neural semantic parsing

CS585, Fall 2019

Deep Learning for Natural Language Processing

Mohit Iyyer

College of Information and Computer Sciences

University of Massachusetts Amherst

many slides from Matt Gardner

questions from last time

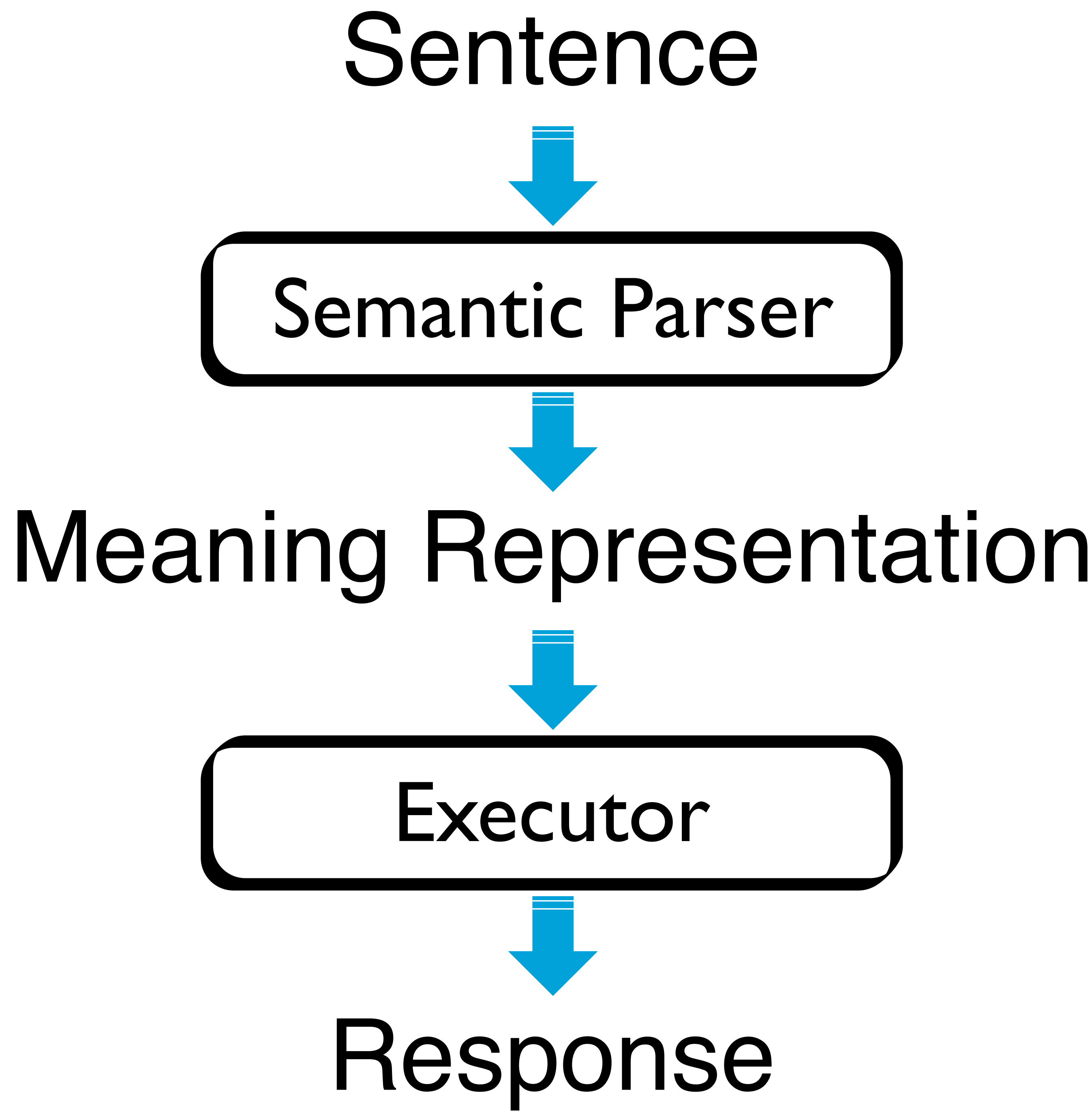
- Milestone 2 due thus Thurs Nov 21
- HW3 to be released tonight, will be due Dec 5
    - there is only one small coding part, which we will do in class on Thursday
    - the rest is just running cells and comparing numbers
- Project presentations Dec 10,1-4pm, CS atrium

midterm!!!!!

- Mean was ~55
  - roughly the same as previous iterations of the class

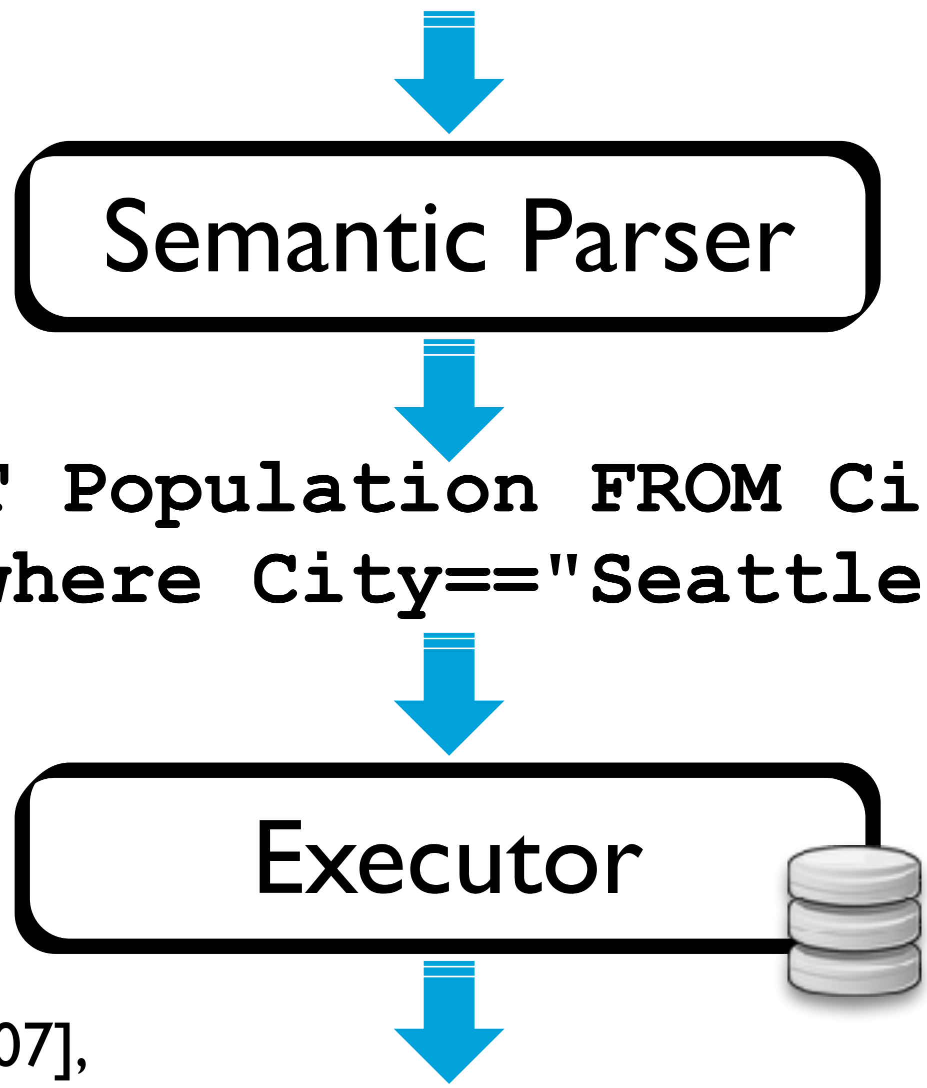
histogram of last year's 585 grades:

Semantic Parsing



Semantic Parsing: QA

How many people live in Seattle?



620,778

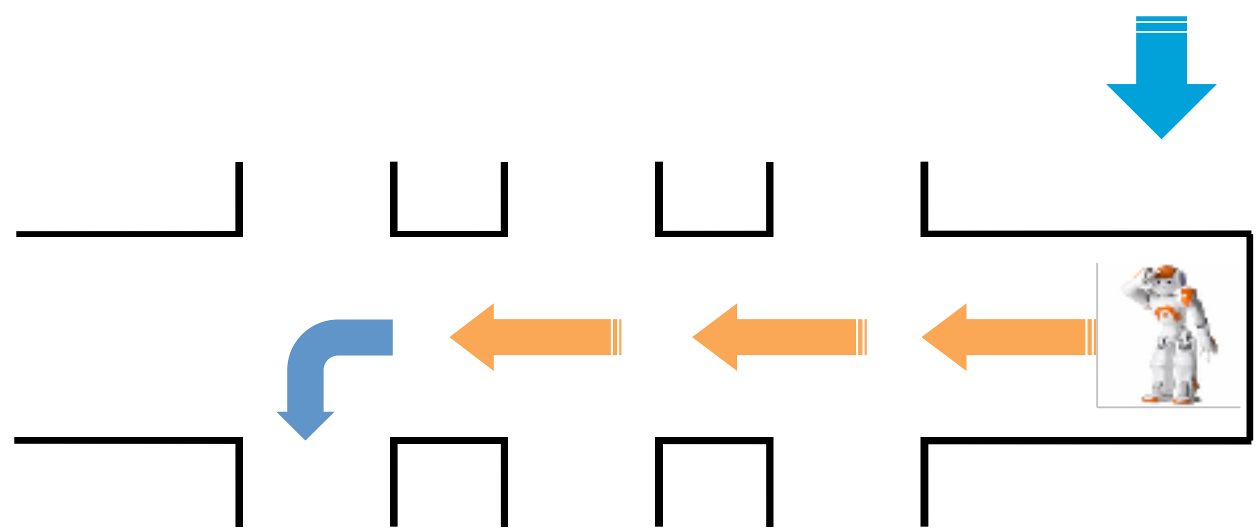
[Wong & Mooney 2007],
[Zettlemoyer & Collins 2005, 2007],
[Kwiatkowski et.al 2010, 2011],
[Liang et.al. 2011],[Berant et.al. 2013,2014],[Reddy et.al, 2014,2016],
[Dong and Lapata, 2016] .....

Semantic Parsing: Instructions

Go to the third junction and take a left

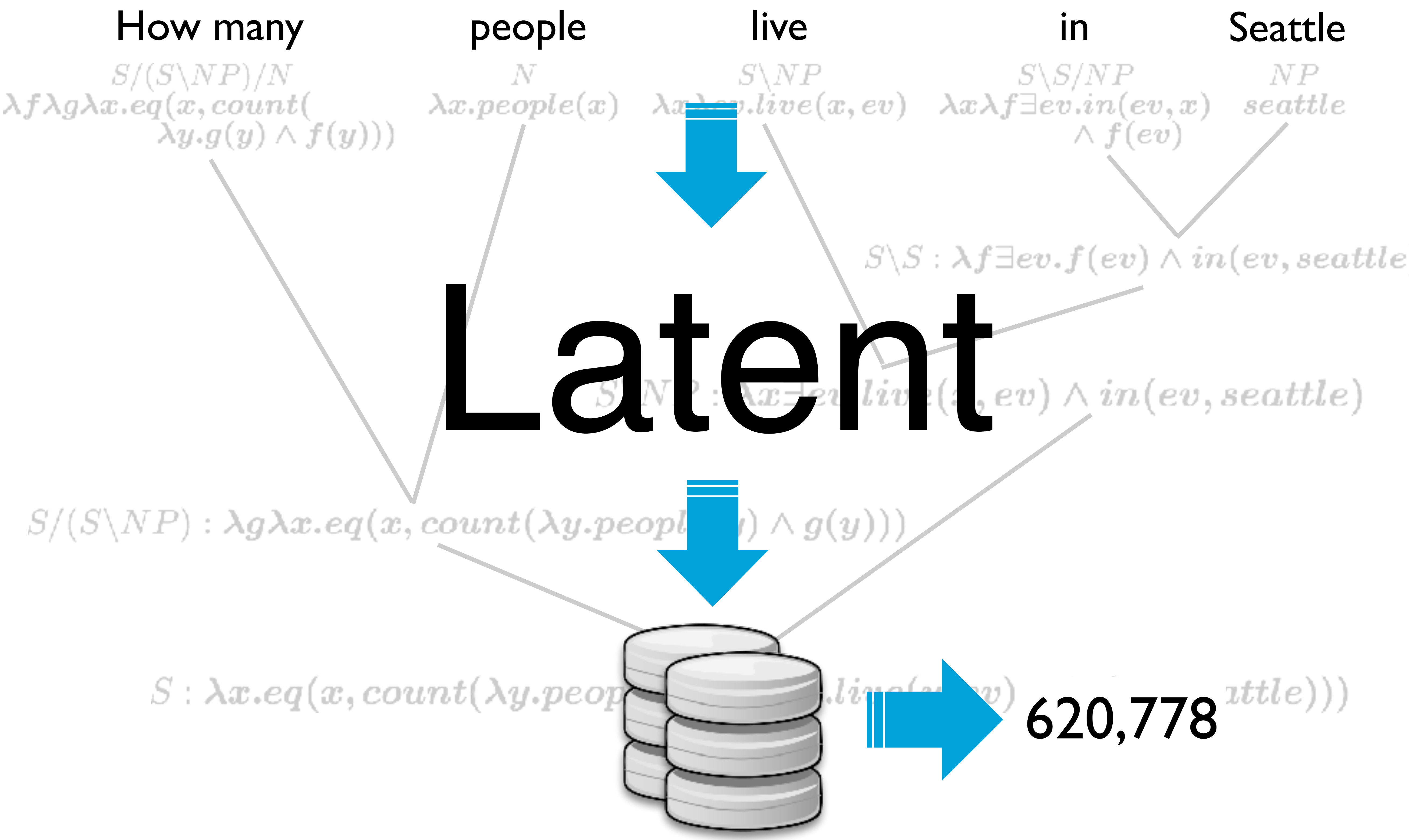
Semantic Parser

```
(do-seq(do-n-times 3
  (move-to forward-loc
    (do-until
      (junction current-loc
        (move-to forward-loc))))
  (turn-right))
```



[Chen & Mooney 2011]
[Matuszek et al 2012]
[Artzi & Zettlemoyer 2013]
[Mei et.al. 2015][Andreas et al, 2015]
[Fried at al, 2018] ....

Semantic Parsing: Complex Structure



CCG Semantic Parsing

$$
\begin{array}{cccc}
\text{move} & \text{to} & \text{the} & \text{chair} \\
\hline
S & AP/NP & NP/N & N \\
\lambda a.move(a) & \lambda x.\lambda a.to(a,x) & \lambda f.\iota x.f(x) & \lambda x.chair(x)
\end{array}
$$

$$
\frac{NP/N : \lambda f.\iota x.f(x) \quad N : \lambda x.chair(x)}{NP : \iota x.chair(x)}>
$$

$$
\frac{AP/NP : \lambda x.\lambda a.to(a,x) \quad NP : \iota x.chair(x)}{AP : \lambda a.to(a, \iota x.chair(x))}>
$$

$$
\frac{AP : \lambda a.to(a, \iota x.chair(x))}{S\backslash S : \lambda f.\lambda a.f(a) \wedge to(a, \iota x.chair(x))}
$$

$$
\frac{S : \lambda a.move(a) \quad S\backslash S : \lambda f.\lambda a.f(a) \wedge to(a, \iota x.chair(x))}{S : \lambda a.move(a) \wedge to(a, \iota x.chair(x))}>
$$

[Zettlemoyer & Collins 2005, 2007]

CCG Semantic Parsing

move to the chair

"The classic approach"

-Mark Johnson (~2016)

S

$\lambda a.move(a) \wedge to(a, \iota x.chair(x))$

[Zettlemoyer & Collins 2005, 2007]

CCG Semantic Parsing

move to the chair

- Complex discrete learning algorithms
- But, grammars hopefully generalize to unseen data well!
- Difficult to engineer: few people can do it and it takes a lot of time

$\lambda a.move(a) \wedge to(a, \iota x.chair(x))$

[Zettlemoyer & Collins 2005, 2007]

Enter seq2seq... (Dong & Lapata, 2016)

- Treat meaning as a string...
- Apply NMT
- Close to SOTA performance!!!
- Much easier to build (with toolkits)

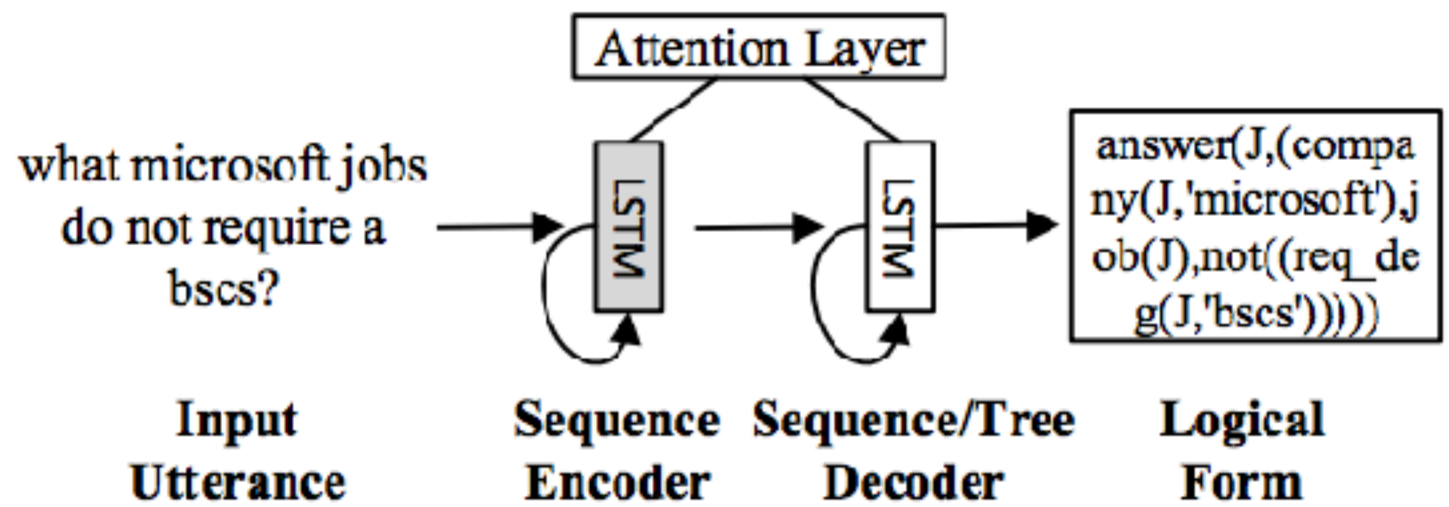


Figure 1: Input utterances and their logical forms are encoded and decoded with neural networks. An attention layer is used to learn soft alignments.

issues with vanilla seq2seq?

Example from WikiTableQuestions

Athlete	Nation	Olympics	Medals
Gillis Grafström	Sweden (SWE)	1920–1932	4
Evgeni Plushenko	Russia (RUS)	2002–2014	4
Karl Schäfer	Austria (AUT)	1928–1936	2
Katarina Witt	East Germany (GDR)	1984–1988	2
Tenley Albright	United States (USA)	1952-1956	2
Kim Yu-na	South Korea (KOR)	2010–2014	2
Patrick Chan	Canada (CAN)	2014	2

Question:

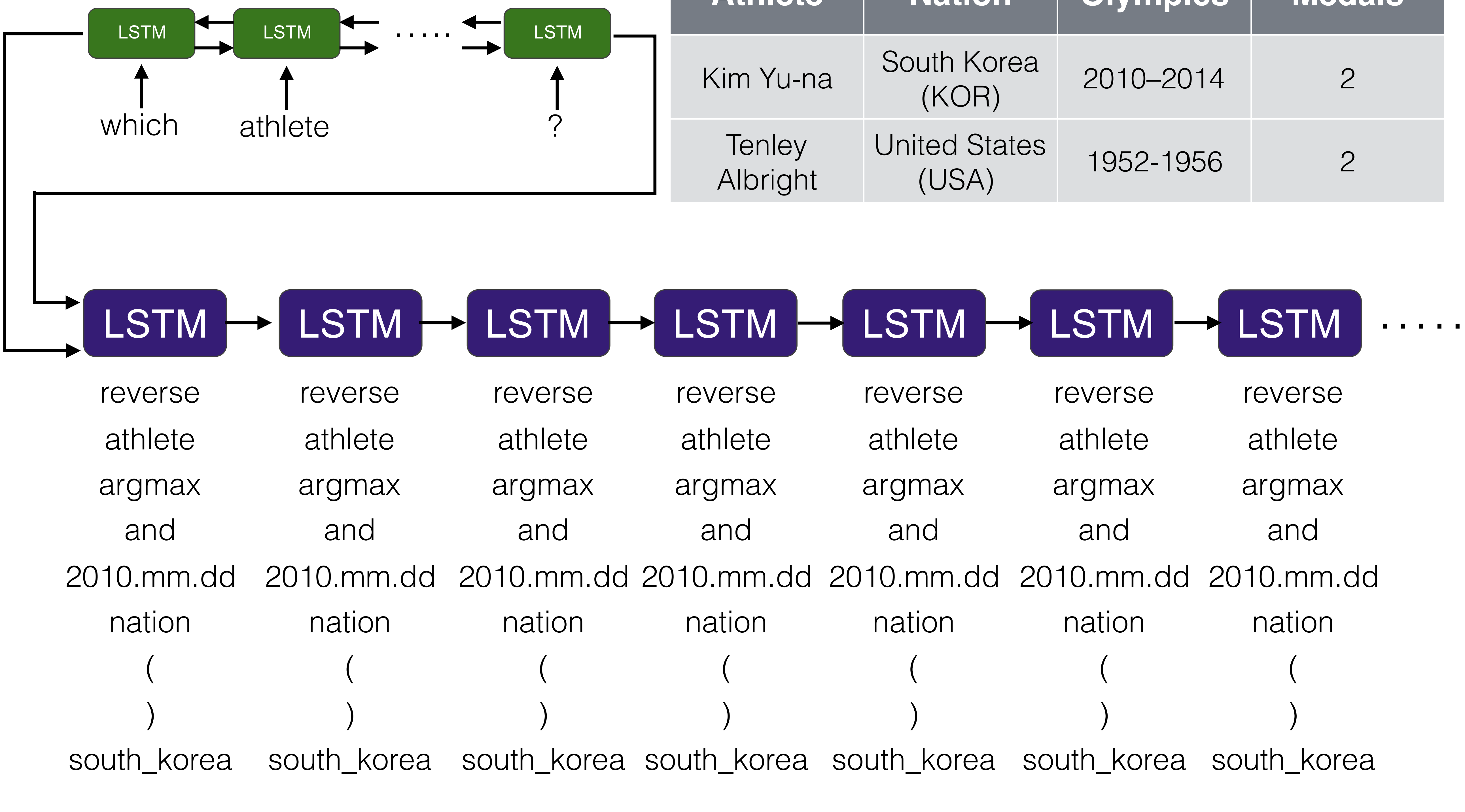
Which athlete was from South Korea after 2010?

```
((reverse athlete)
 (and
  (nation south_korea)
  (year ((reverse date)
     (>= 2010-mm-dd)))
```

[Pasupat and Liang, 2015]

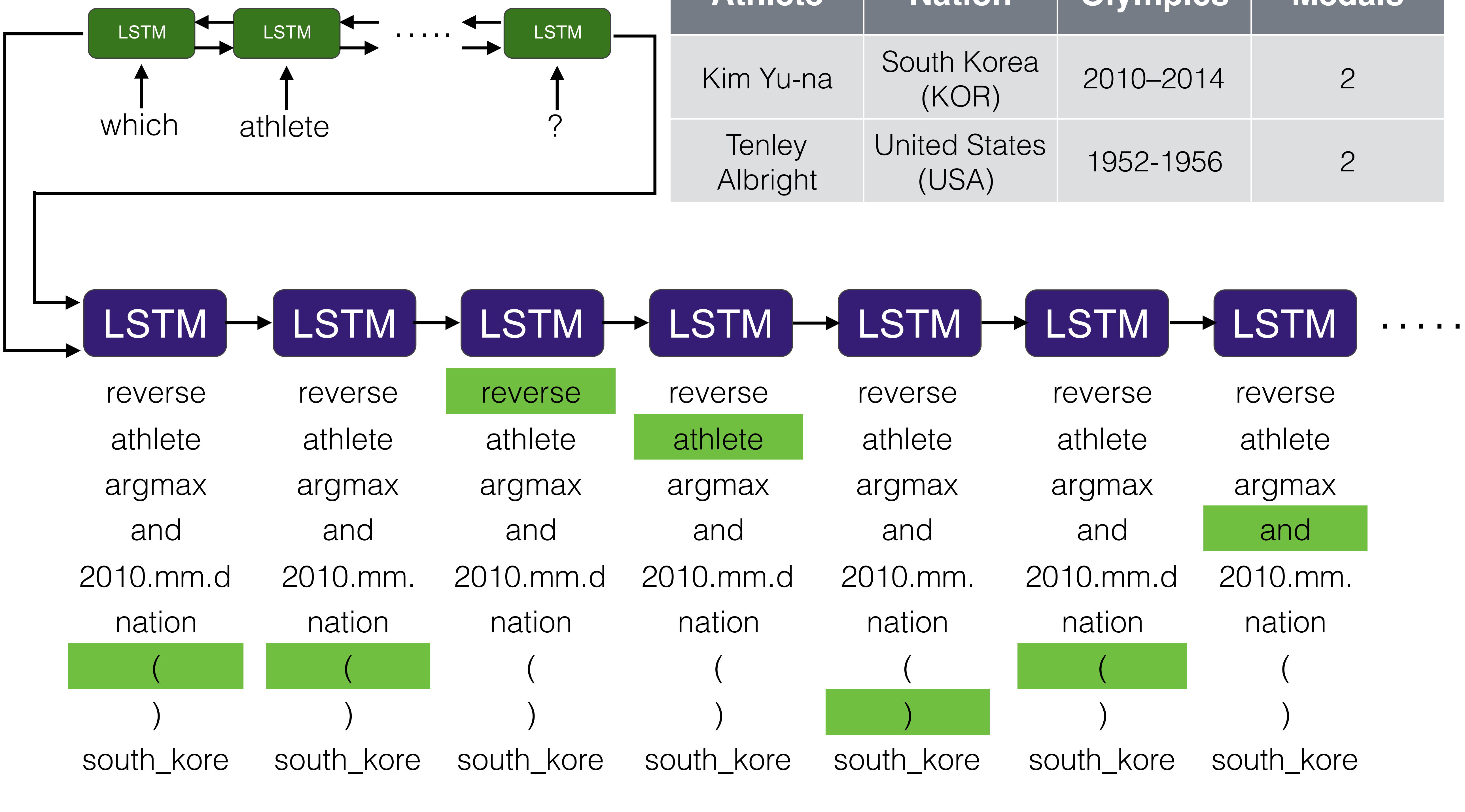
Seq2Seq Output Space

Athlete	Nation	Olympics	Medals
Kim Yu-na	South Korea (KOR)	2010–2014	2
Tenley Albright	United States (USA)	1952-1956	2

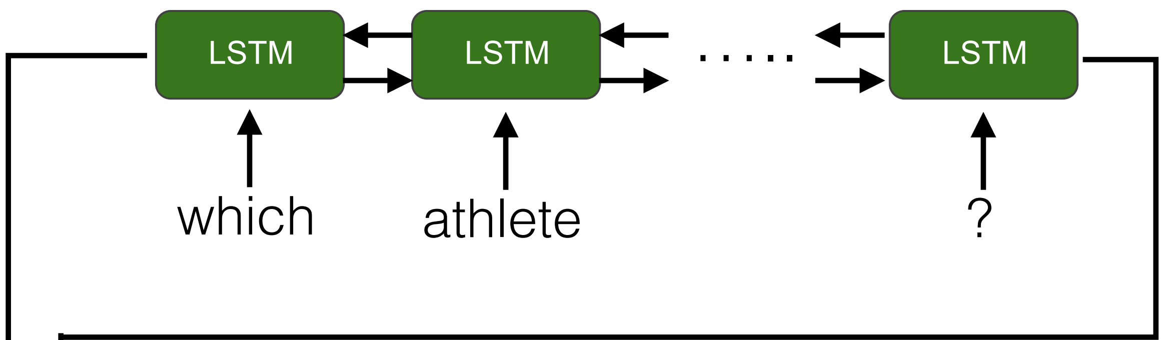


Seq2Seq Output Space

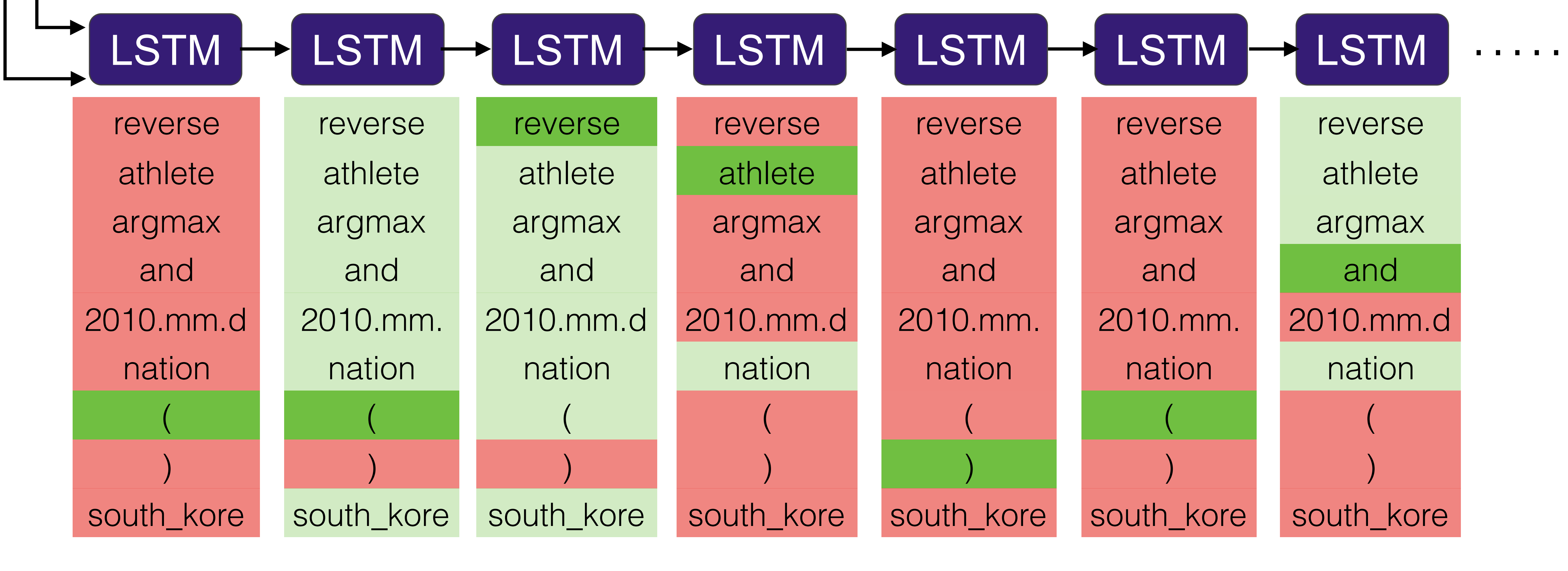
Athlete	Nation	Olympics	Medals
Kim Yu-na	South Korea (KOR)	2010–2014	2
Tenley Albright	United States (USA)	1952-1956	2



Seq2Seq Output Space



Athlete	Nation	Olympics	Medals
Kim Yu-na	South Korea (KOR)	2010–2014	2
Tenley Albright	United States (USA)	1952-1956	2



Constrained Decoding

- Constrain the output space to selections that matter
- **Inference:** Avoid invalid parses
- **Training:** Do not waste modeling power in distinguishing invalid parses from valid ones!

Token-based Decoding:

The output space is tokens, but they are constrained to be relevant at each time step.

Grammar-based Decoding:

The output space is production rules, and a grammar defines the constraints.

Constrained Decoding

- Constrain the output space to selections that matter
- **Inference:** Avoid invalid parses
- **Training:** Do not waste modeling power in distinguishing invalid parses from valid ones!

Token-based Decoding

Dong and Lapata. 2016. Language to Logical Form with Neural Attention. In ACL.

Dong and Lapata. 2018. Coarse-to-Fine Decoding for Neural Semantic Parsing. In ACL.

Goldman, Latcinnik, Naveh, Globerson and Berant. 2018. Weakly-supervised Semantic Parsing with Abstract Examples. In ACL.

Grammar-based Decoding:

Xiao, Dymetman, and Gardent. 2016. Sequence-based Structured Prediction for Semantic Parsing. In ACL.

Yin and Neubig. 2017. A Syntactic Neural Model for General Purpose Code Generation. In ACL.

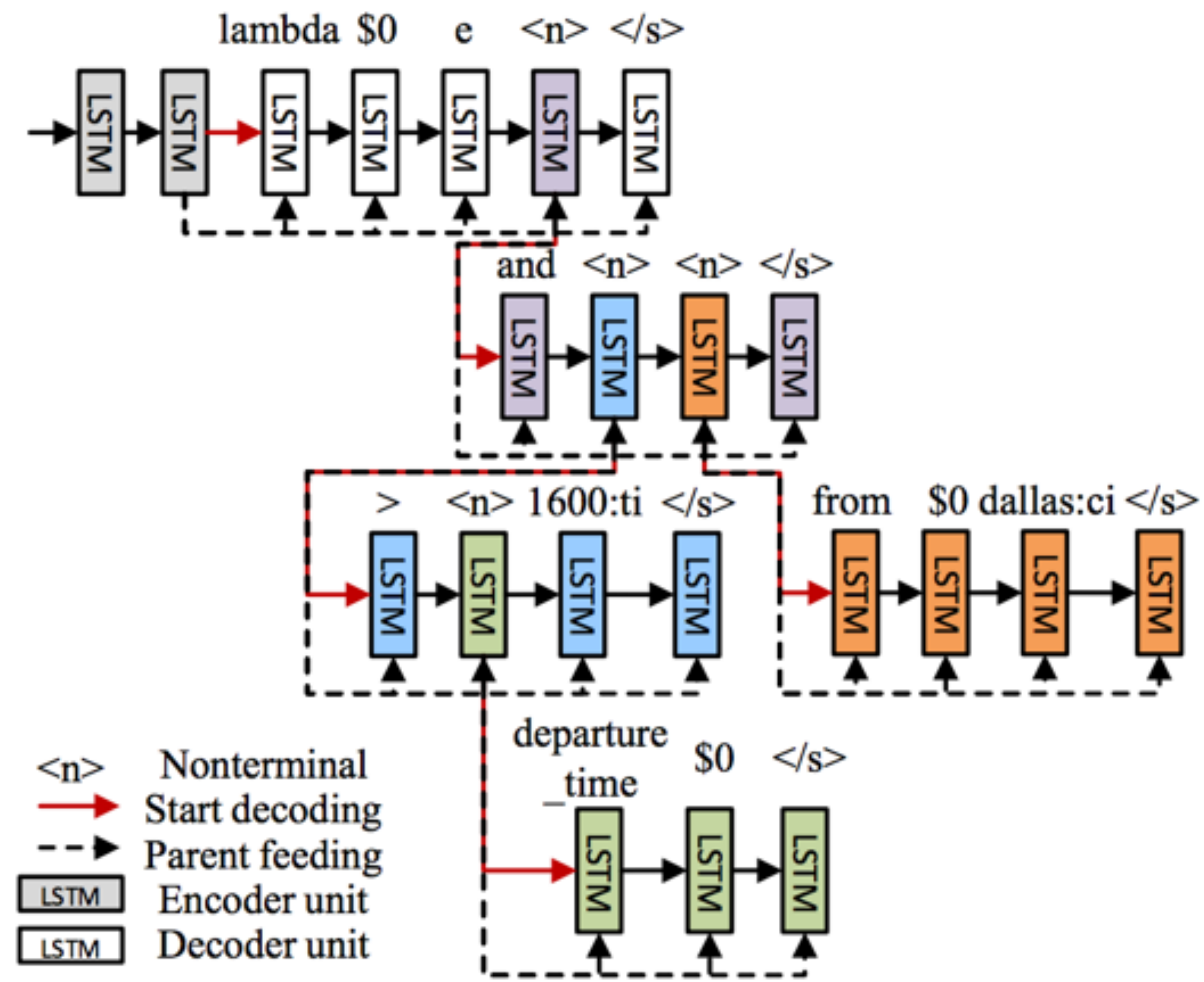
Krishnamurthy, Dasigi, and Gardner. 2017. Neural Semantic Parsing with Type Constraints for Semi-Structured Tables. In EMNLP.

Token-based Constrained Decoding

Constraining output structure: Seq2Tree

Flights from Dallas leaving after 4 in the afternoon

(lambda $0 e
  (and
    (> (departure_time $0) 1600:ti)
    (from $0 dallas:ci)))

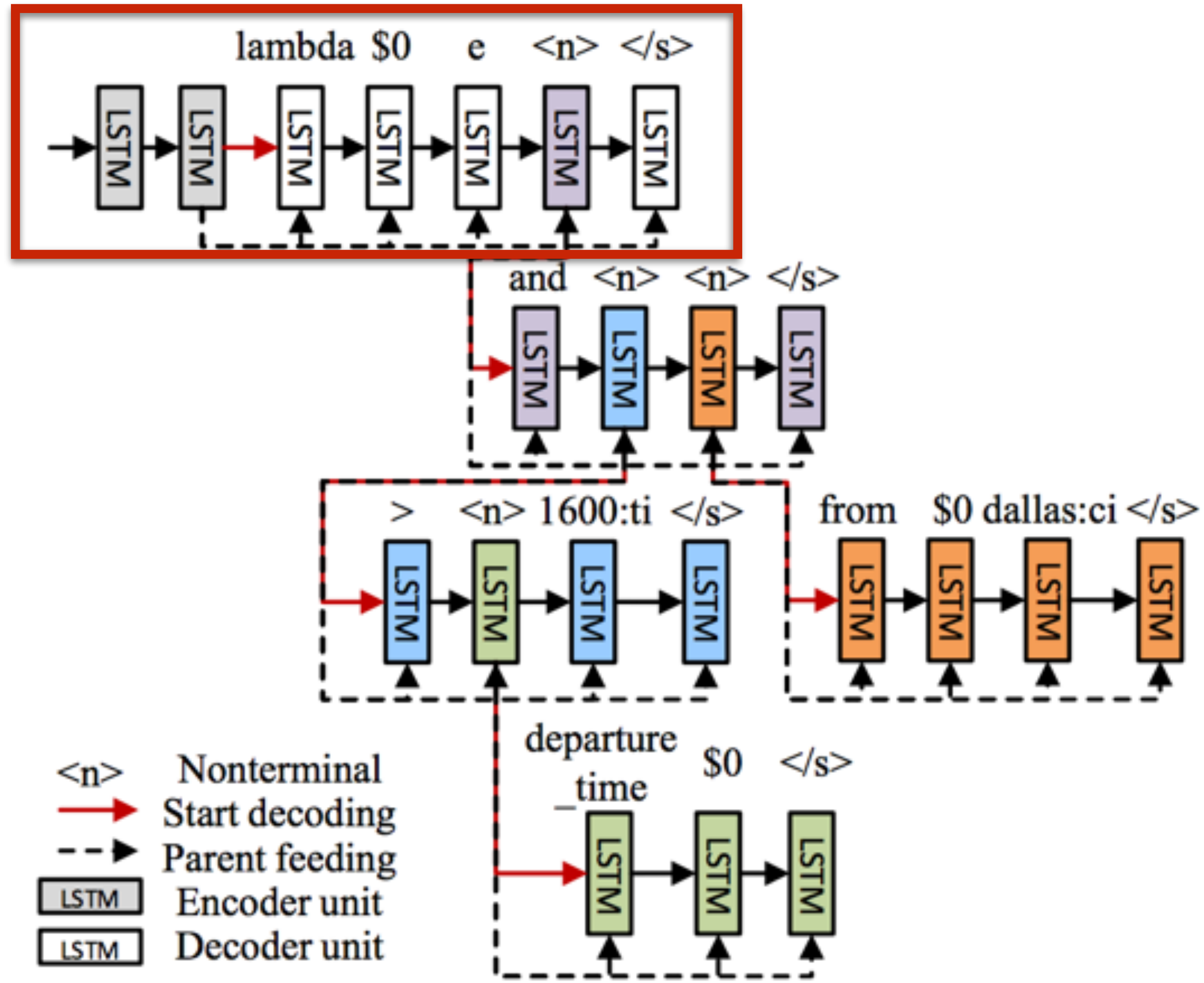


[Dong and Lapata, 2016]

Constraining output structure: Seq2Tree

Flights from Dallas leaving after 4 in the afternoon

(lambda $0 e **<n>**)

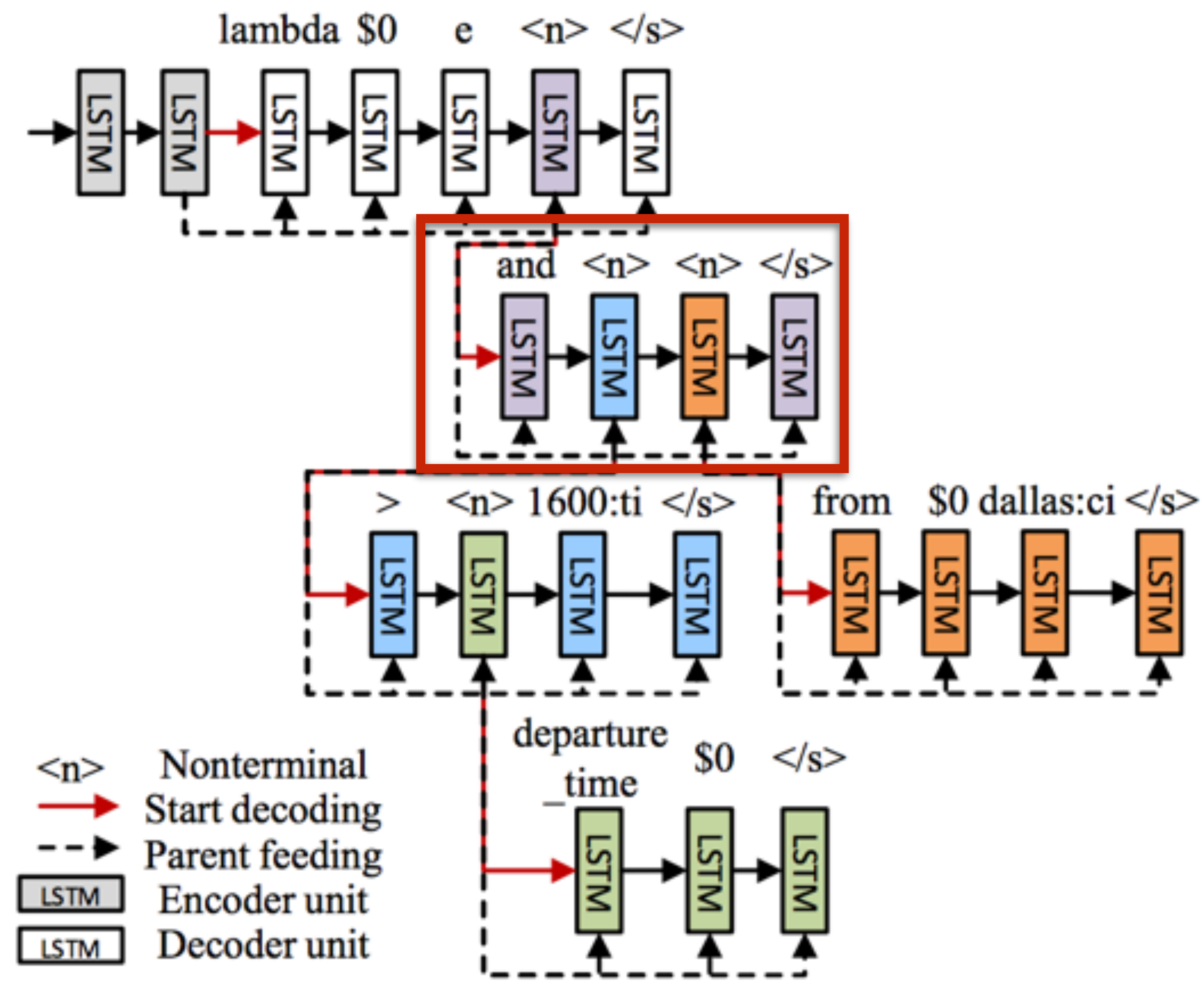


[Dong and Lapata, 2016]

Constraining output structure: Seq2Tree

Flights from Dallas leaving after 4 in the afternoon

(lambda $0 e
  (and **<n>** <n>))

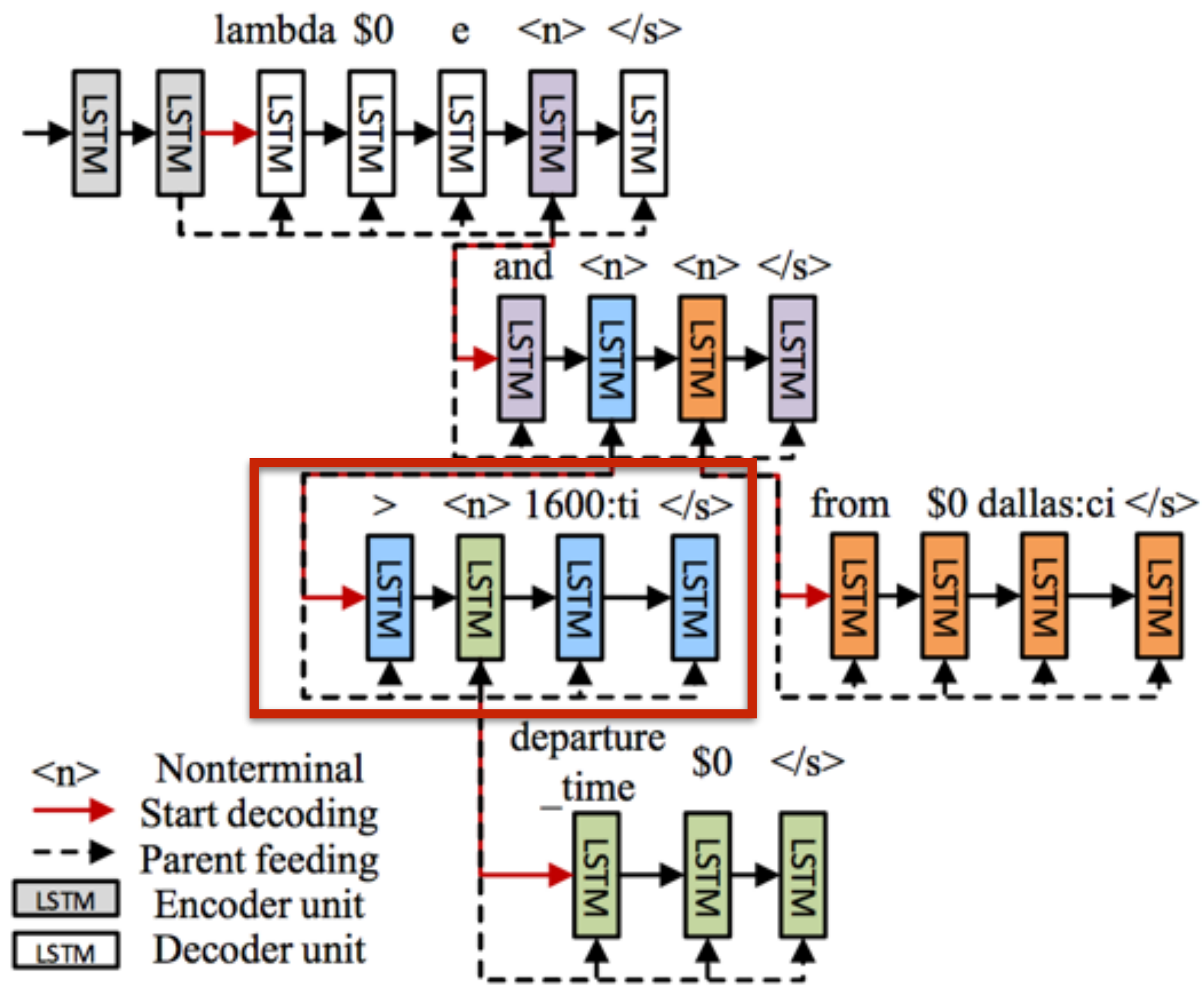


[Dong and Lapata, 2016]

Constraining output structure: Seq2Tree

Flights from Dallas leaving after 4 in the afternoon

(lambda $0 e
  (and
    (> **<n>** 1600:ti)
     <n>))

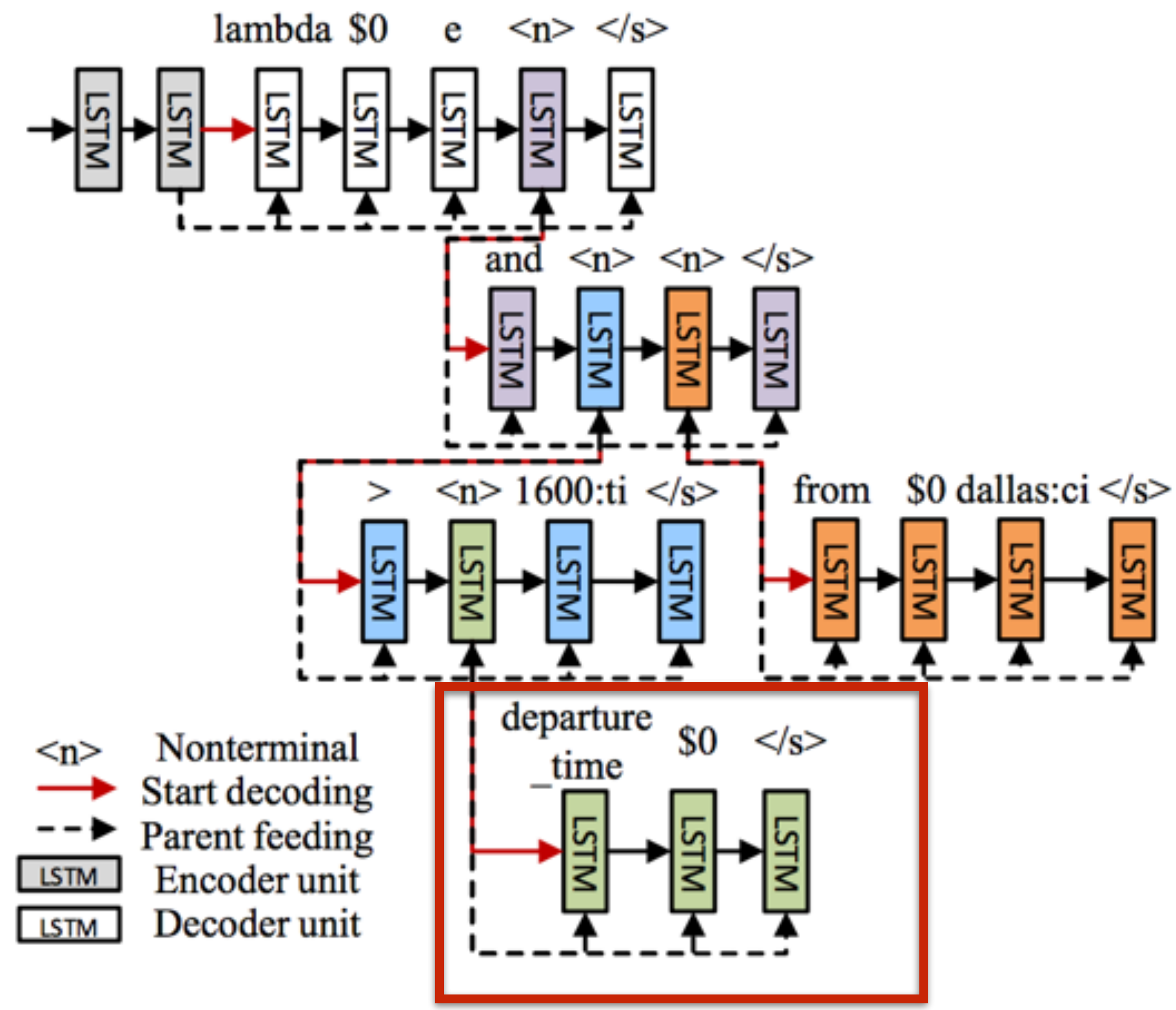


[Dong and Lapata, 2016]

Constraining output structure: Seq2Tree

Flights from Dallas leaving after 4 in the afternoon

(lambda $0 e
(and
(> (departure_time $0) 1600:ti)
<n>))

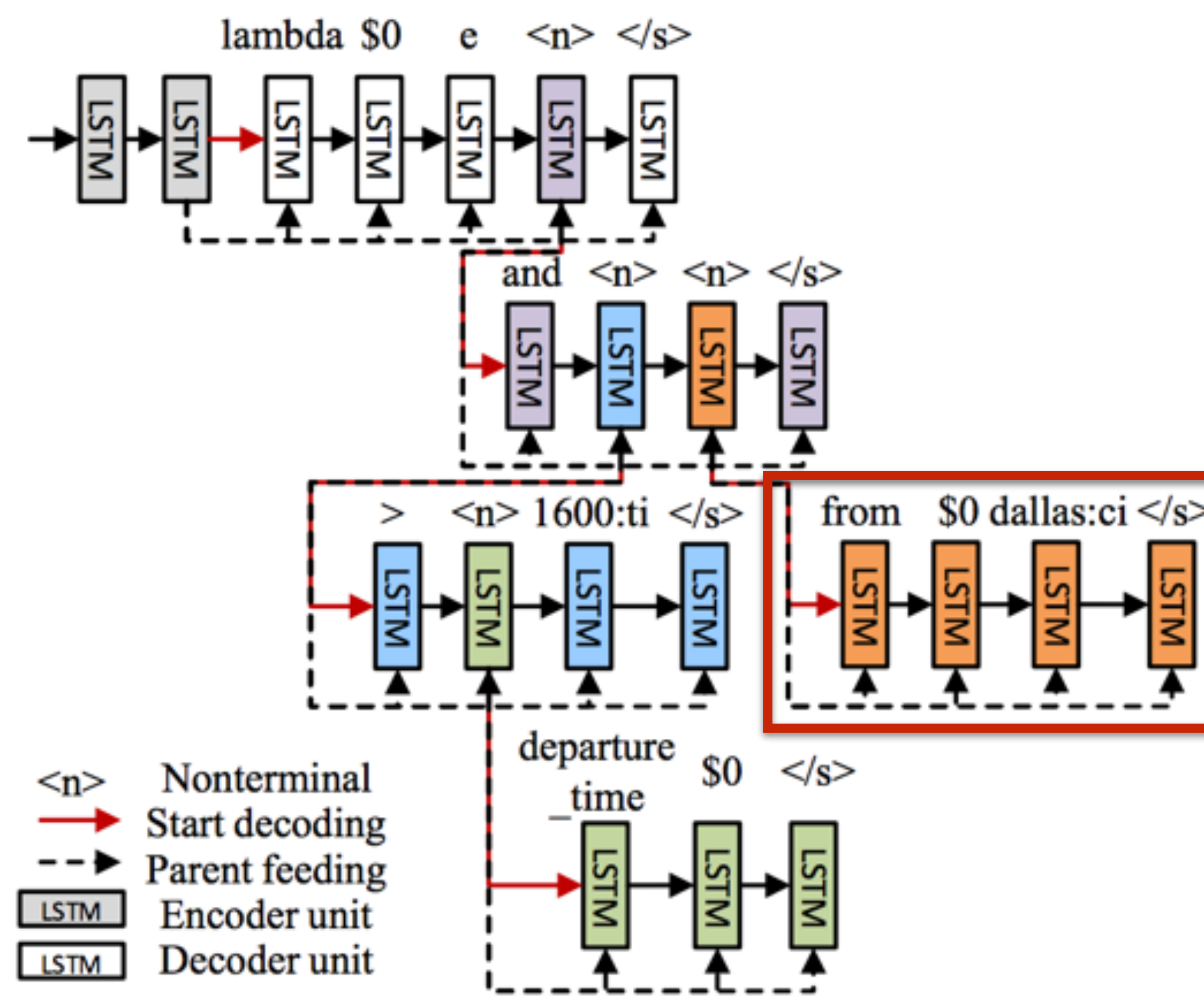


[Dong and Lapata, 2016]

Constraining output structure: Seq2Tree

Flights from Dallas leaving after 4 in the afternoon

(lambda $0 e
  (and
    (> (departure_time $0) 1600:ti)
    (from $0 dallas:ci)))



[Dong and Lapata, 2016]

How do I train a semantic parser?

Got Supervision?

x_i: flights from Dallas leaving after 4 in the afternoon

y_i: (lambda $0 e
(and
(>(departure_time $0) 1600:ti)
(from $0 dallas:ci)))

$$D = \{x_i, y_i\}_{i=1}^{N}$$

Task: Given x_{N+k} find y_{N+k}

Fully supervised

Got Supervision?

x_i: flights from Dallas leaving after 4 in the afternoon

y_i: (lambda $0 e
(and
(>(departure_time $0) 1600:ti)
(from $0 dallas:ci)))

$$D = \{x_i, y_i\}_{i=1}^{N}$$

Task: Given x_{N+k} find y_{N+k}

Fully supervised

x_i: Which athlete was from South Korea after 2010?

~~y_i: ((reverse athlete)~~
~~(and~~
~~(nation south_korea)~~
~~(year ((reverse date) (>= 2010)))~~

z_i: Kim Yu-Na

w_i:

Athlete	Nation	Olympics	Medals
Kim Yu-na	South Korea	2010–2014	2
Tenley Albright	United States	1952-1956	2

$$D = \{x_i, w_i, z_i\}_{i=1}^{N}$$

Task: Given x_{N+k}, w_{N+k} find y_{N+k}
such that $[\![y_{N+k}]\!]^{w_{N+k}} = z_{N+k}$

Weakly supervised

Three common training methods

- Maximum Marginal Likelihood
- Structured Learning Methods
- Reinforcement Learning Methods

And some hybrid approaches..

Maximum Marginal Likelihood

- Given $D = \{x_i, w_i, z_i\}_{i=1}^{N}$
- We want to optimize $\max_{\theta} \prod_{x_i, z_i \in D} p(z_i | x_i; \theta)$
- But the semantic parser defines a distribution over logical forms.
- So we marginalize over logical forms that yield z_i

$$\max_{\theta} \prod_{x_i, w_i, z_i \in D} \sum_{y_i \in Y \,|\, [\![y_i]\!]^{w_i} = z_i} p(y_i | x_i; \theta)$$

- Y could be the set of all valid logical forms, if we are using constrained decoding during training
- Even then, the summation could be intractable!

Structured Learning Methods

- More commonly used with traditional semantic parsers
    - Eg. Margin based models and Latent variable structured perceptron (Zettlemoyer and Collins 2007)
- Typically involve heuristic search over the state space like MML methods
- Unlike MML, can use arbitrary cost function
- Training typically maximizes margins or minimizes expected risks

MML: Approximating $\mathbf{Y}$

- Perform heuristic search
- Search may be bounded, by length or otherwise
- $\mathbf{Y}$ is approximated as a subset of retrieved logical forms

Two options for search:

Online Search	Offline Search
Search for consistent logical forms during training, as per model scores	Search for consistent logical forms before training
Candidate set changes as training progresses	Candidate set is static
Less efficient	More efficient

Reinforcement Learning Methods

- Comparison with MML:
  - Like MML Y is approximated
  - Unlike MML, the approximation is done using sampling techniques
- Comparison with structured learning methods
  - Like structured learning methods, the reward function can be arbitrary
  - Unlike structured learning methods, reward is directly maximized
- Training typically uses policy gradient methods

  Example from Liang et al., 2017, using REINFORCE

$$\max_{\theta} \sum_{x} \mathbb{E}_{P_{\theta}(a_{0:T}|x)}[R(x, a_{0:T})]$$

What you need on top of seq2seq

1. Convert programs to action sequences

2. What actions are valid at every timestep?

3. Convert action sequences back to programs

4. (sometimes) A way to execute programs

5. If you don’t have labeled logical forms: a different way to train

let’s look at a method for **sequential semantic parsing** that combines structured learning and RL!

Iyyer et al., ACL 2017

conversational contexts are hard!

How much protein is in an egg?
And how many carbohydrates?

Are eggs on my shopping list?
What about butter?

Do I need an umbrella today?
Where can I buy one?

What's 42 plus 8 minus 13?
Is the answer divisible by 4?

conversational contexts are hard!

How much protein is in an egg?
And how many carbohydrates?

Are eggs on my shopping list?
What about butter?

Do I need an umbrella today?
Where can I buy one?

What's 42 plus 8 minus 13?
Is the answer divisible by 4?

the follow-up question can only be answered by resolving either an *explicit* or *implied* reference to the previous question

FINA Women’s Water Polo World Cup

Rank	Nation	Gold	Silver
1	Netherlands	8	3
2	Australia	3	3
3	USA	2	5
4	Hungary	1	1
5	Canada	0	0

FINA Women's Water Polo World Cup

Rank	Nation	Gold	Silver
1	Netherlands	8	3
2	Australia	3	3
3	USA	2	5
4	Hungary	1	1
5	Canada	0	0

1. Which nations competed in the FINA women's water polo cup?

FINA Women's Water Polo World Cup

Rank	Nation	Gold	Silver
1	Netherlands	8	3
2	Australia	3	3
3	USA	2	5
4	Hungary	1	1
5	Canada	0	0

1. Which nations competed in the FINA women's water polo cup?

SELECT Nation

semantic parse:
a logical form executed on table to yield answer

FINA Women's Water Polo World Cup

Rank	Nation	Gold	Silver
1	Netherlands	8	3
2	Australia	3	3
3	USA	2	5
4	Hungary	1	1
5	Canada	0	0

1. Which nations competed in the FINA women's water polo cup?

SELECT Nation

2. Of these nations, which ones took home at least one gold medal?

SUBSEQUENT WHERE Gold != 0

FINA Women's Water Polo World Cup

Rank	Nation	Gold	Silver
1	Netherlands	8	3
2	Australia	3	3
3	USA	2	5
4	Hungary	1	1
5	Canada	0	0

1. Which nations competed in the FINA women's water polo cup?

SELECT Nation

2. Of these nations, which ones took home at least one gold medal?

SUBSEQUENT WHERE Gold != 0

SUBSEQUENT: handles references between questions

FINA Women's Water Polo World Cup

Rank	Nation	Gold	Silver
1	Netherlands	8	3
2	Australia	3	3
3	USA	2	5
4	Hungary	1	1
5	Canada	0	0

1. Which nations competed in the FINA women's water polo cup?

SELECT Nation

2. Of these nations, which ones took home at least one gold medal?

SUBSEQUENT WHERE Gold != 0

3. Of those, which ranked in the top 2 positions?

SUBSEQUENT WHERE Rank <= 2

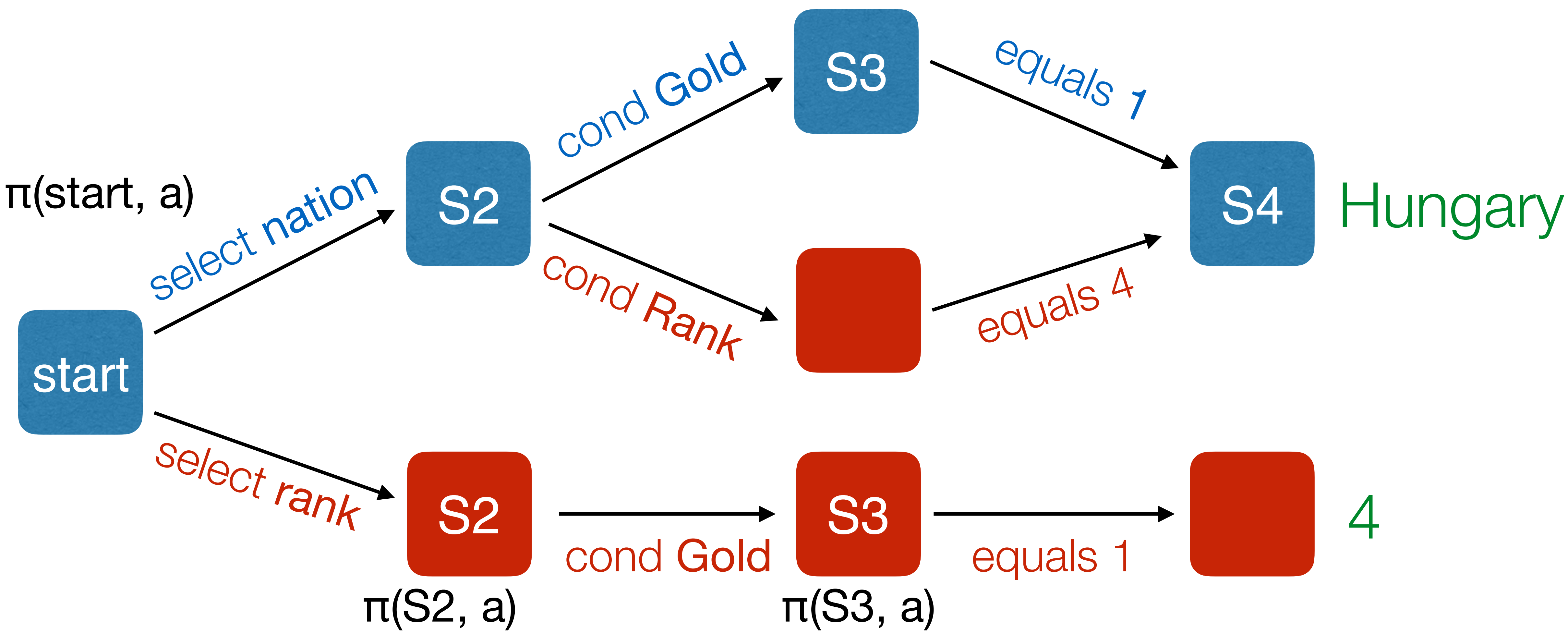
dynamic semantic parsing

- We collect **SQA**, a dataset of ~6000 question/answer sequences
- Since we only know the answer to a question and *not* its ground-truth logical form, this problem is only weakly supervised.
- To solve it, we use **reward-guided structured-output learning**

dynamic semantic parsing

Q: which nations won exactly one gold medal? A: Hungary

1. select-column Nation
2. cond-column Gold
3. op-equal 1

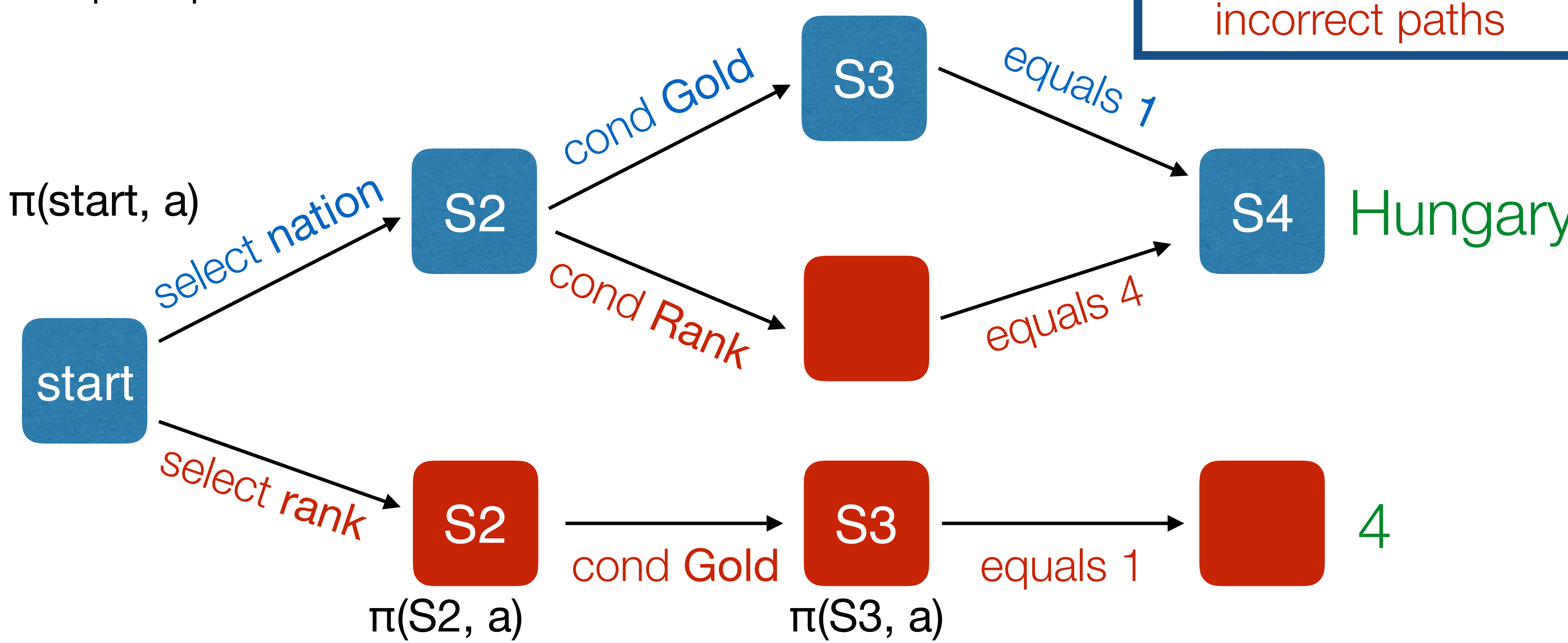


dynamic semantic parsing

Q: which nations won exactly one gold medal? A: **Hungary**

1. select-column **Nation**
2. cond-column **Gold**
3. op-equal **1**

maximize score of
(approx) correct path
minimize score of
incorrect paths



dynamic semantic parsing

- neural network modules output scalar values which we use in the **value function**
  π(current parse, next operation)
- end-to-end training algorithm: approximate a reference parse and train the value function to favor that parse
- discourse-level information incorporated with SUBSEQUENT statements, which have their own action semantics

ex: module implementation