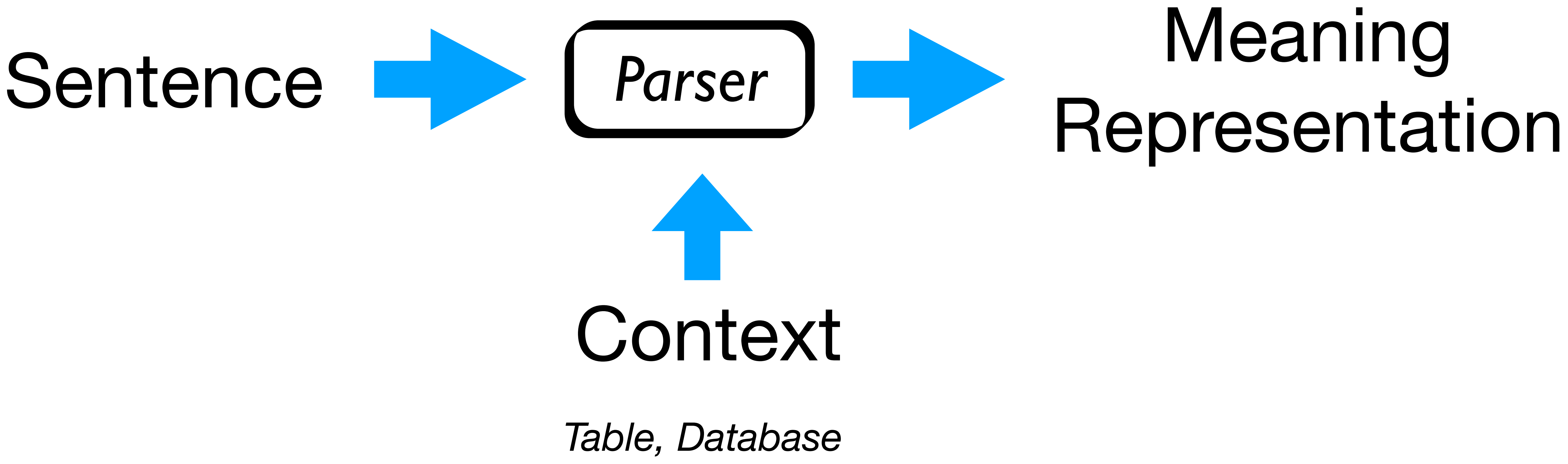
Learning Semantic Parsers from Denotations with Latent Structured Alignments and Abstract Programs

Bailin Wang, Ivan Titov, Mirella Lapata

bailin.wang@ed.ac.uk *{ititov,mlap}@inf.ed.ac.uk*

Edinburgh
University of Edinburgh
Natural Language Processing
NLP

Background



Sentence

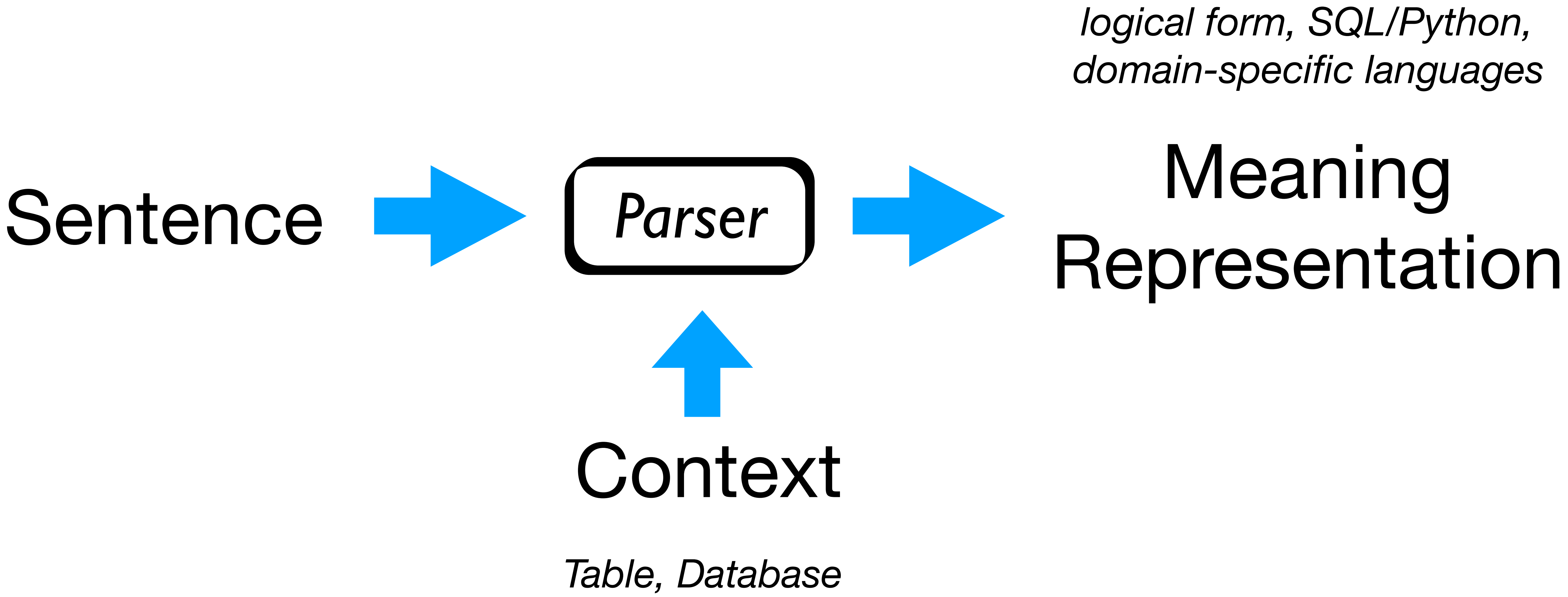
Parser

Meaning Representation

Context

Table, Database

Background



logical form, SQL/Python, domain-specific languages

Sentence

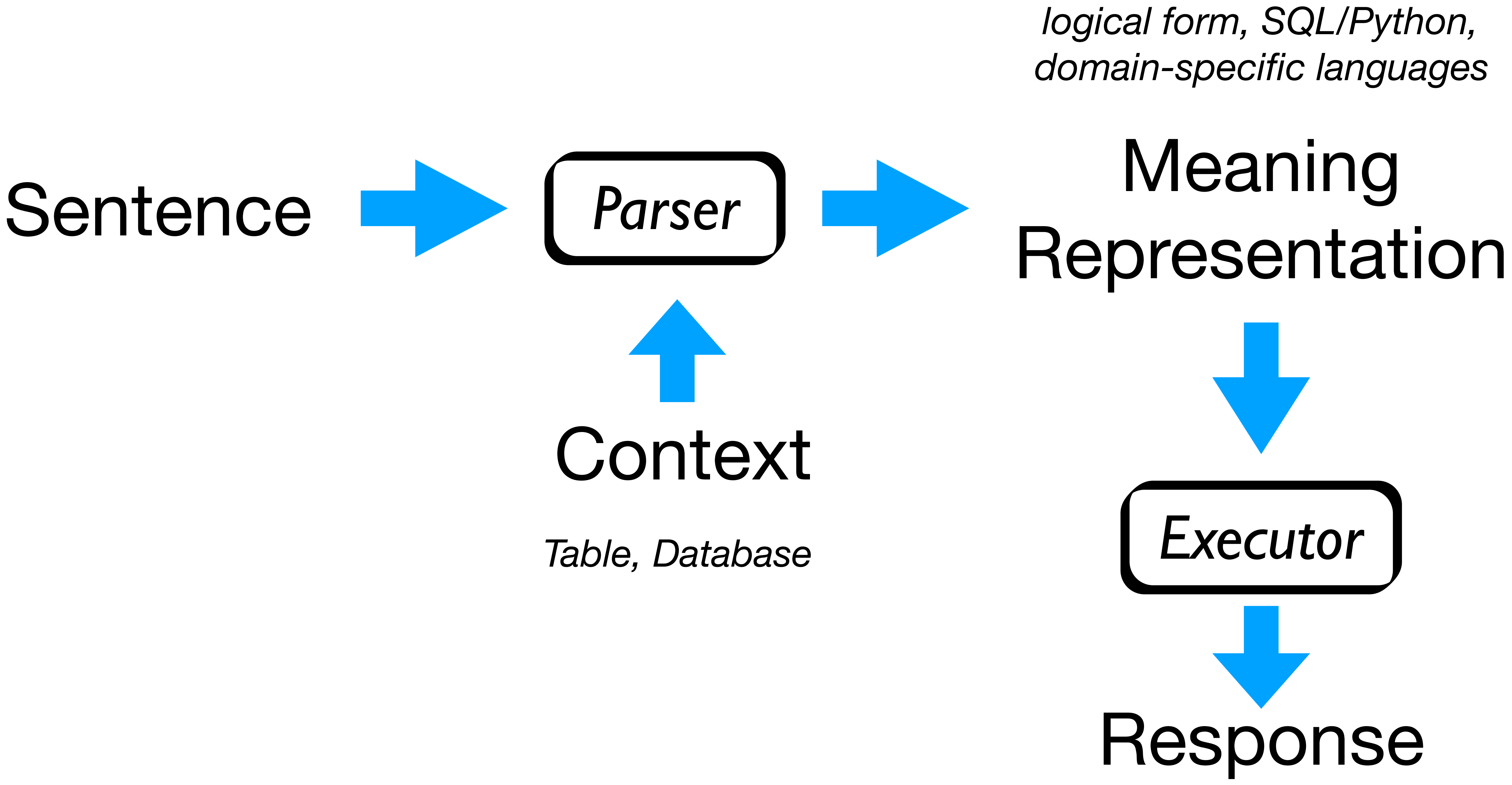
Parser

Meaning Representation

Context

Table, Database

Background



Sentence

Parser

Context

Table, Database

logical form, SQL/Python, domain-specific languages

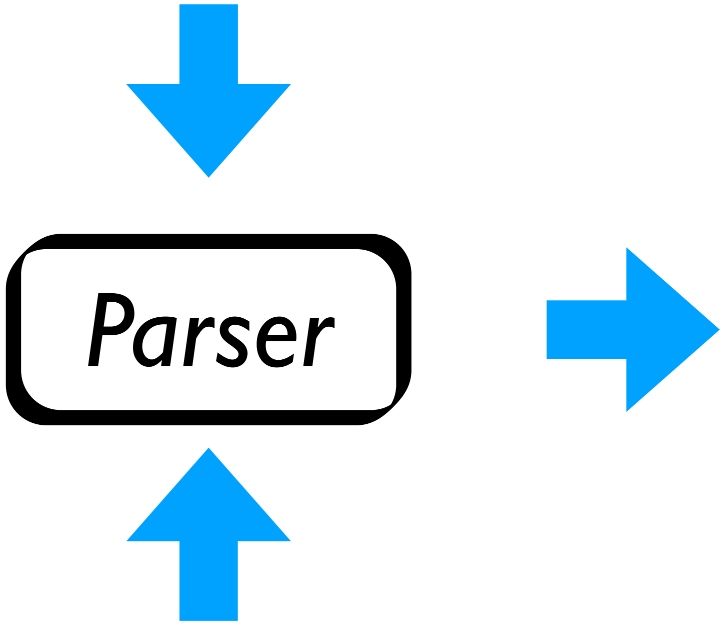
Meaning Representation

Executor

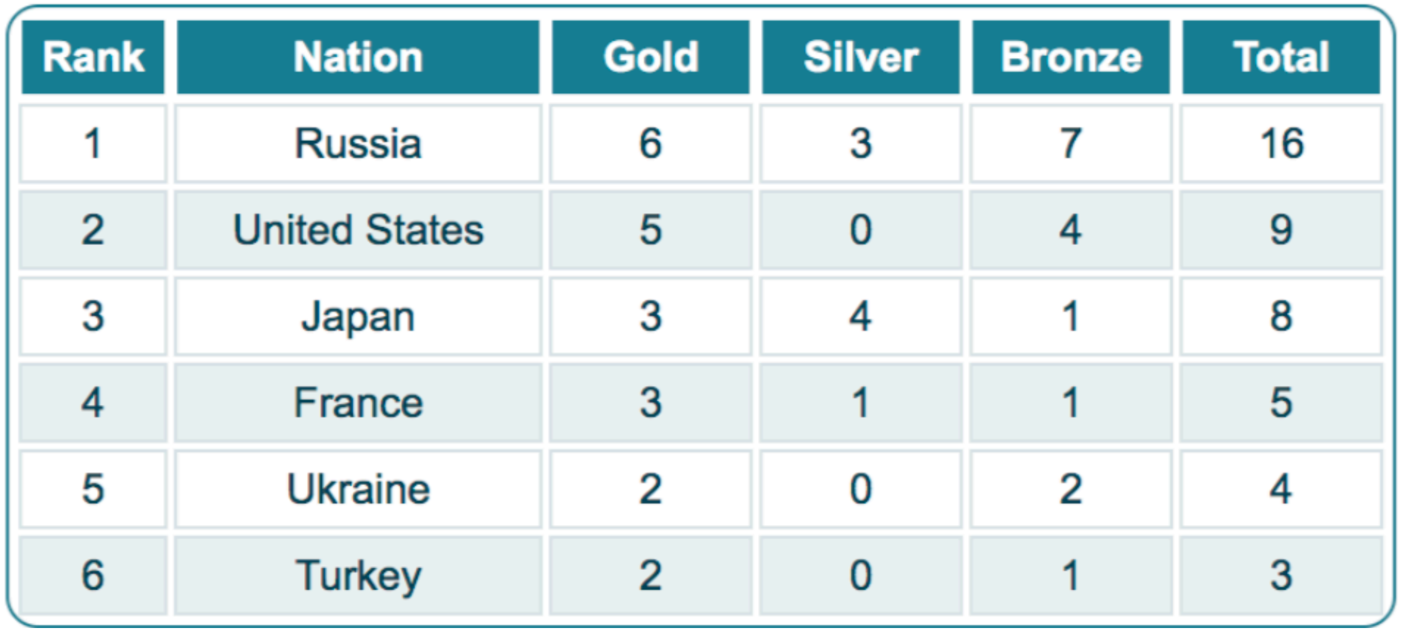
Response

Executable Semantic Parsing

how many silver medals did the nation of Turkey win?



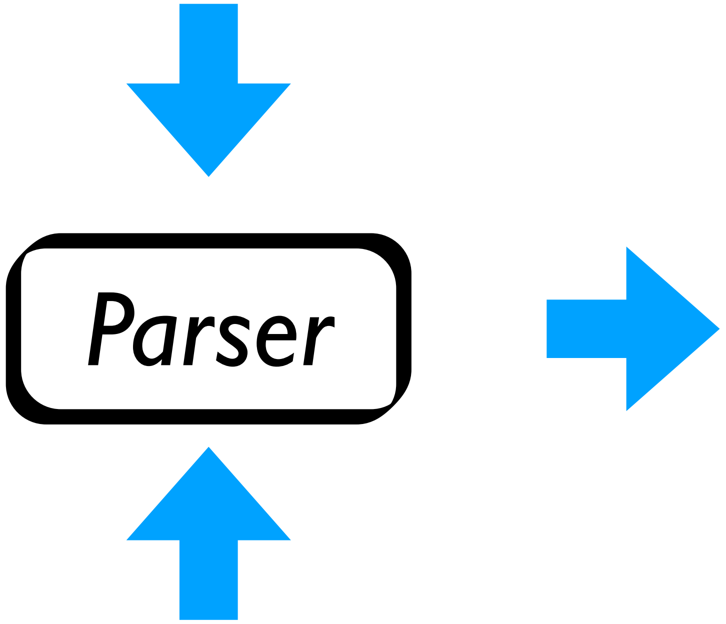
select([nation =“Turkey”], silver)



Rank	Nation	Gold	Silver	Bronze	Total
1	Russia	6	3	7	16
2	United States	5	0	4	9
3	Japan	3	4	1	8
4	France	3	1	1	5
5	Ukraine	2	0	2	4
6	Turkey	2	0	1	3

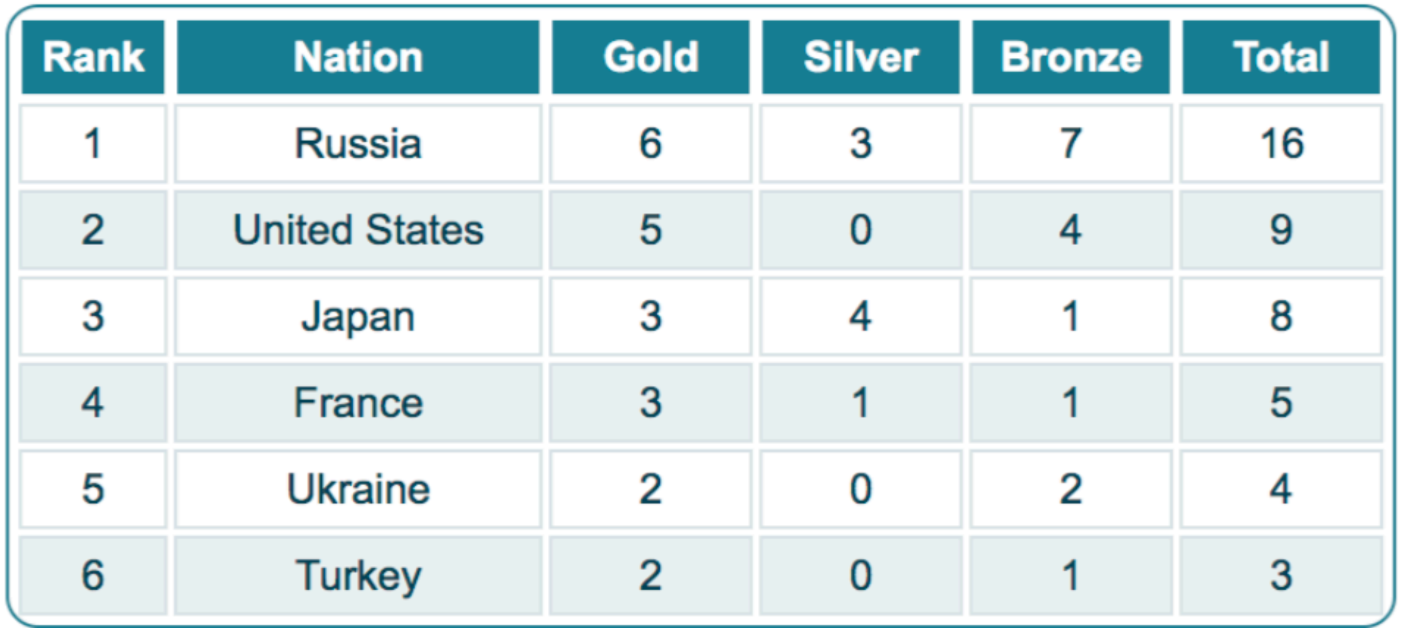
Executable Semantic Parsing

how many silver medals did the nation of Turkey win?



Parser

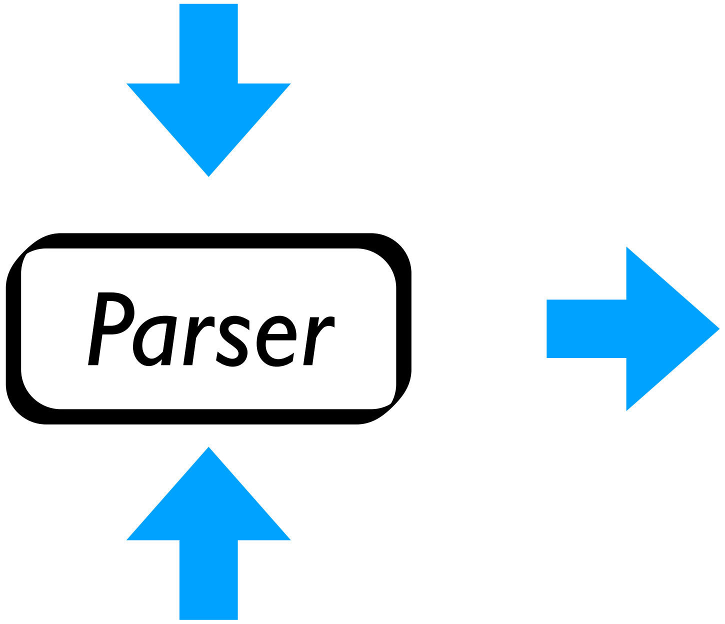
[nation =“Turkey”]



Rank	Nation	Gold	Silver	Bronze	Total
1	Russia	6	3	7	16
2	United States	5	0	4	9
3	Japan	3	4	1	8
4	France	3	1	1	5
5	Ukraine	2	0	2	4
6	Turkey	2	0	1	3

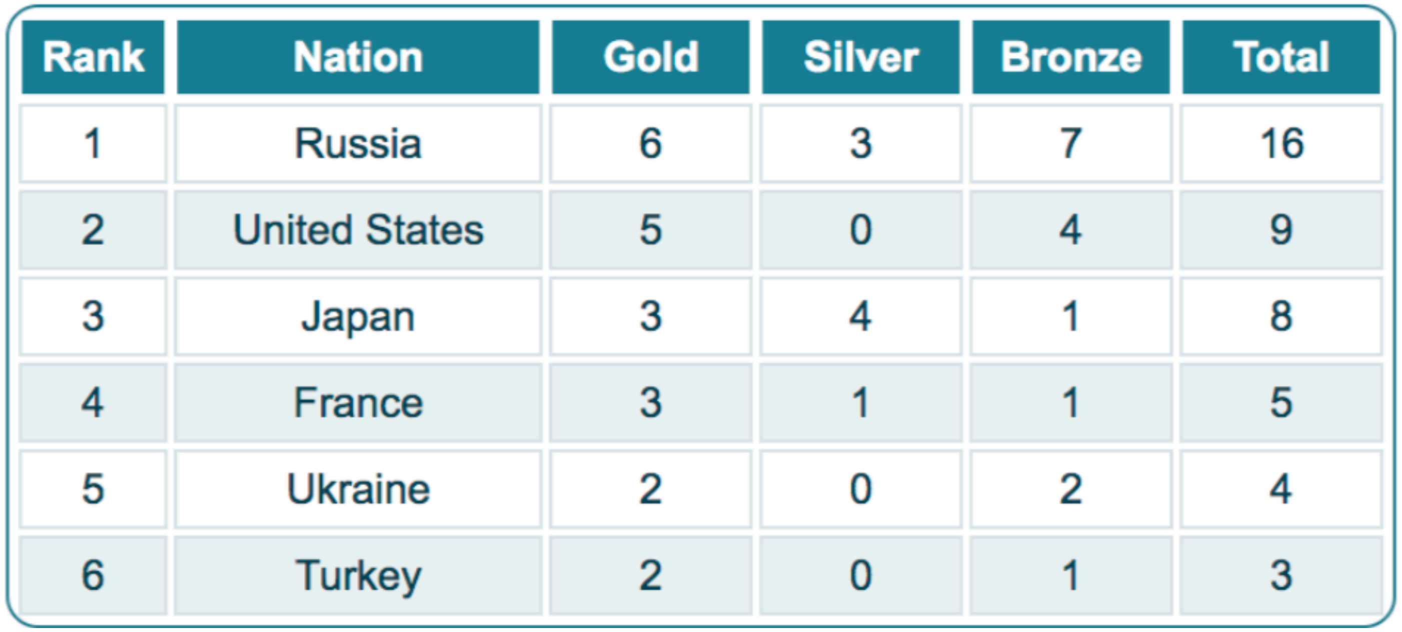
Executable Semantic Parsing

how many silver medals did the nation of Turkey win?



Parser

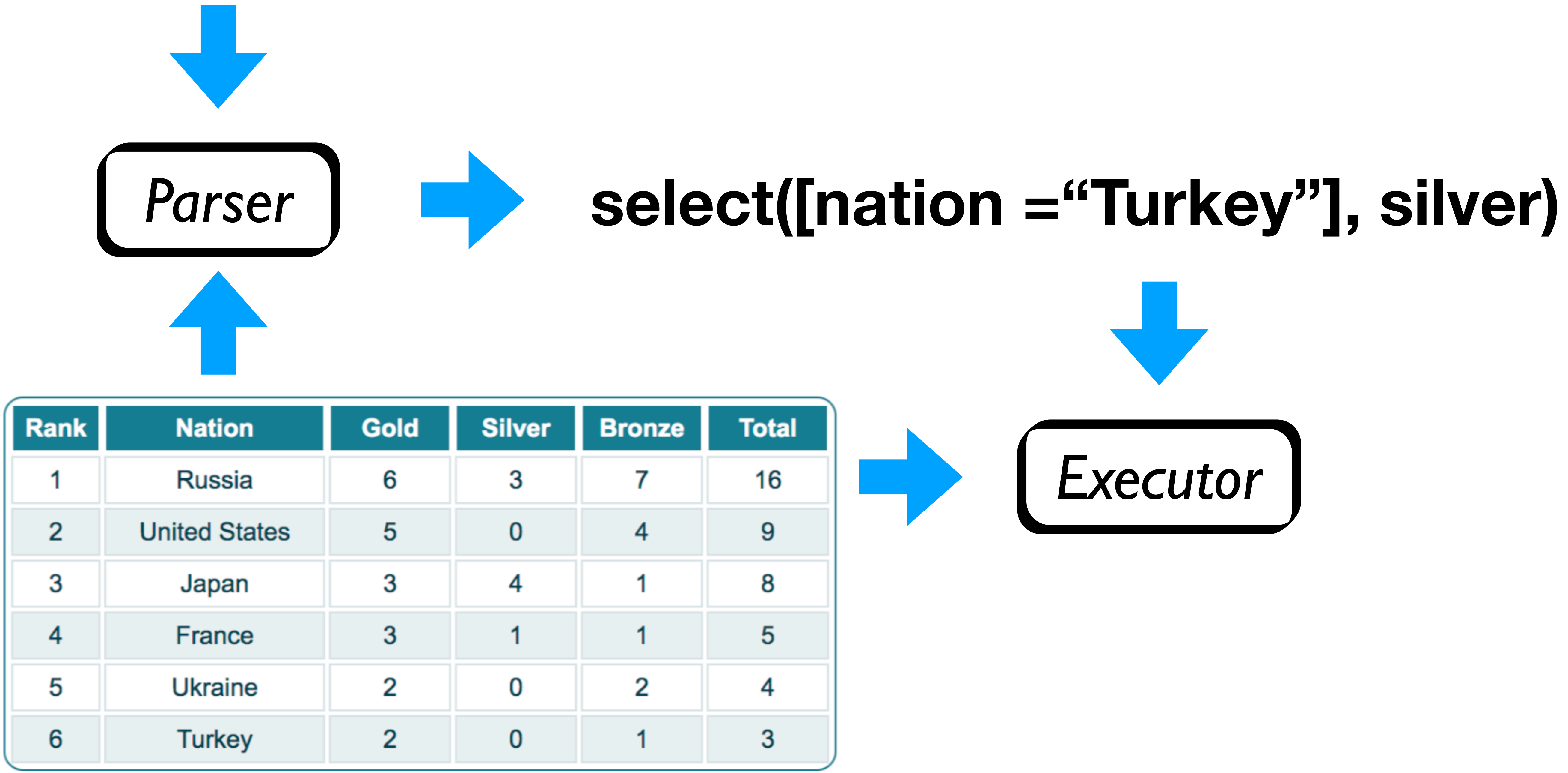
select([nation =“Turkey”], silver)



Rank	Nation	Gold	Silver	Bronze	Total
1	Russia	6	3	7	16
2	United States	5	0	4	9
3	Japan	3	4	1	8
4	France	3	1	1	5
5	Ukraine	2	0	2	4
6	Turkey	2	0	1	3

Executable Semantic Parsing

how many silver medals did the nation of Turkey win?

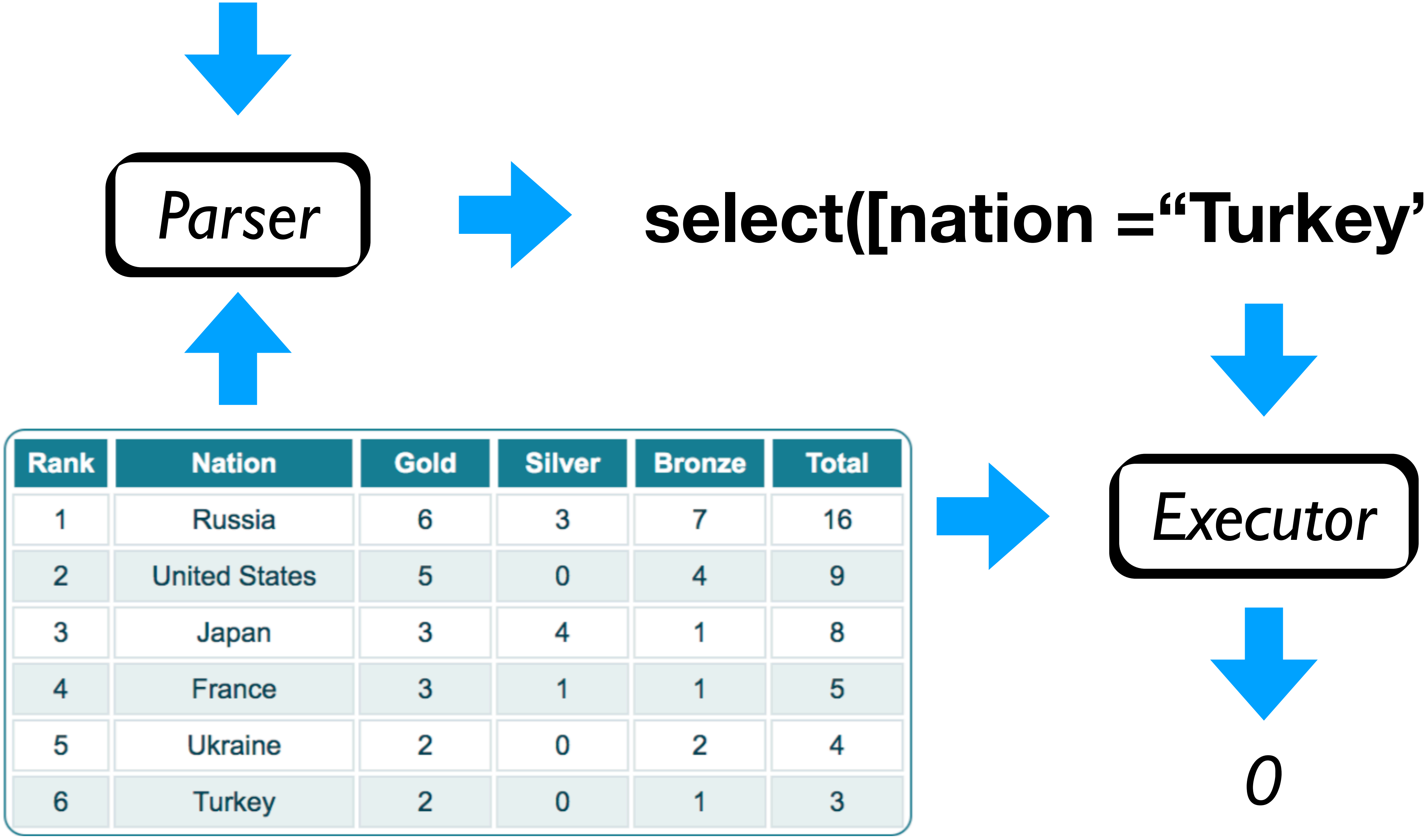


select([nation =“Turkey”], silver)

Rank	Nation	Gold	Silver	Bronze	Total
1	Russia	6	3	7	16
2	United States	5	0	4	9
3	Japan	3	4	1	8
4	France	3	1	1	5
5	Ukraine	2	0	2	4
6	Turkey	2	0	1	3

Executable Semantic Parsing

how many silver medals did the nation of Turkey win?



Parser

select([nation =“Turkey”], silver)

Executor

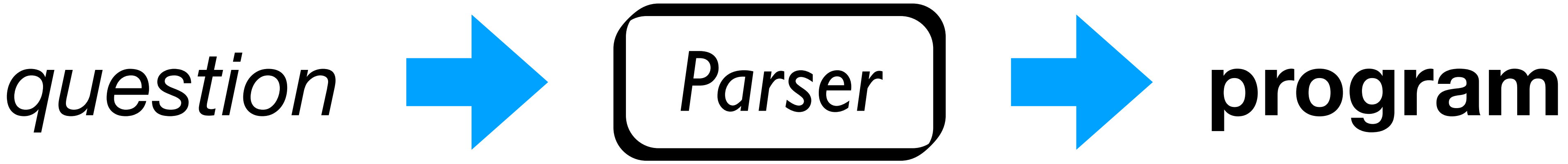
Rank	Nation	Gold	Silver	Bronze	Total
1	Russia	6	3	7	16
2	United States	5	0	4	9
3	Japan	3	4	1	8
4	France	3	1	1	5
5	Ukraine	2	0	2	4
6	Turkey	2	0	1	3

0

Question-Program Pairs

how many silver medals did the nation of Turkey win?

select([nation =“Turkey”], silver)



question *Parser* **program**

[Zettlemoyer and Collins, 2005; Wong and Mooney, 2007; Lu et al., 2008; Jia and Liang, 2016]

Question-Program Pairs

how many silver medals did the nation of Turkey win?

select([nation =“Turkey”], silver)

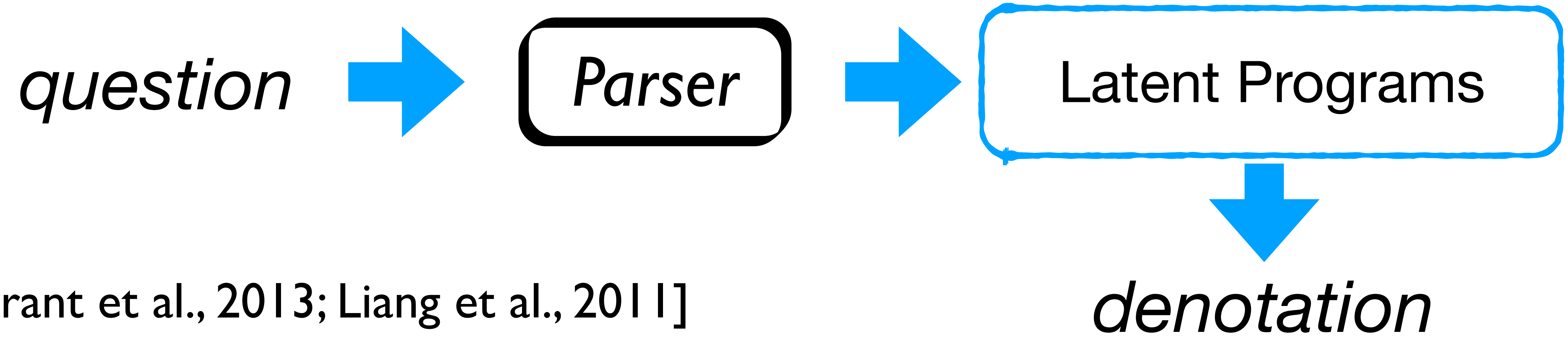
Expensive to collect!

[Zettlemoyer and Collins, 2005; Wong and Mooney, 2007; Lu et al., 2008; Jia and Liang, 2016]

Question-Denotation Pairs

how many silver medals did the nation of Turkey win?

Denotation: 0



question

Parser

Latent Programs

denotation

[Berant et al., 2013; Liang et al., 2011]

Structure

- **Challenges with learning from denotation**
- Our approach
- Experiments and analysis

Large Search Space

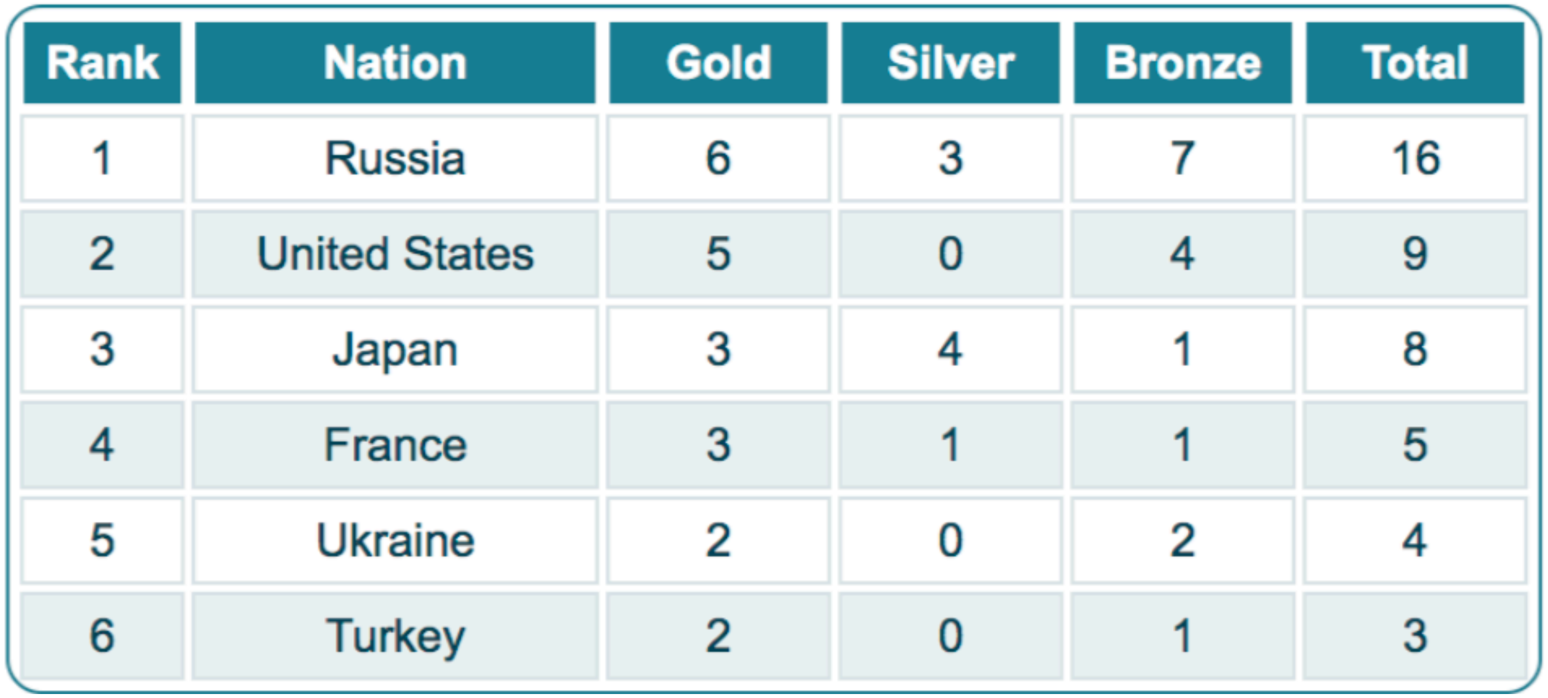
how many silver medals did the nation of Turkey win?

Program		***Denotation***	
count([nation =“Turkey”])	→	**1**	✗
sum(all_rows, silver)	→	**8**	✗
select([nation =“Turkey”], silver)	→	**0**	✓
select([nation =“Turkey”], nation)	→	**Turkey**	✗
select(argmax([all_rows], silver), silver)	→	**4**	✗

⋮

Spurious Programs

how many silver medals did the nation of Turkey win?



Rank	Nation	Gold	Silver	Bronze	Total
1	Russia	6	3	7	16
2	United States	5	0	4	9
3	Japan	3	4	1	8
4	France	3	1	1	5
5	Ukraine	2	0	2	4
6	Turkey	2	0	1	3

Consistent Programs

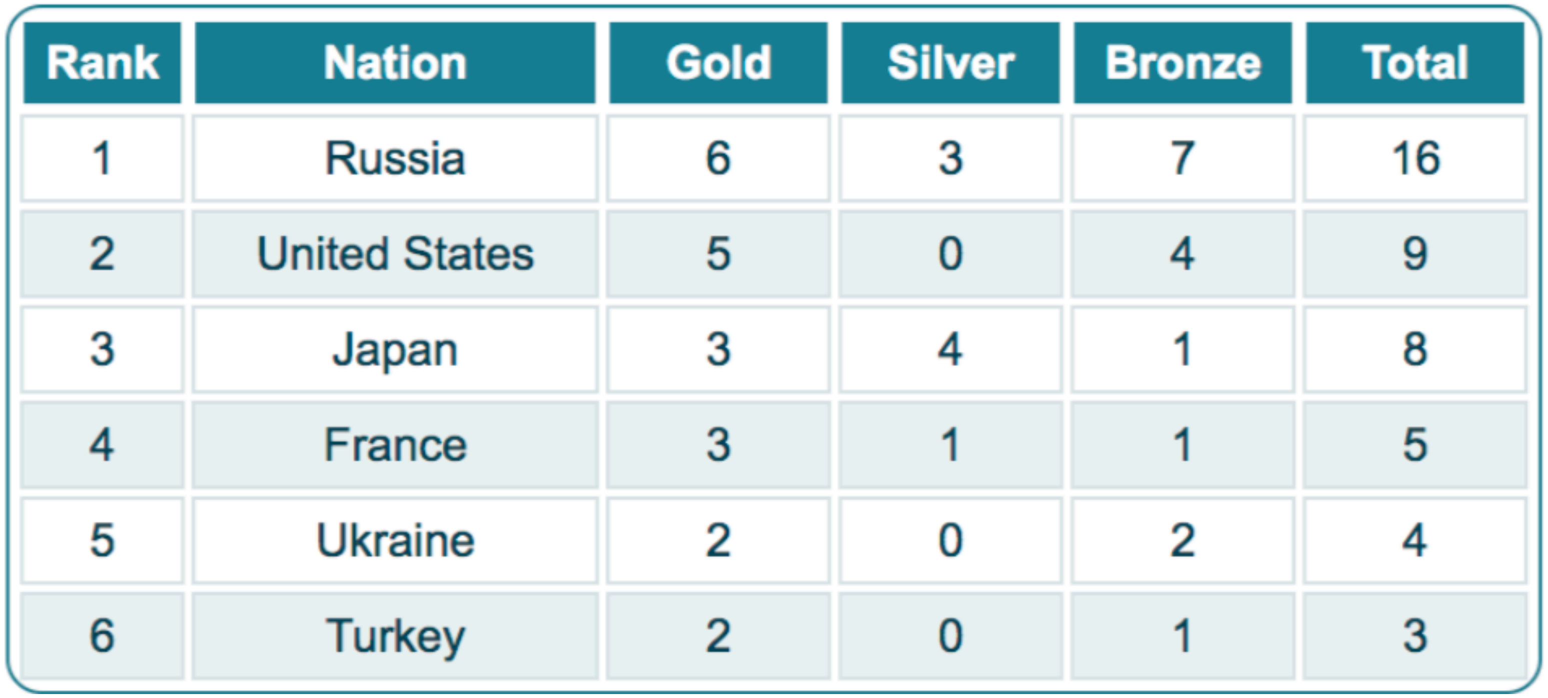
select([nation =“Turkey”], silver) → **0** ✓

select(previous(argmax([all_rows], silver), silver) → **0** ✓

select(argmin([all_rows], silver))), silver) → **0** ✓

Spurious Programs

how many silver medals did the nation of Turkey win?



Rank	Nation	Gold	Silver	Bronze	Total
1	Russia	6	3	7	16
2	United States	5	0	4	9
3	Japan	3	4	1	8
4	France	3	1	1	5
5	Ukraine	2	0	2	4
6	Turkey	2	0	1	3

Consistent Programs = **Spurious Programs**

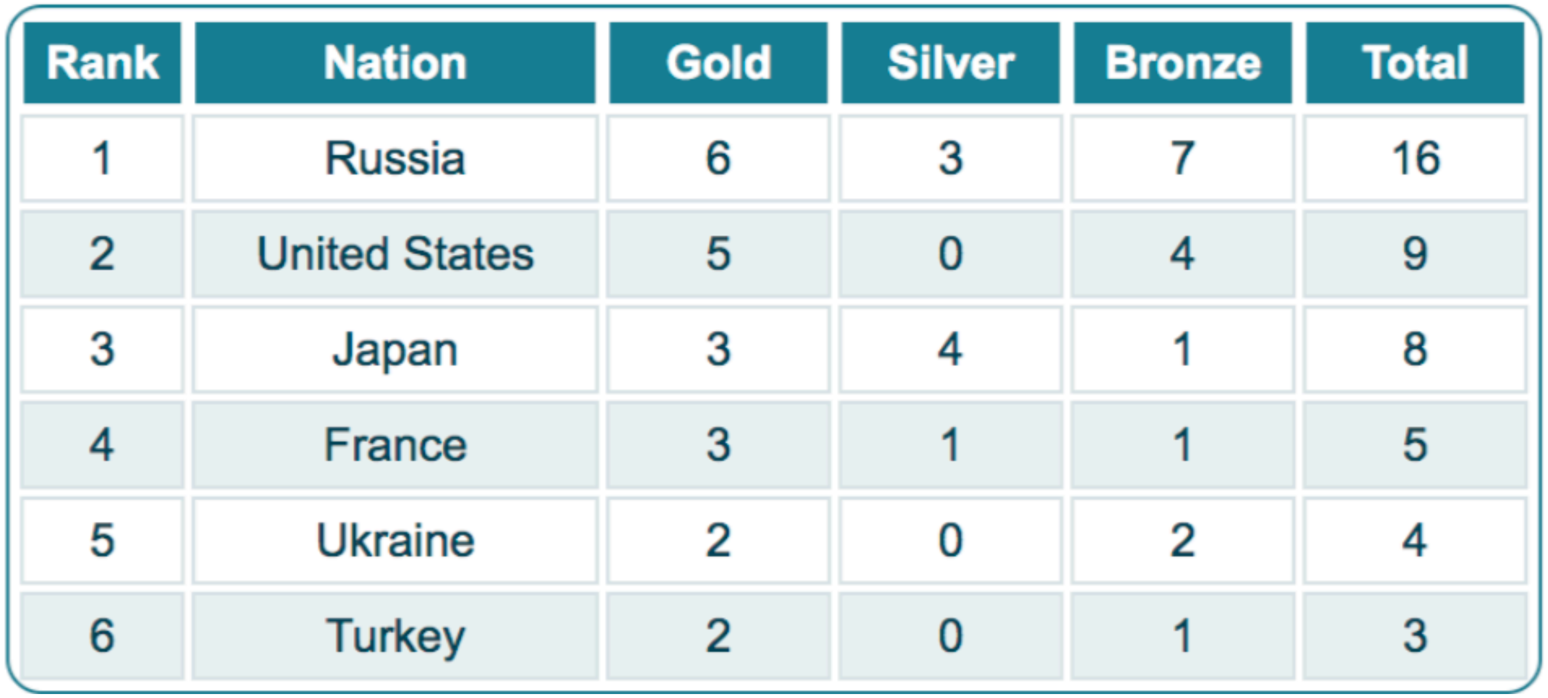
select([nation ="Turkey"], silver) ⟶ **0** ✓

select(previous(argmax([all_rows], silver), silver) ⟶ **0** ✓

select(argmin([all_rows], silver))), silver) ⟶ **0** ✓

Spurious Programs

how many silver medals did the nation of Turkey win?



Rank	Nation	Gold	Silver	Bronze	Total
1	Russia	6	3	7	16
2	United States	5	0	4	9
3	Japan	3	4	1	8
4	France	3	1	1	5
5	Ukraine	2	0	2	4
6	Turkey	2	0	1	3

Consistent Programs = Spurious Programs + Correct Programs

select([nation =“Turkey”], silver) → 0 ✓

select(previous(argmax([all_rows], silver), silver) → 0 ✓

select(argmin([all_rows], silver))), silver) → 0 ✓

Spurious Programs

Learning Objective:

p_θ**(Consistent Programs | Question, Context)**

||

p_θ **(Spurious Programs|…)** **+** p_θ **(Correct Programs|…)**

Spurious Programs

Learning Objective:

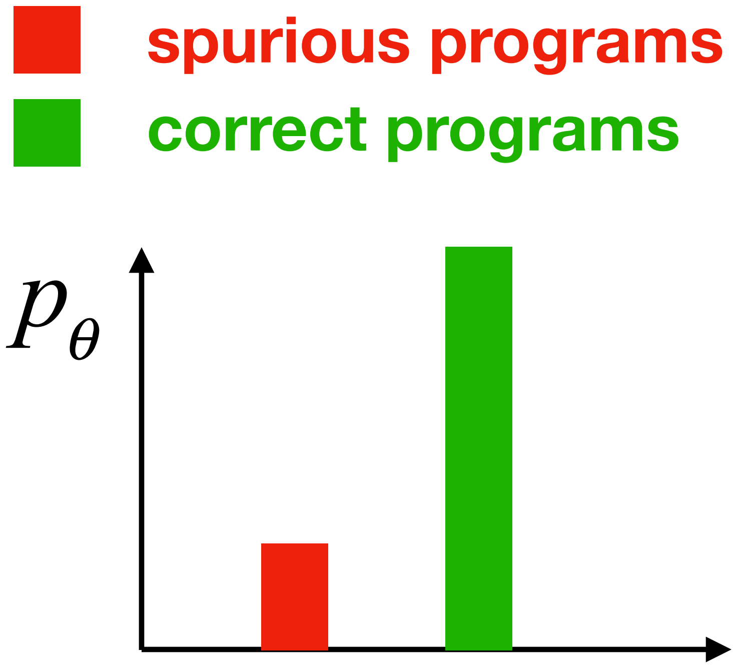
$$p_\theta(\textbf{Consistent Programs} \mid \textbf{Question, Context})$$

$$\parallel$$

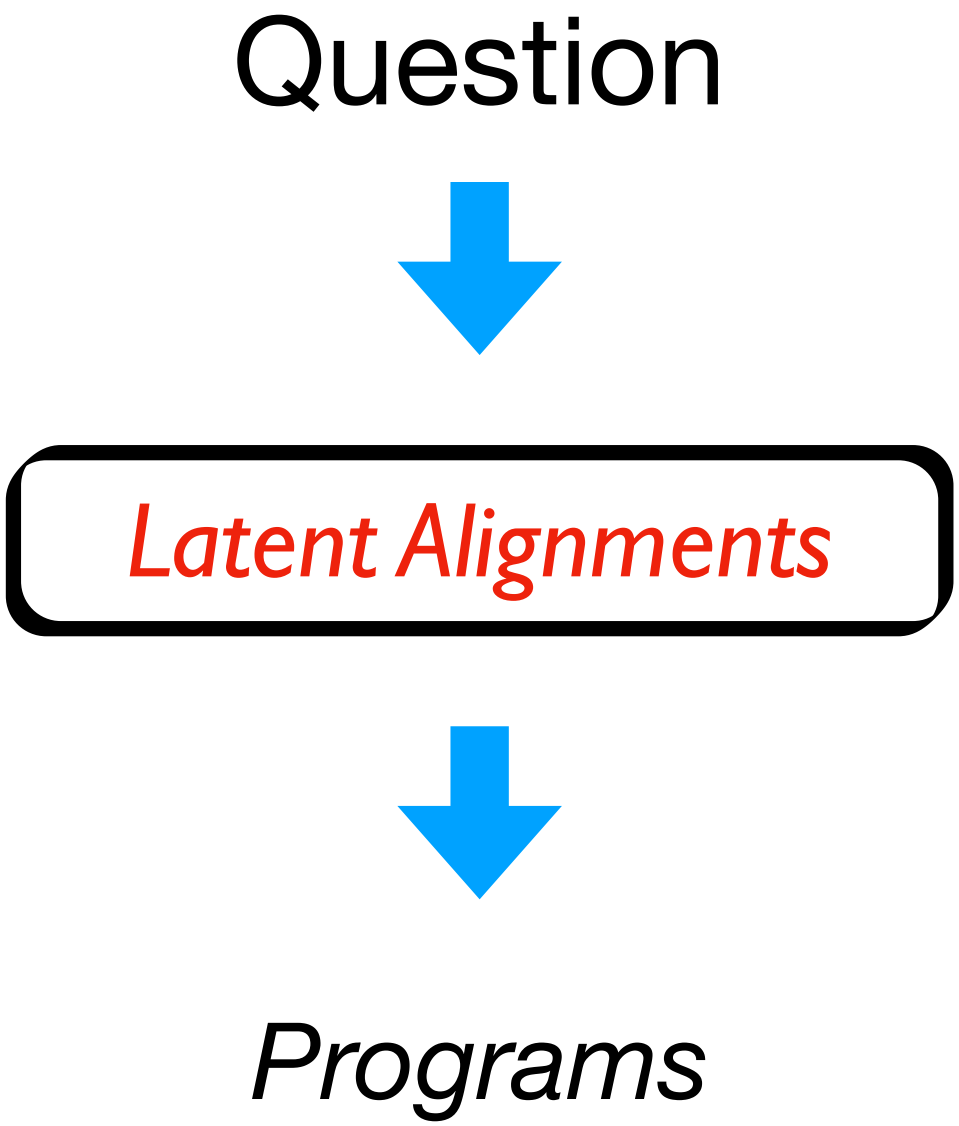
$$\underbrace{p_\theta(\textbf{Spurious Programs}|\ldots)}_{\textit{harmful !}} + p_\theta(\textbf{Correct Programs}|\ldots)$$

Our Approach

Inductive Bias



Latent Alignments

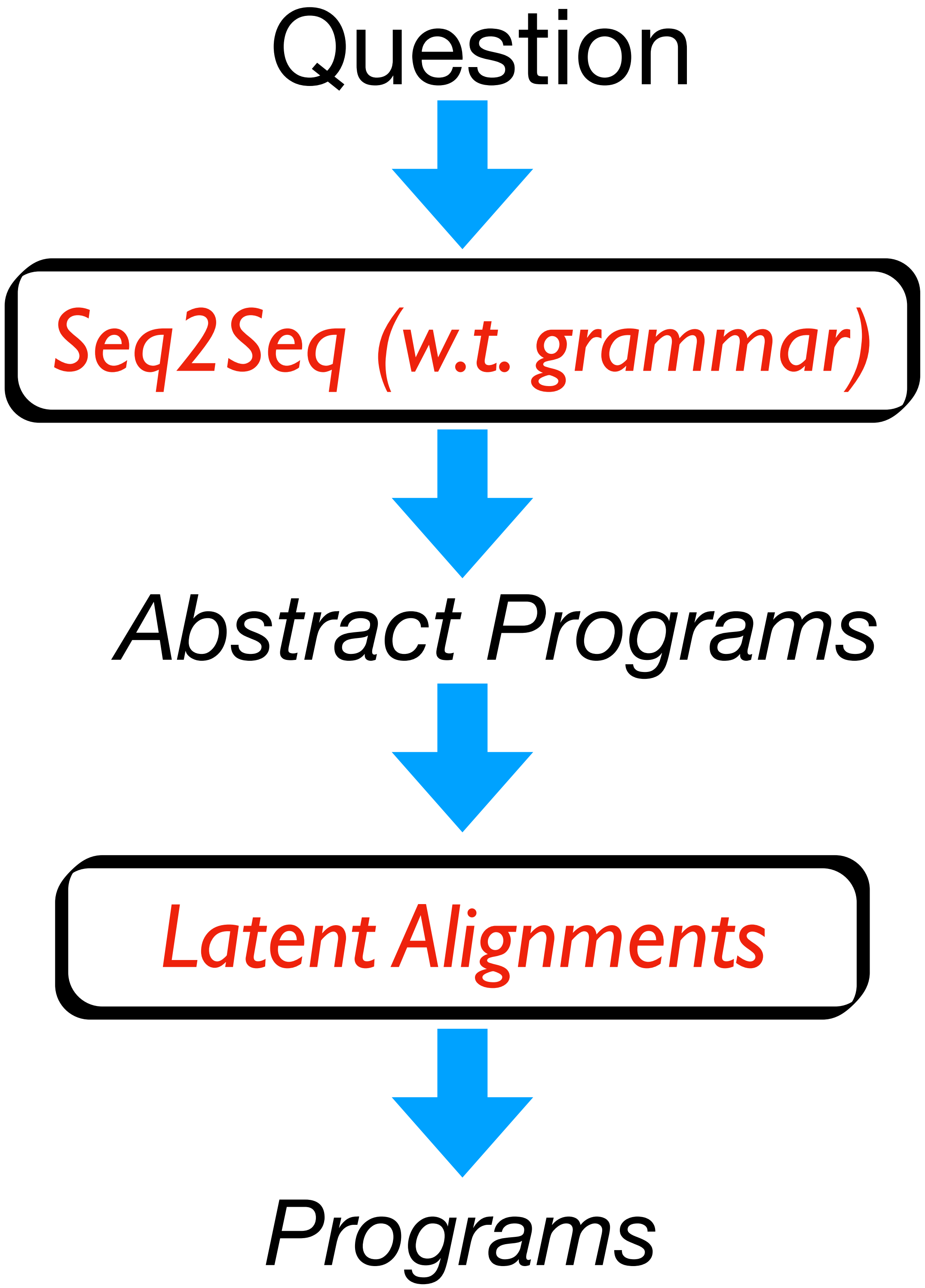


Question

Latent Alignments

Programs

Two-Stage Search



Question

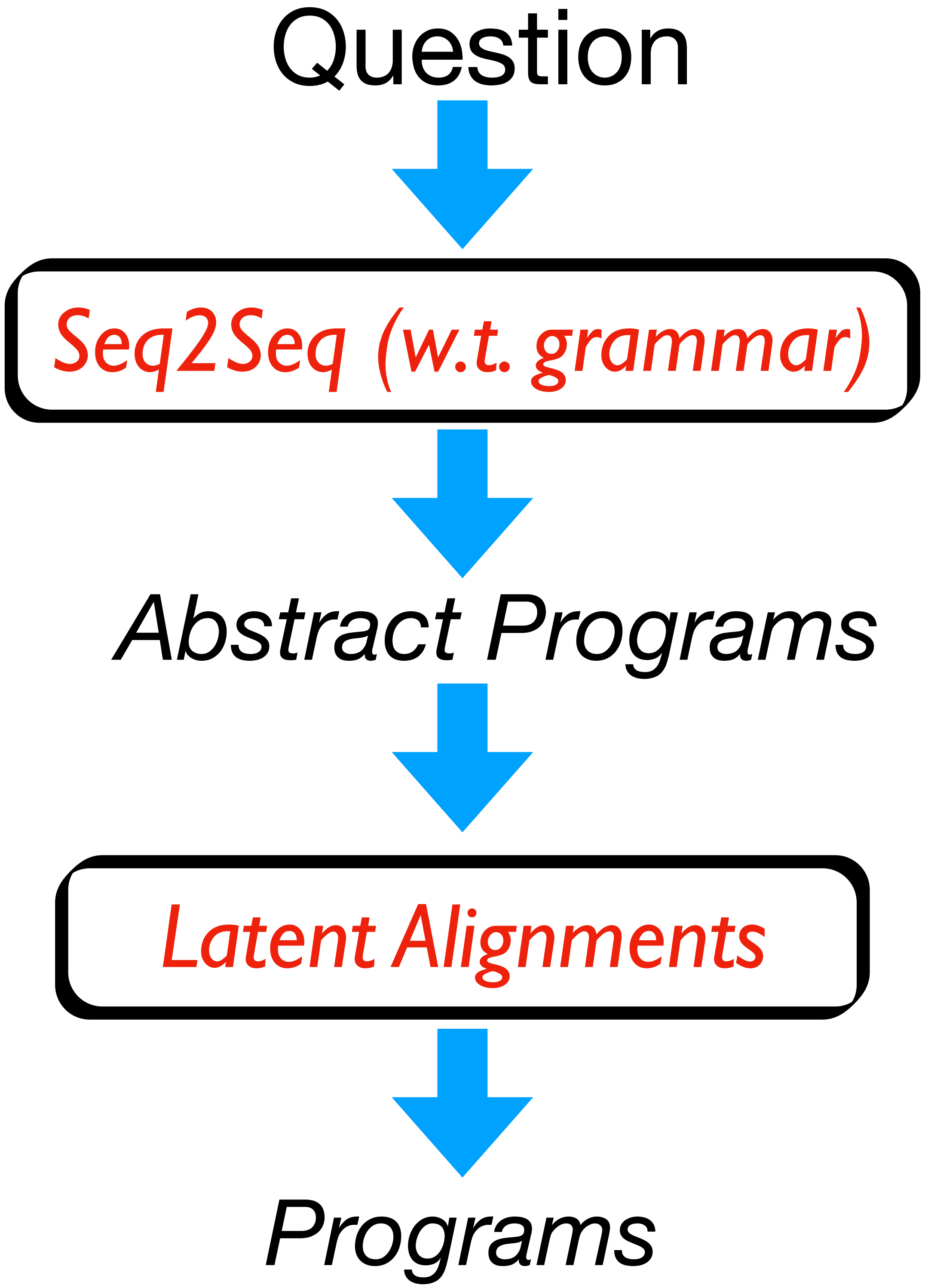
Seq2Seq (w.t. grammar)

Abstract Programs

Latent Alignments

Programs

Two-Stage Search



Question

Seq2Seq (w.t. grammar)

Abstract Programs

Latent Alignments

Programs

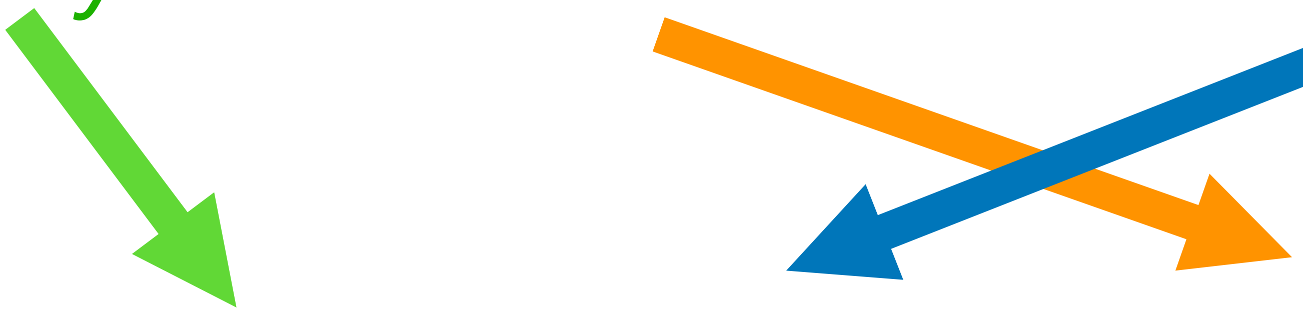
how many silver medals did the nation of Turkey win?

e.g. **select(***#rows***),** *#column***)**

e.g. **select([nation ="Turkey"], silver)**

Observation: Need Alignments

One-to-one alignment

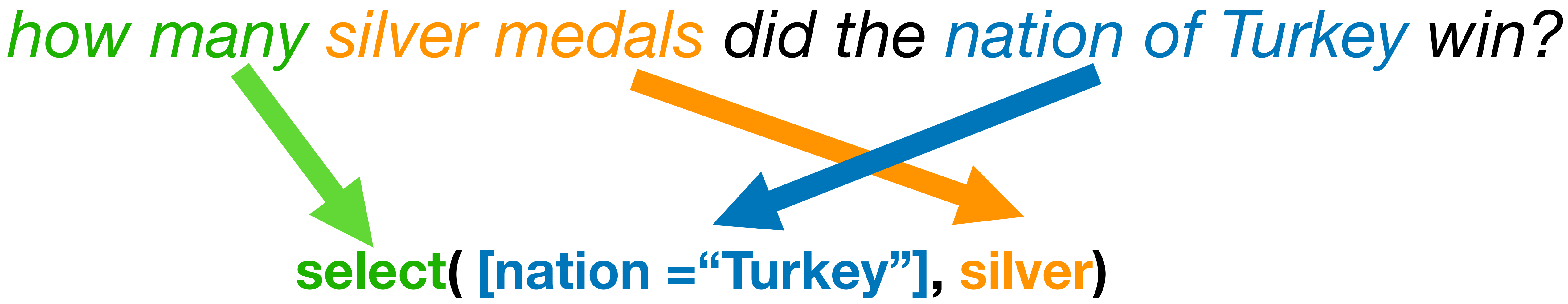


how many silver medals did the nation of Turkey win?

select([nation =“Turkey”], silver)

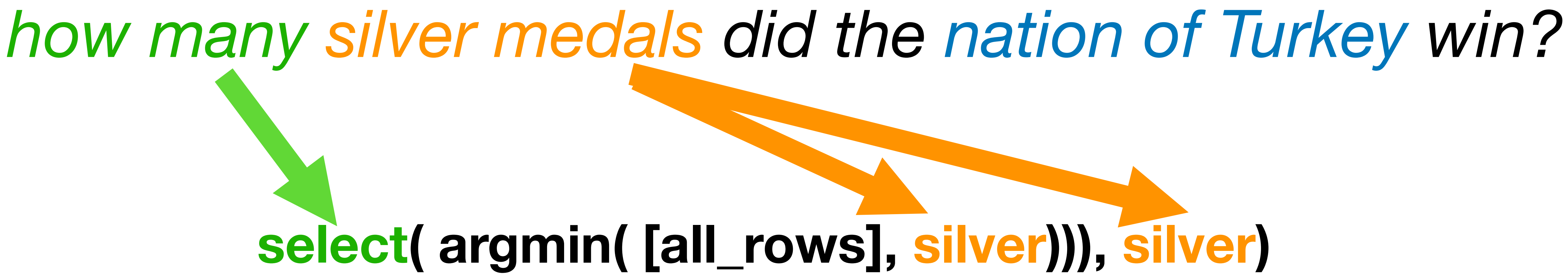
Observation: Need Alignments

One-to-one alignment



how many silver medals did the nation of Turkey win?

select([nation =“Turkey”], silver)



how many silver medals did the nation of Turkey win?

select(argmin([all_rows], silver))), silver)

Alignment

SCFG [Wong and Mooney, 2007]
CCG [Zettlemoyer & Collins, 2007]
Hybrid Tree [Lu et al., 2008]

- Cover alignments
- Full correspondence are expensive to model

Alignment

SCFG [Wong and Mooney, 2007]
CCG [Zettlemoyer & Collins, 2007]
Hybrid Tree [Lu et al., 2008]

- Cover alignments
- Full correspondence are expensive to model

Seq2Seq [Dong & Lapata, 2016]

- No structural alignments
- Efficient

Alignment

SCFG [Wong and Mooney, 2007]
CCG [Zettlemoyer & Collins, 2007]
Hybrid Tree [Lu et al., 2008]

- Cover alignments
- Full correspondence are expensive to model

Seq2Seq [Dong & Lapata, 2016]

- No structural alignments
- Efficient

Get the best of both worlds?

Alignment

SCFG [Wong and Mooney, 2007]
CCG [Zettlemoyer & Collins, 2007]
Hybrid Tree [Lu et al., 2008]

- Cover alignments
- Full correspondence are expensive to model

Seq2Seq [Dong & Lapata, 2016]

- No structural alignments
- Efficient

Get the best of both worlds?

Only consider alignments of low-level operations

Decomposition

Program:

Abstract Program + *Instantiation*

(high-level operations and slots)

(low-level operations)

select(*#rows***,** *#column***)**

select([rank = 2], total)

select([nation =“Turkey”], silver)

select([nation =“United States”], gold)

Abstract Programs

```
select(#rows, #column)

select(argmax( #rows),
  #column), #column)
select(argmin( #rows),
  #column), #column)

⋮
```

- Retain high-level operations
- Generalise across different context

[Dong and Lapata, 2018; Catherine Finegan-Dollak and Radev, 2018; Goldman et al., 2018; Herzig and Berant, 2018; Nye et al., 2019]

Abstract Programs

select(*#rows***,** *#column***)**

select(argmax(*#rows***),** *#column***),** *#column***)**

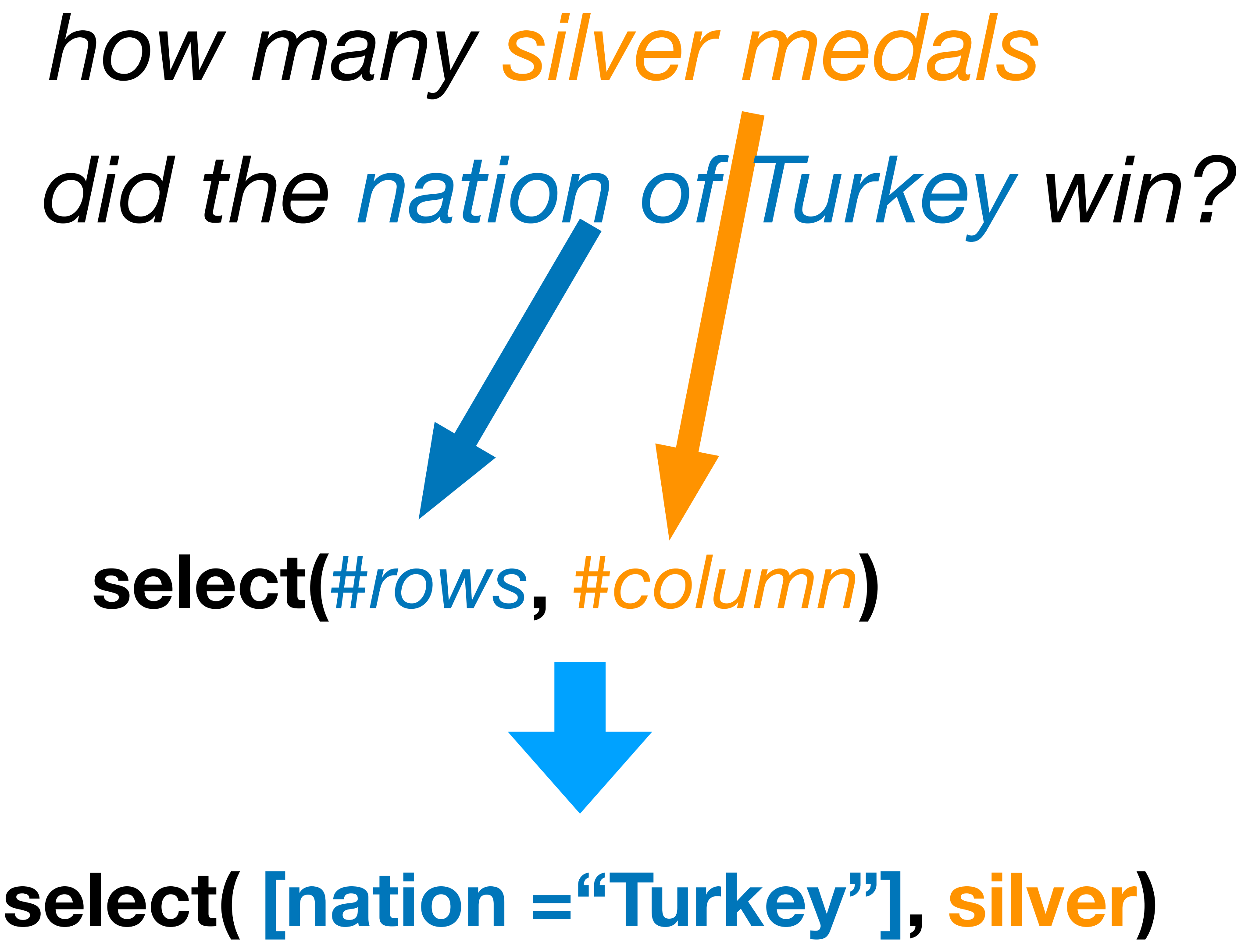
select(argmin(*#rows***),** *#column***),** *#column***)**

⋮

- Retain high-level operations
- Generalise across different context
- Facilitate tractable probabilistic alignment model

[Dong and Lapata, 2018; Catherine Finegan-Dollak and Radev, 2018; Goldman et al., 2018; Herzig and Berant, 2018; Nye et al., 2019]

Model Alignments



Question

how many silver medals did the nation of Turkey win?

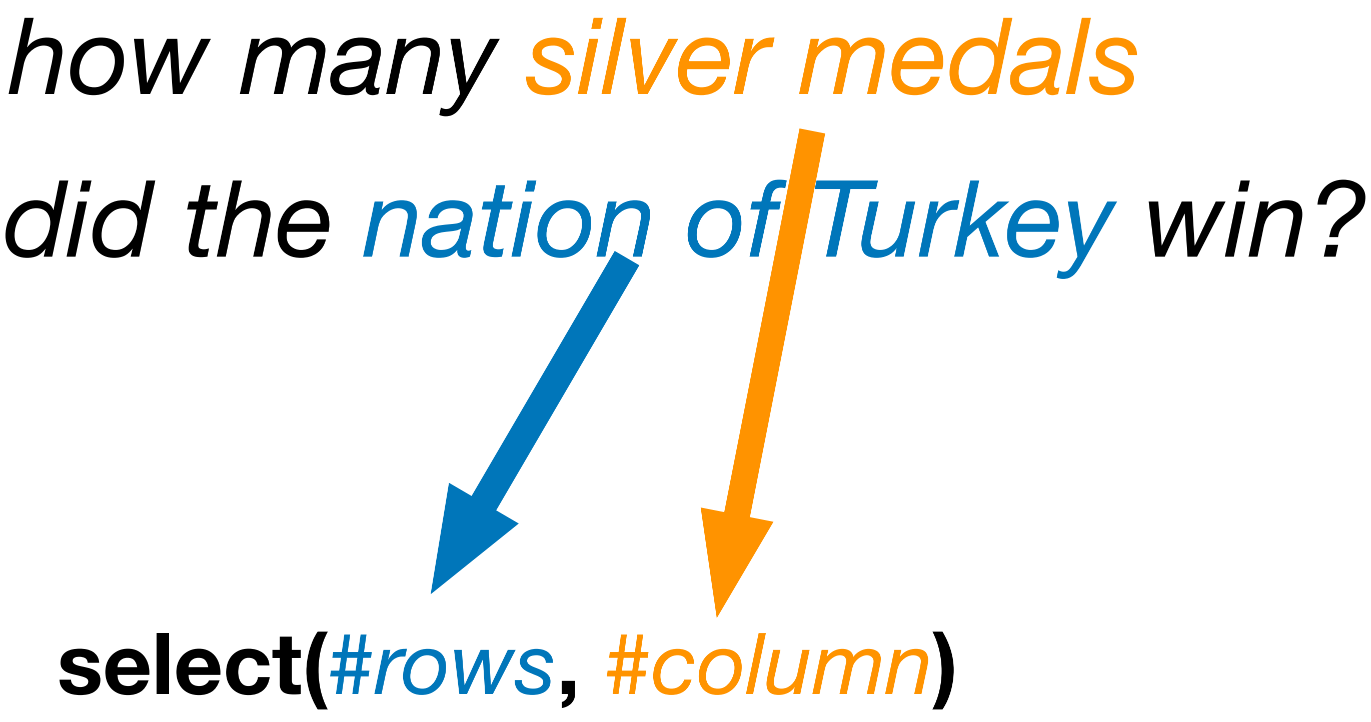
Abstract Program

select(*#rows*, *#column***)**

Program

select([nation =“Turkey”], silver)

Structural Alignment



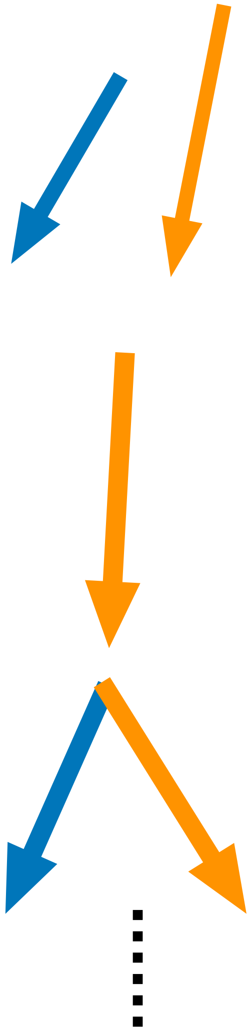
how many silver medals did the nation of Turkey win?

select(*#rows***,** *#column***)**

- One-to-one alignments between spans and slots
- Marginal probability can be computed with dynamic programming [Tackstrom et al. 2015]

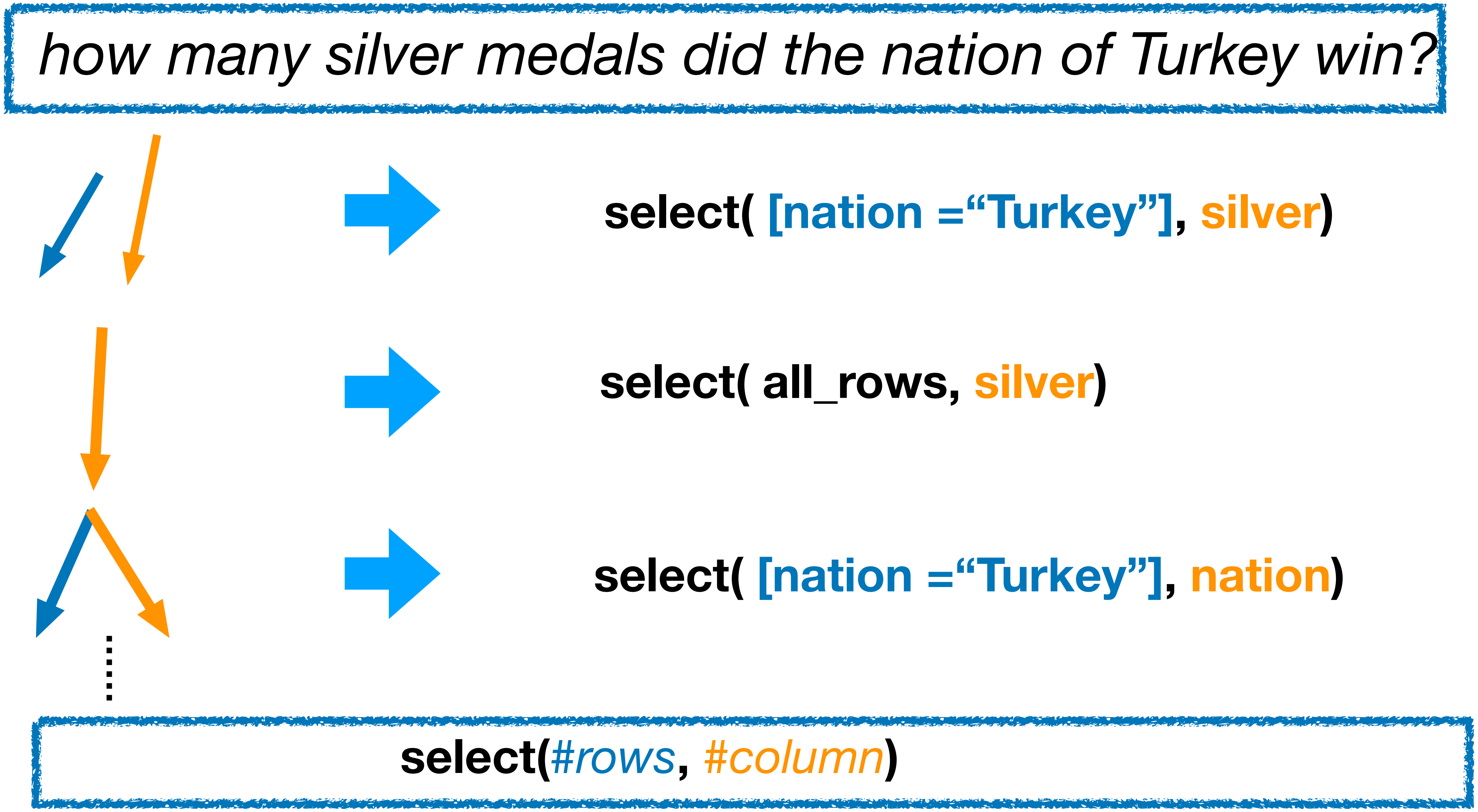
Structural Alignment

how many silver medals did the nation of Turkey win?



select(*#rows***,** *#column***)**

Structural Alignment



how many silver medals did the nation of Turkey win?

select([nation =“Turkey”], silver)

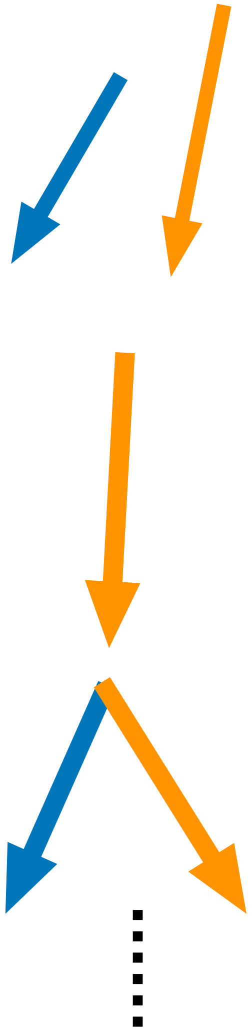
select(all_rows, silver)

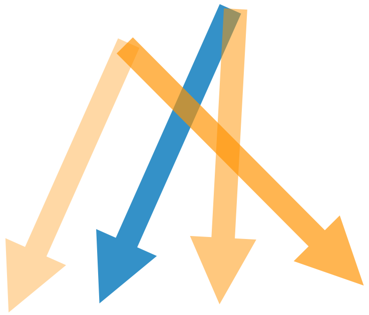
select([nation =“Turkey”], nation)

select(*#rows*, *#column*)

Structural Alignment

how many silver medals did the nation of Turkey win?





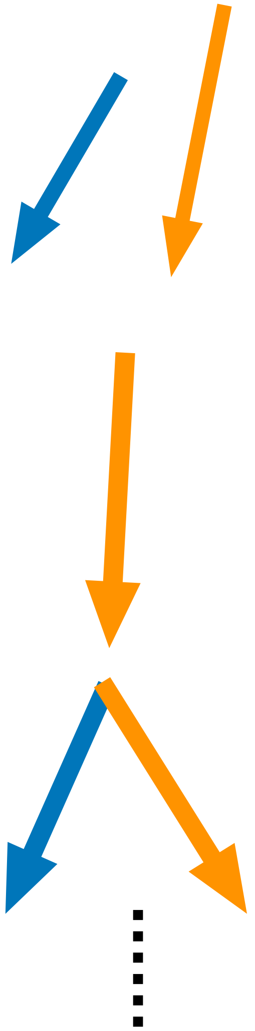
Expected Alignment
(Continuous)

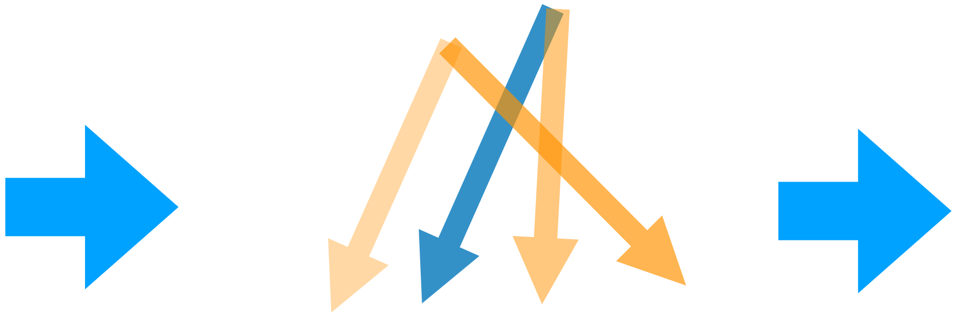
select([nation =“Turkey”],
silver)

select(*#rows*, *#column*)

Structural Alignment

how many silver medals did the nation of Turkey win?





select(**[nation =“Turkey”],**
silver)

Expected Alignment (Continuous)

structured attention [Kim et al., 2017]

select(*#rows*, *#column*)

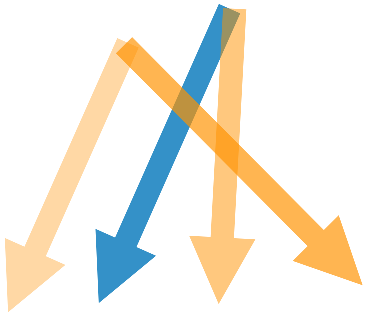
Summary

- A two-stage approach that incorporates:

1. abstract programs

select(*#rows*, *#column*)

2. (expected) structured alignments



Experiments

Datasets

	WikiTableQuestions	WikiSQL
# of question-denotation pairs	18,496	80,654
# of tables	2,018	24,241

- Questions involve a variety of operations
- Unseen tables at ***test*** time

Consistent Programs

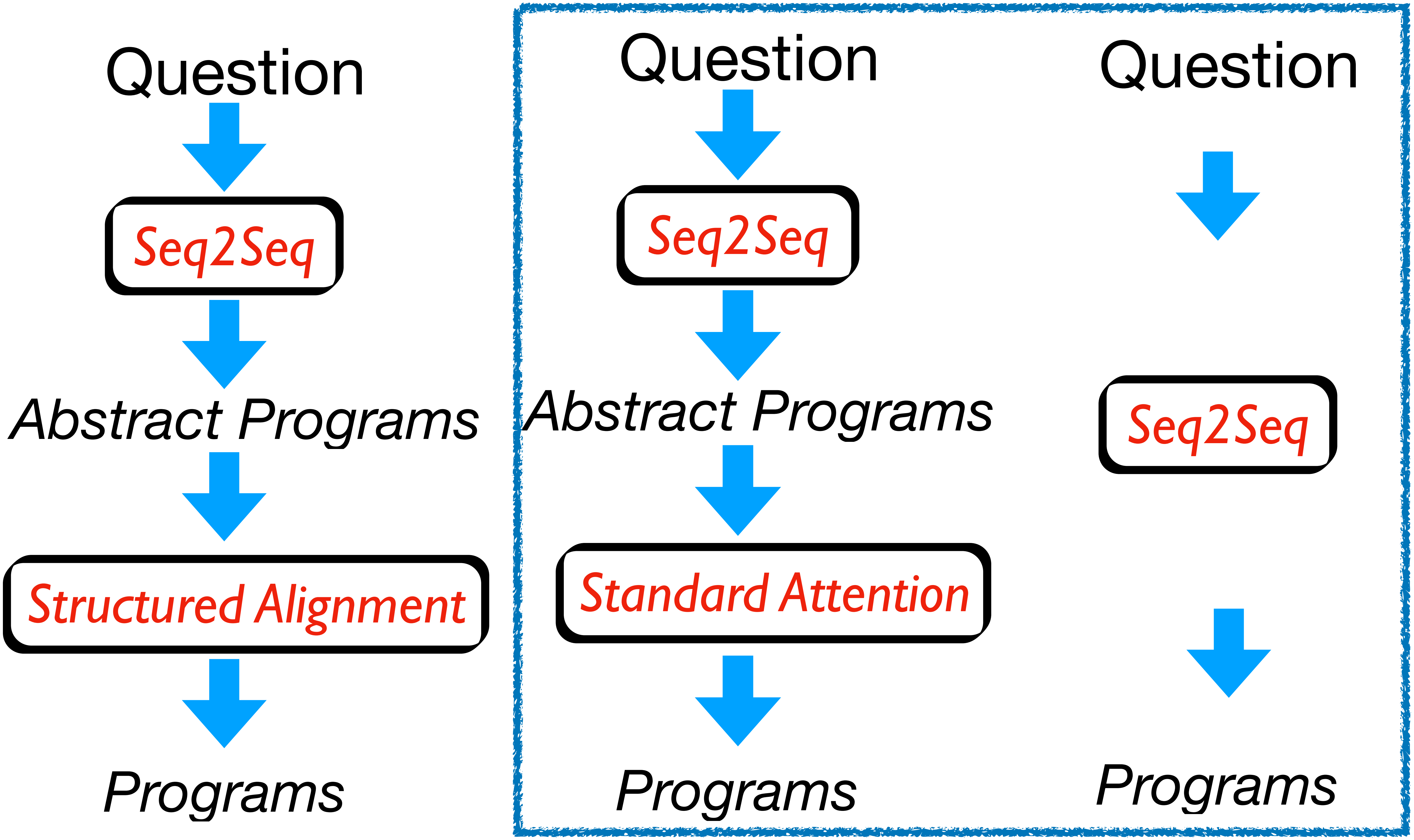
	WikiTableQuestions	WikiSQL
# of consistent programs per question	200	84

Most consistent programs are spurious!

Training and Inference

- Train: ***marginal probability*** of consistent programs
- Inference: ***beam search*** (for abstract programs)
- Evaluation: ***execution accuracy***

Baselines



Question

Seq2Seq

Abstract Programs

Structured Alignment

Programs

Question

Seq2Seq

Abstract Programs

Standard Attention

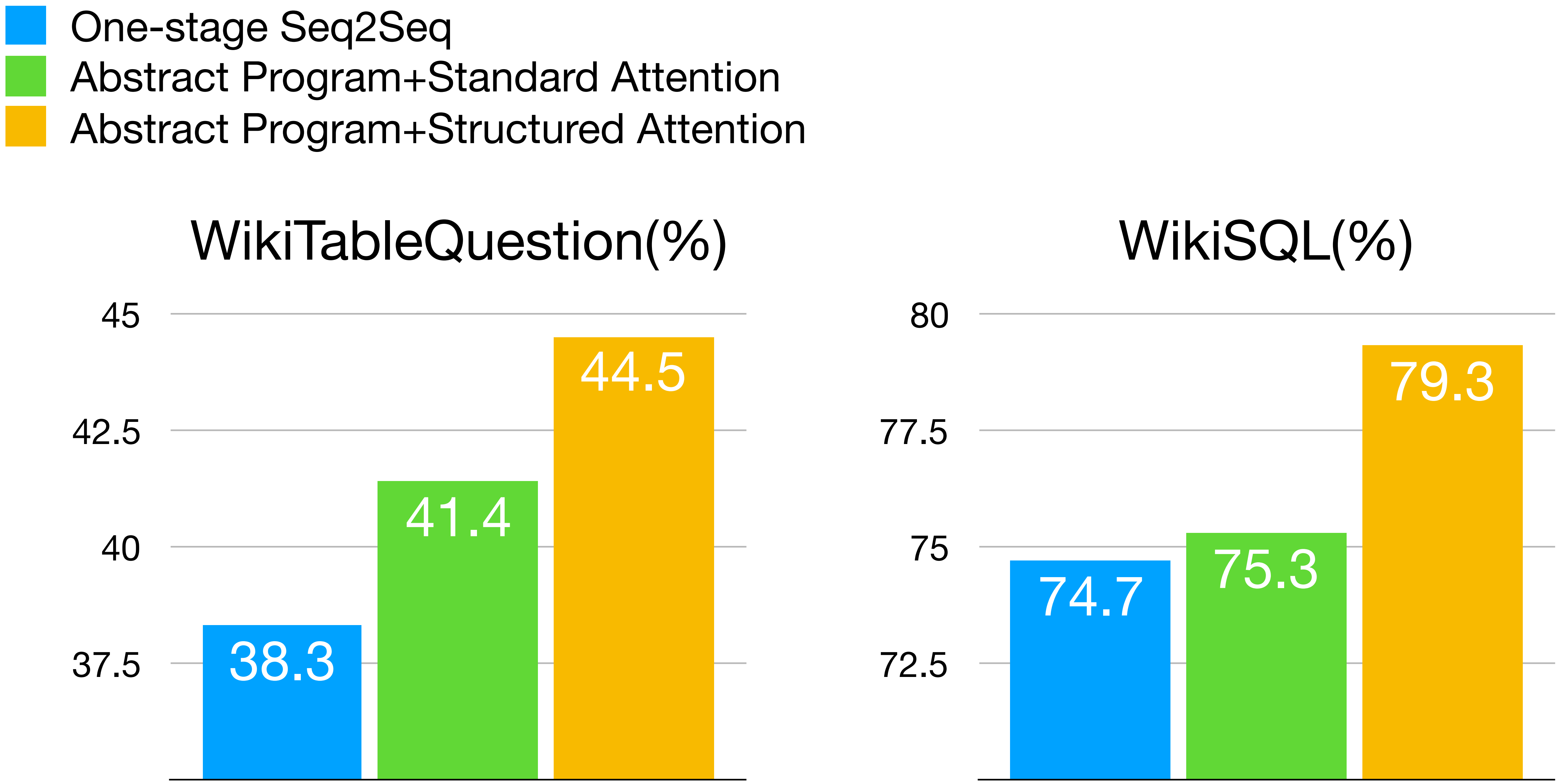
Programs

Question

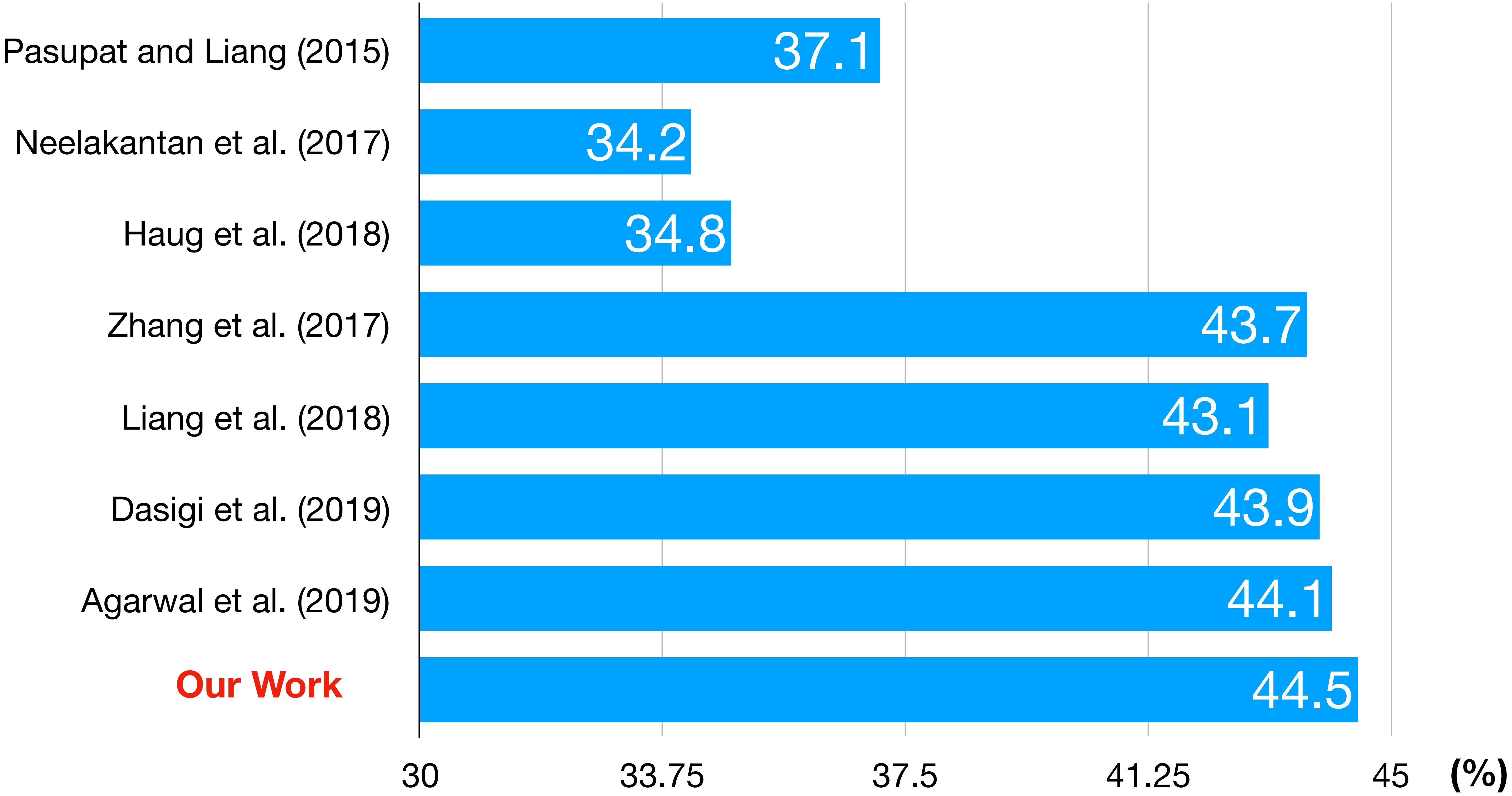
Seq2Seq

Programs

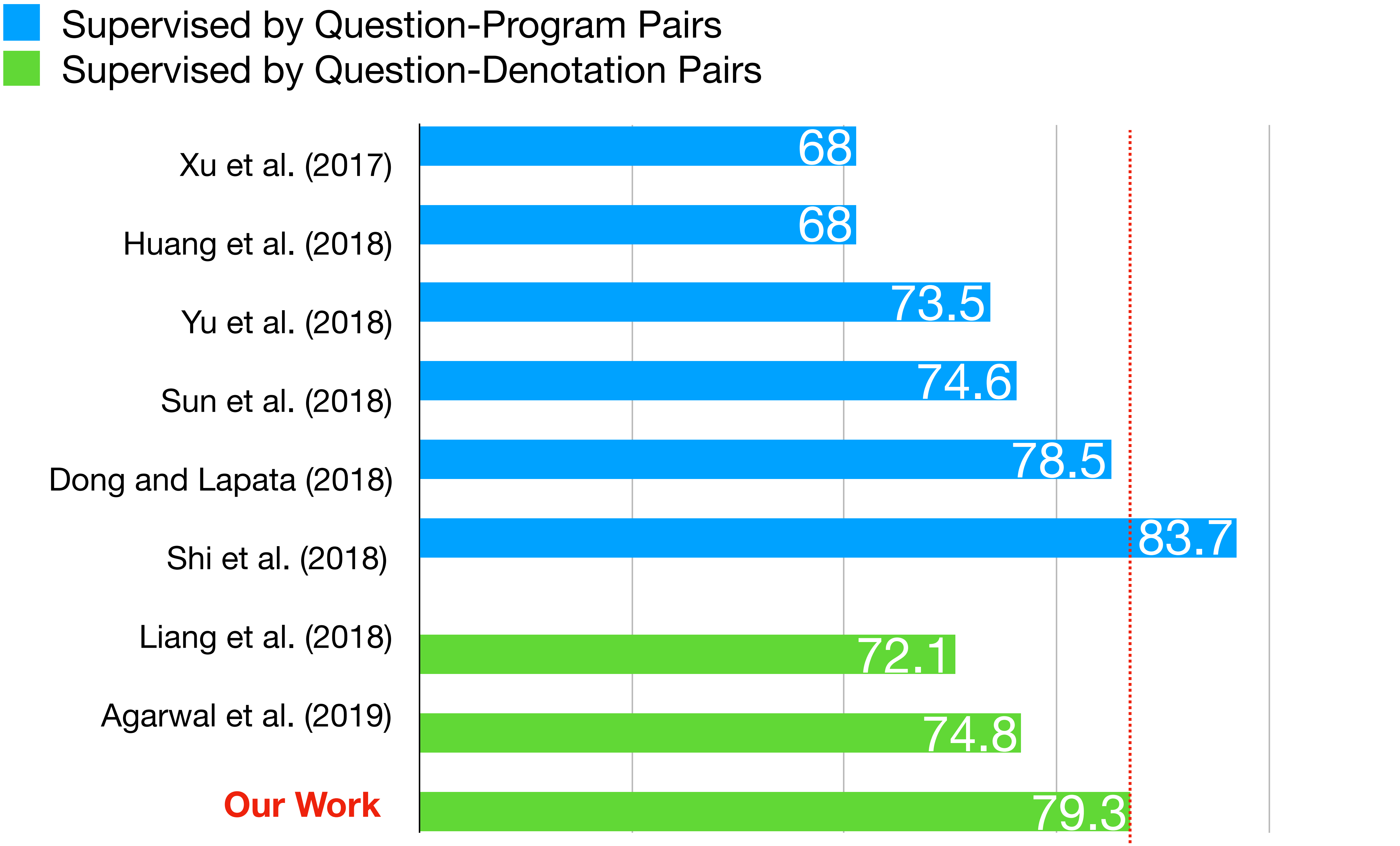
Compared with Baselines



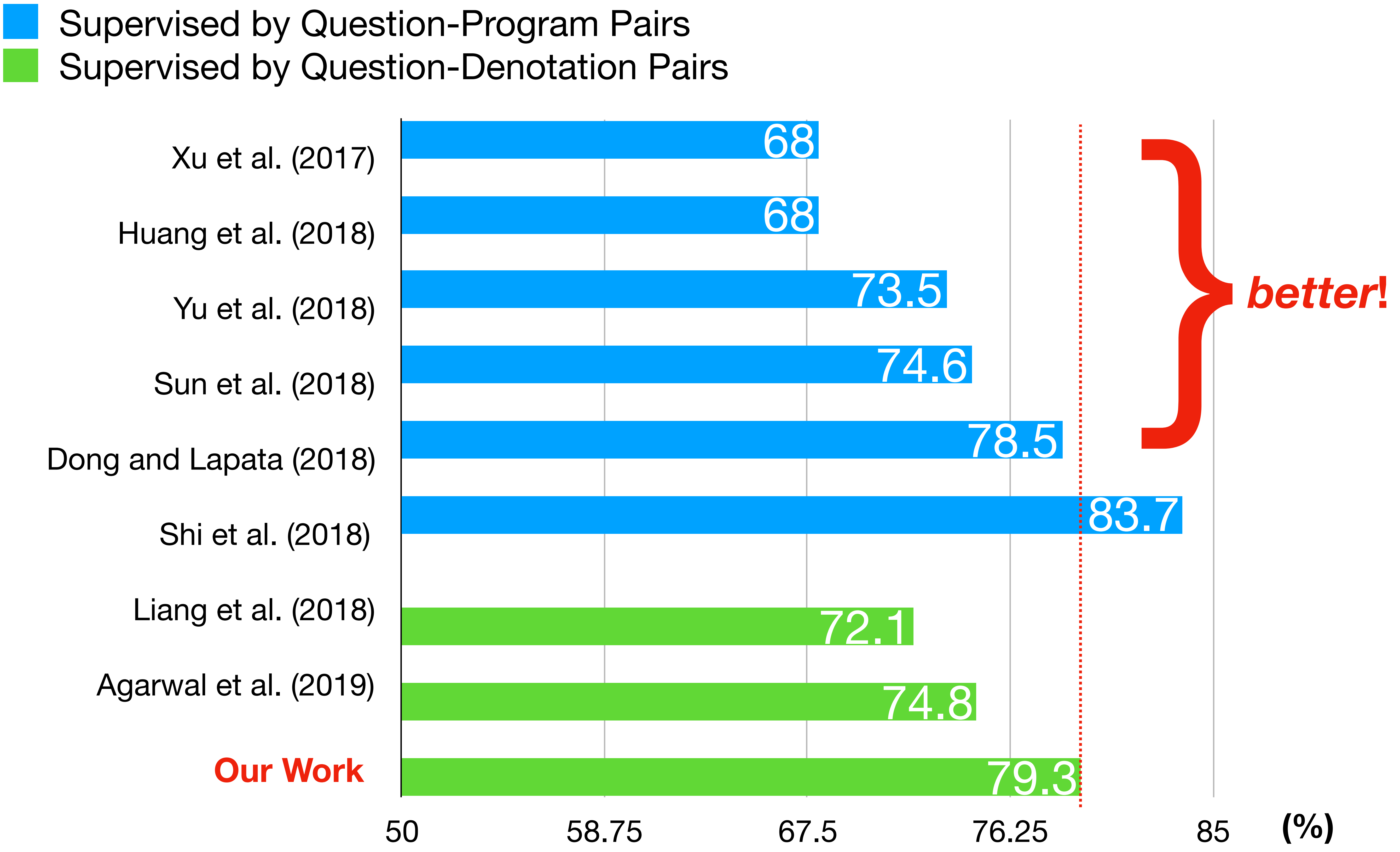
Results on WikiTableQ



Results on WikiSQL



Results on WikiSQL



Does it really handle spuriousness?

Expectation:

p_θ(**Spurious Programs**|...) < p_θ(**Correct Programs**|...)

Does it really handle spuriousness?

Expectation:

p_θ(**Spurious Programs**|...) **<** p_θ(**Correct Programs**|...)

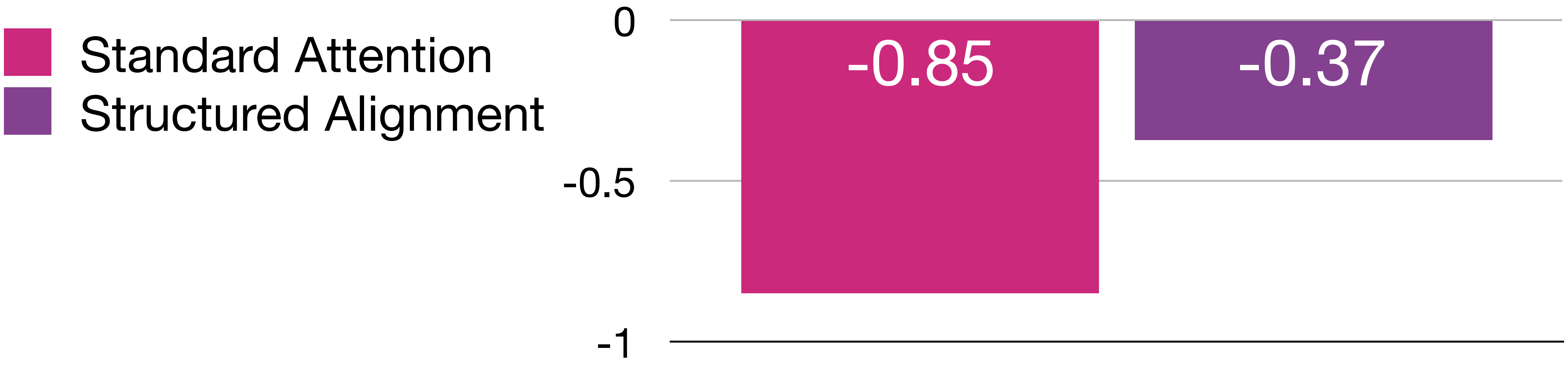
p_θ(**Correct Program**|**Consistent Programs**)

Does it really handle spuriousness?

Expectation:

p_θ(**Spurious Programs**|...) < p_θ(**Correct Programs**|...)

log p_θ(**Correct Program**|**Consistent Programs**)



Generalization

General Form of Two-Stage Framework

Abstract Programs + Probabilistic Alignment

domain specific language

logical forms

SQL/Python

⋮

+

one-to-one

monotonicity

bijectivity

sparsity

⋮

Summary

- We propose a semantic parser that features abstract programs and latent structured alignments.
- State-of-the-art performance on two benchmarks.
- Inductive bias introduced really help handle spuriousness.
- Code available: https://github.com/berlino/weaksp_em19

erc European Research Council

ERC Starting Grant BroadSem