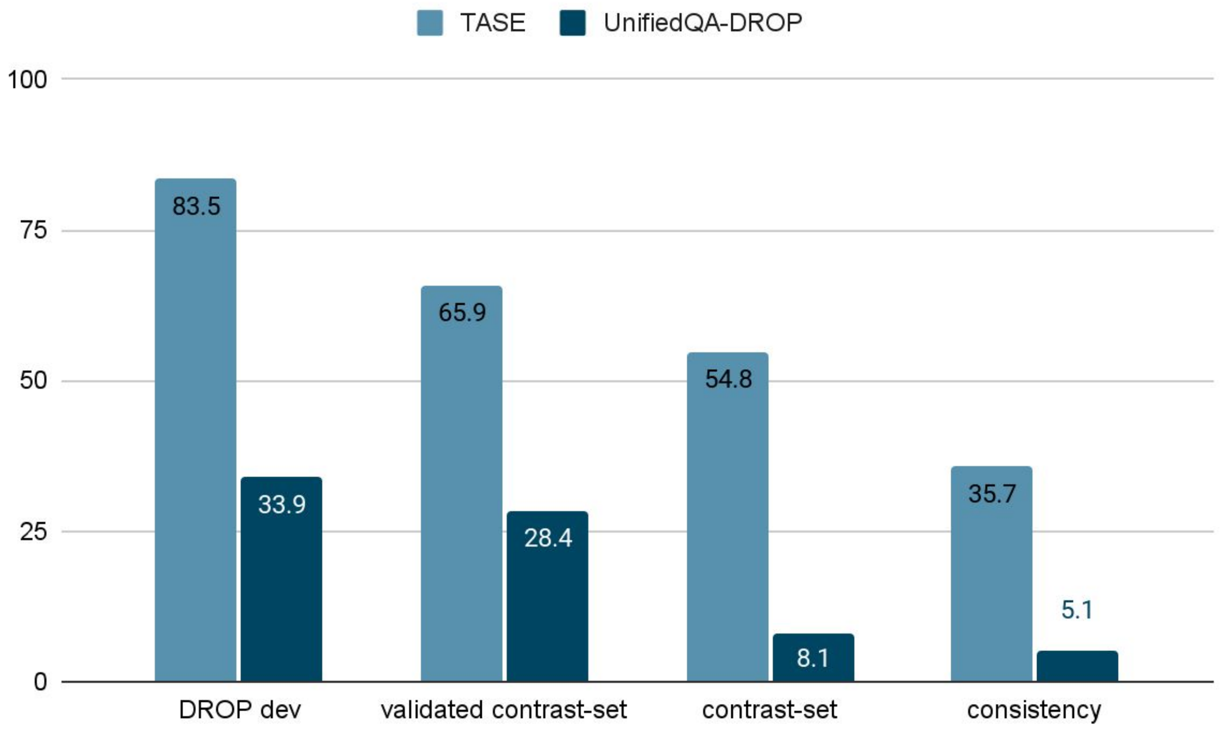
90%

IIRC

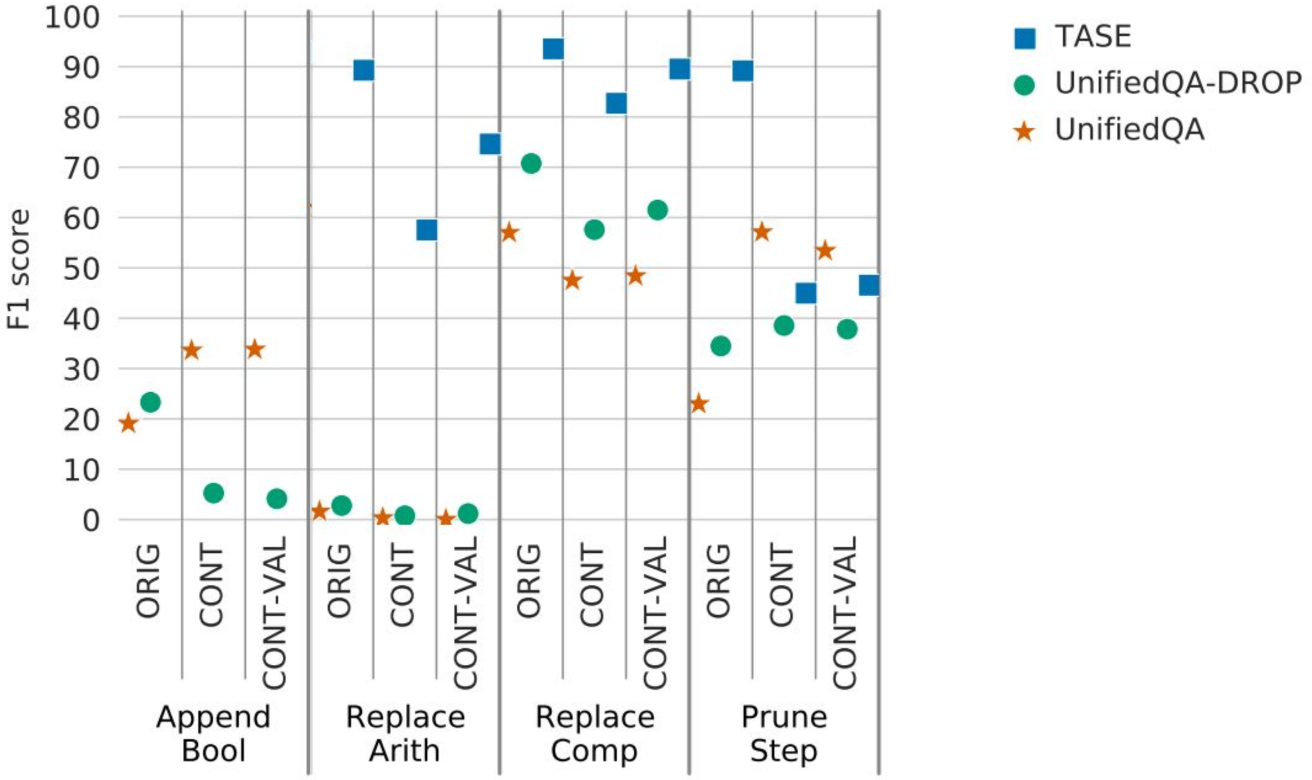
Model performance drops on generated examples (*HotpotQA*)



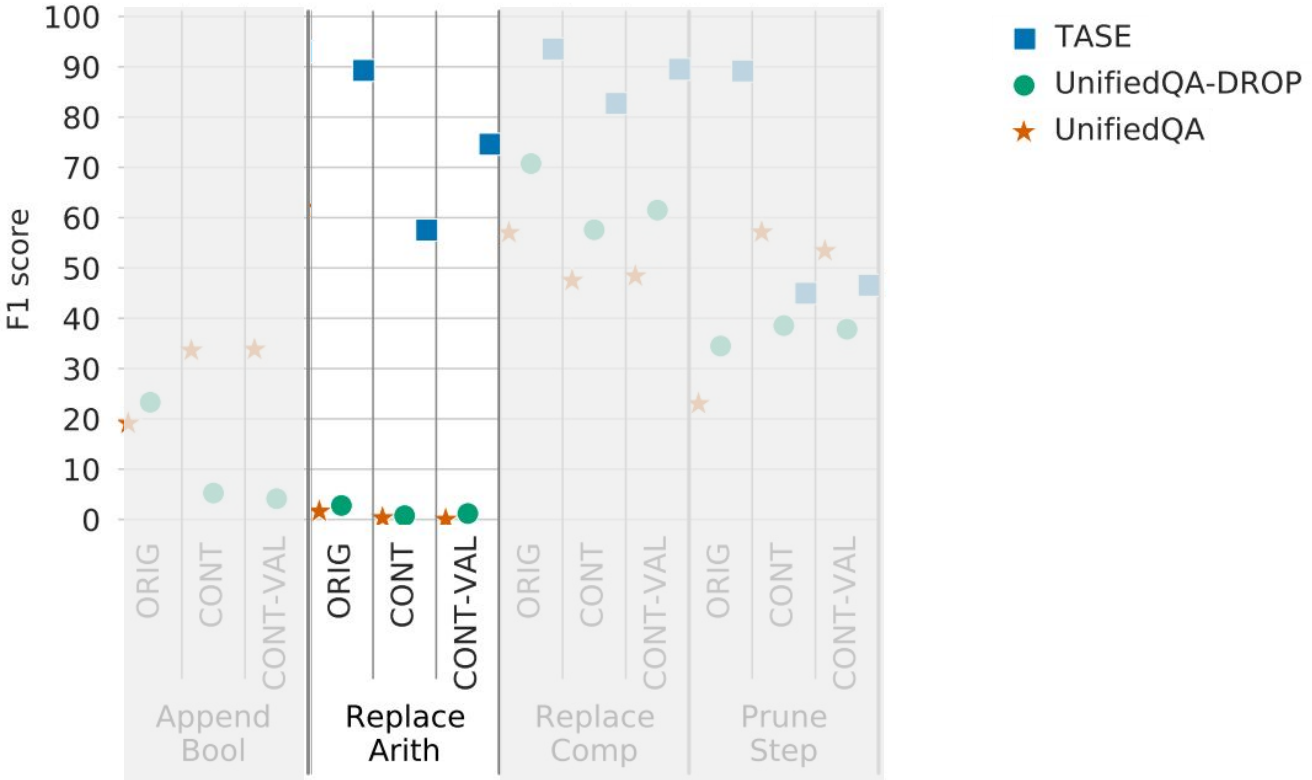
Model performance drops on generated examples (*DROP*)



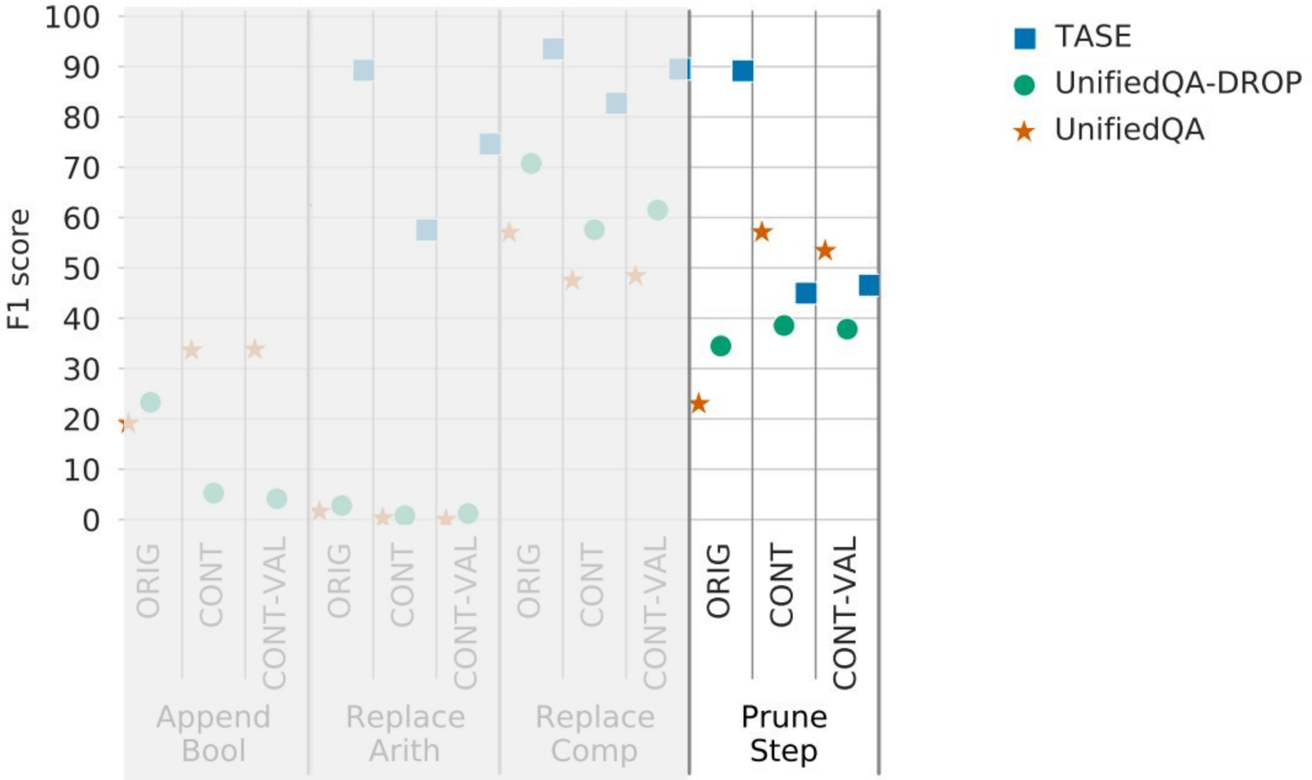
Analysis: performance per transformation (*DROP*)



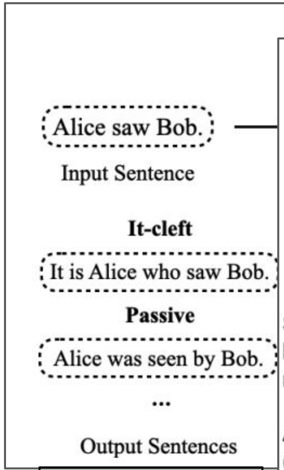
Analysis: performance per transformation (*DROP*)



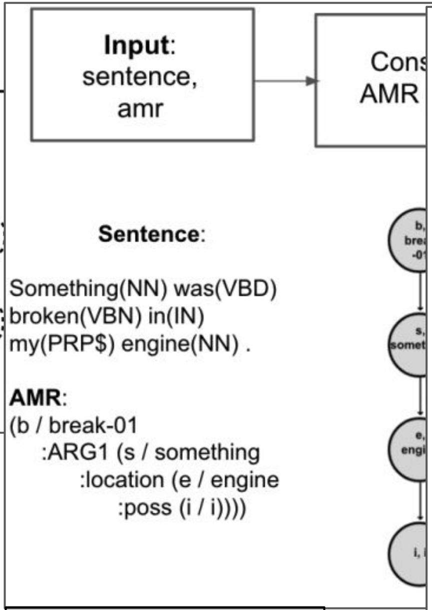
Analysis: performance per transformation (*DROP*)



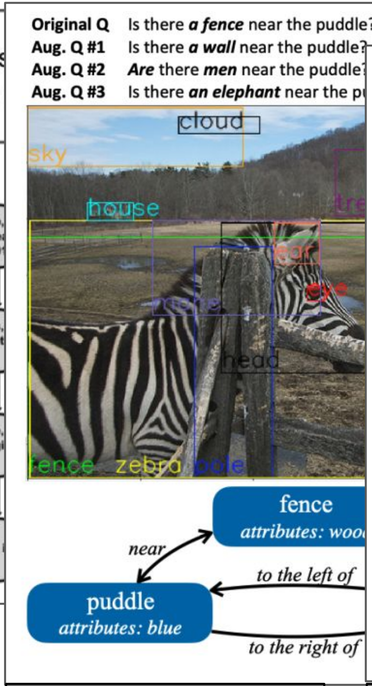
New research program?



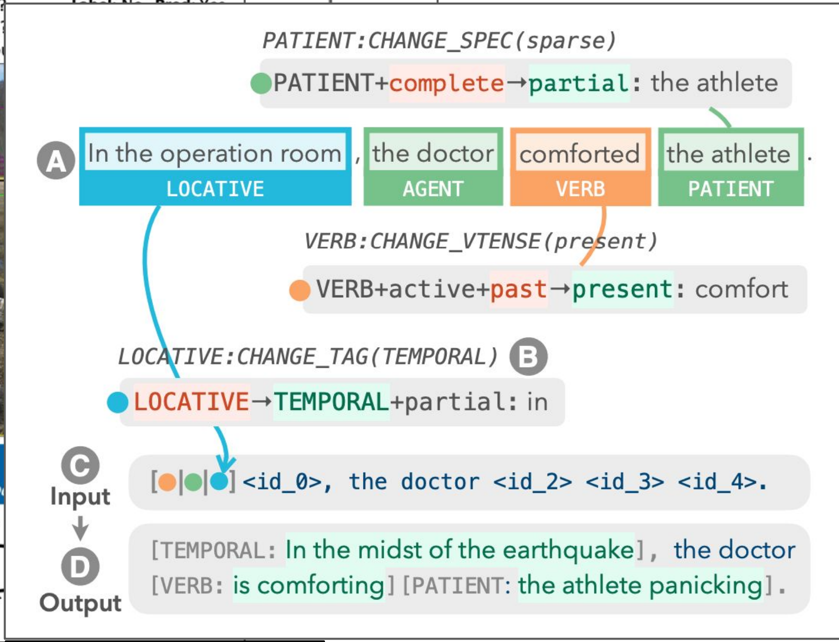
Syntactic parser
Li et al., 2020



AMR, Rashkit and Flanigan, 2021



Scene graphs, Bitton et al., 2021



SRL, Ross et al., 2021

Role of symbolic models

Two roles discussed for symbolic models on two ends of the spectrum

- Explicitly-compositional models for compositional generalization
  - Great performance
  - Interpretability
  - But is it an upper-bound? A guide for more flexible model? Or the key to future models?
- Controlled data generation for evaluating (and improving?) robustness
  - Train with standard models
  - Sim2Real: cover the blind spots of your huge pre-trained models

Thank you! questions?

Mor Geva

Tomer Wolfson

Jonathan Herzig

Ben Bogin

Sanjay Sumbramanian

Matt Gardner