

Neural and neurosymbolic parsing

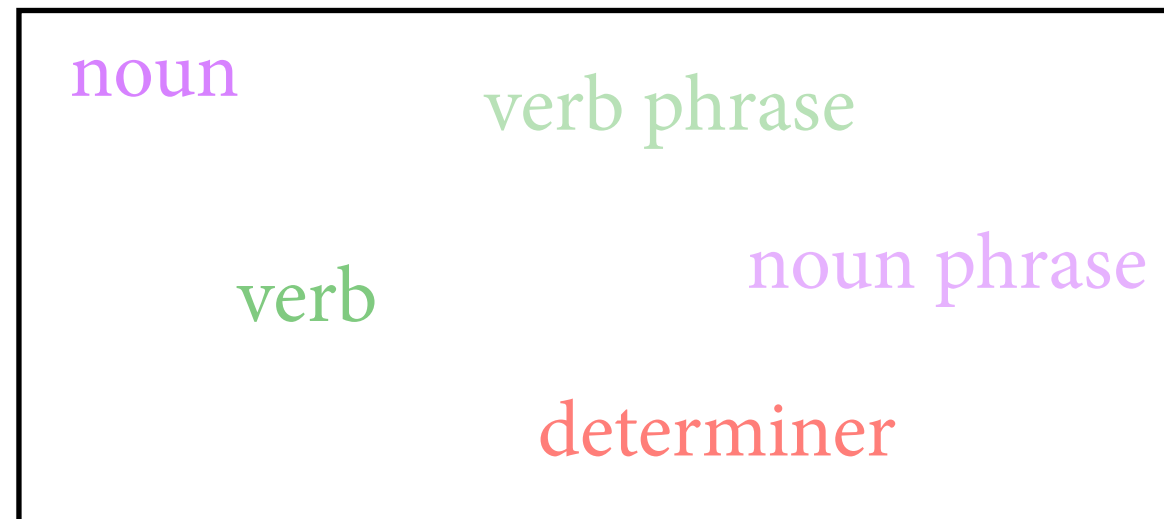
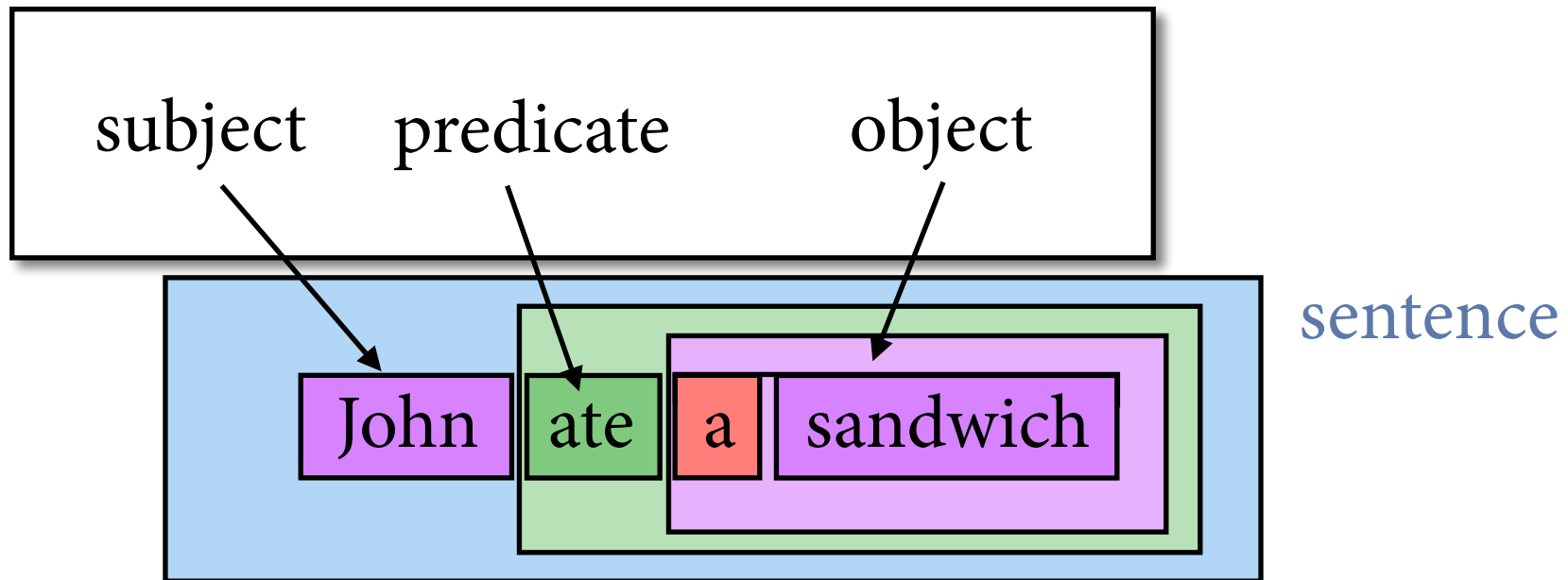
Computational Linguistics

Alexander Koller

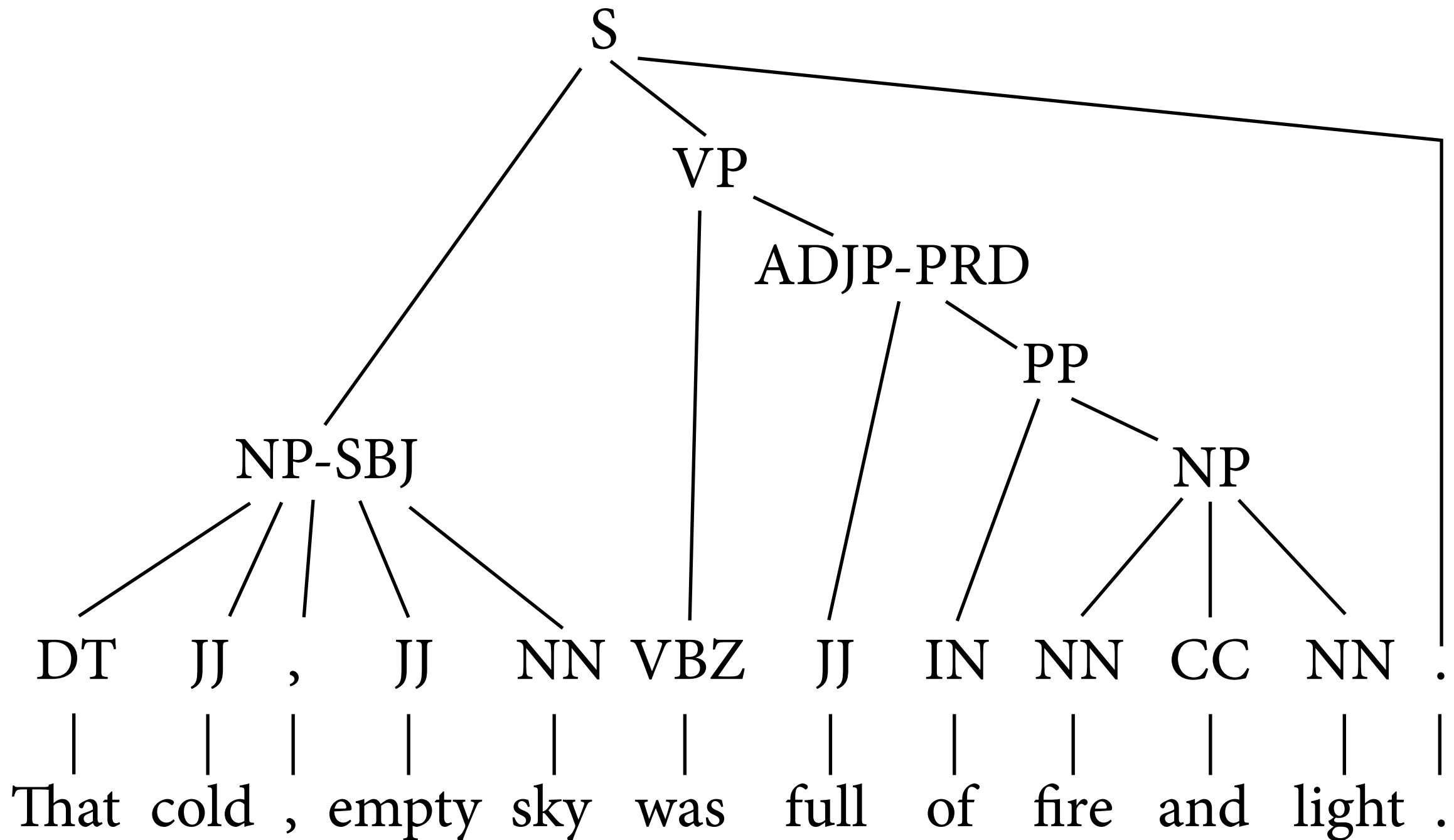
12 December 2023

Sentences have structure

grammatical functions

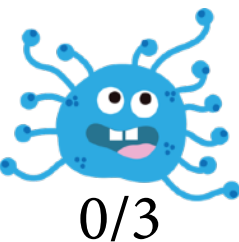


Penn Treebank



Parsing with neural models

- The syntactic structure of a sentence is a tree.
So we need models of syntax that know what a tree is.
- Right?
- Right??

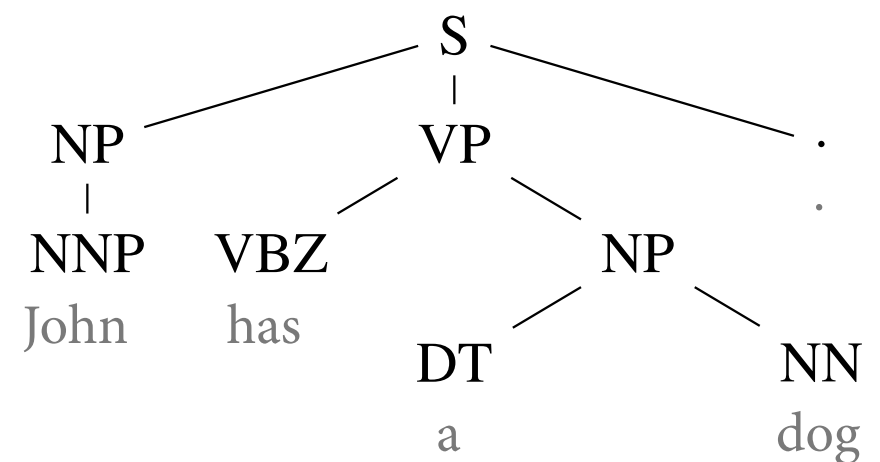


Trees as sequences

- Every tree can be uniquely and reversibly encoded as a sequence of symbols.
- Thus, if we know how to map sequences to sequences, we can also map sequences to trees.

input sequence:

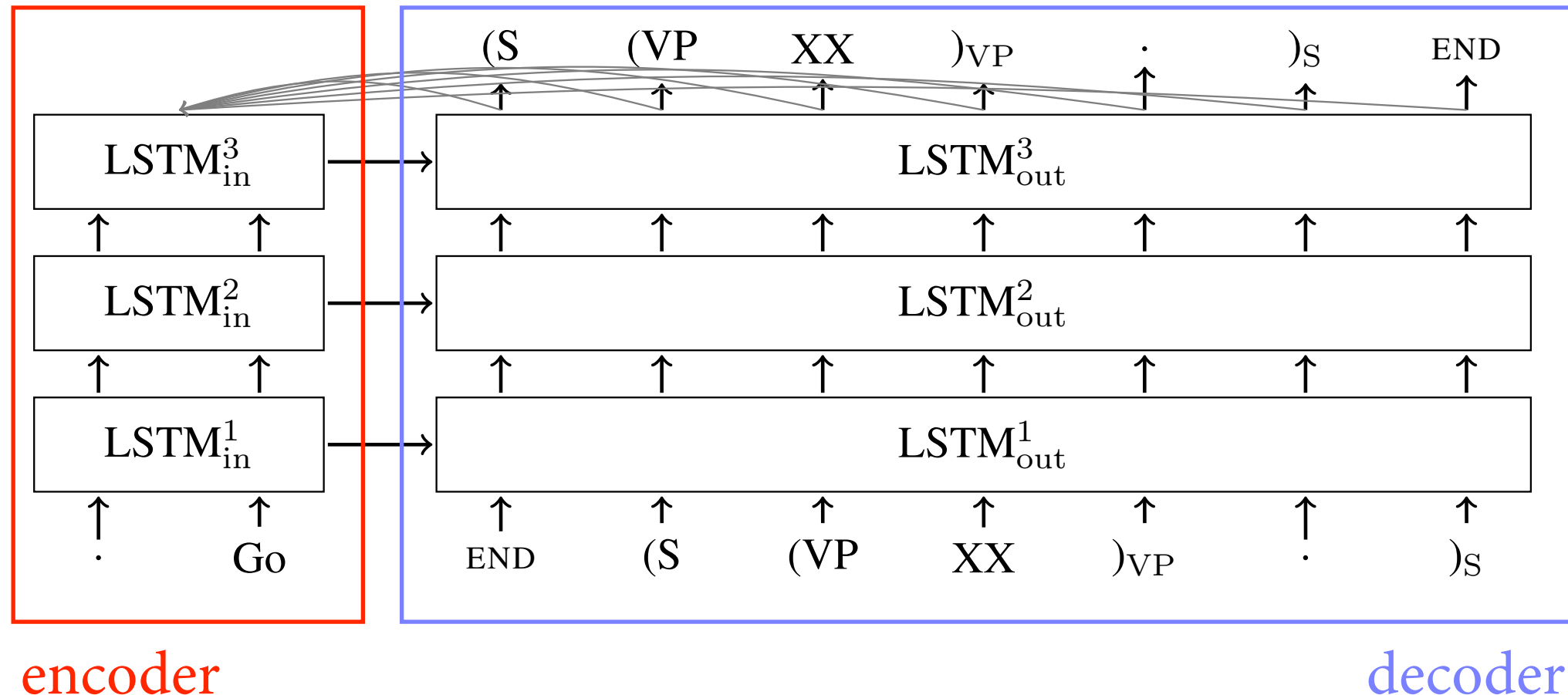
John has a dog .



output sequence:

(S (NP NNP)_{NP} (VP VBZ (NP DT NN)_{NP})_{VP} .)_S

Seq2seq model for parsing



Note: Input sequence is presented in reverse order.
This makes LSTM-based MT models more accurate
and was standard practice at the time.

Evaluation

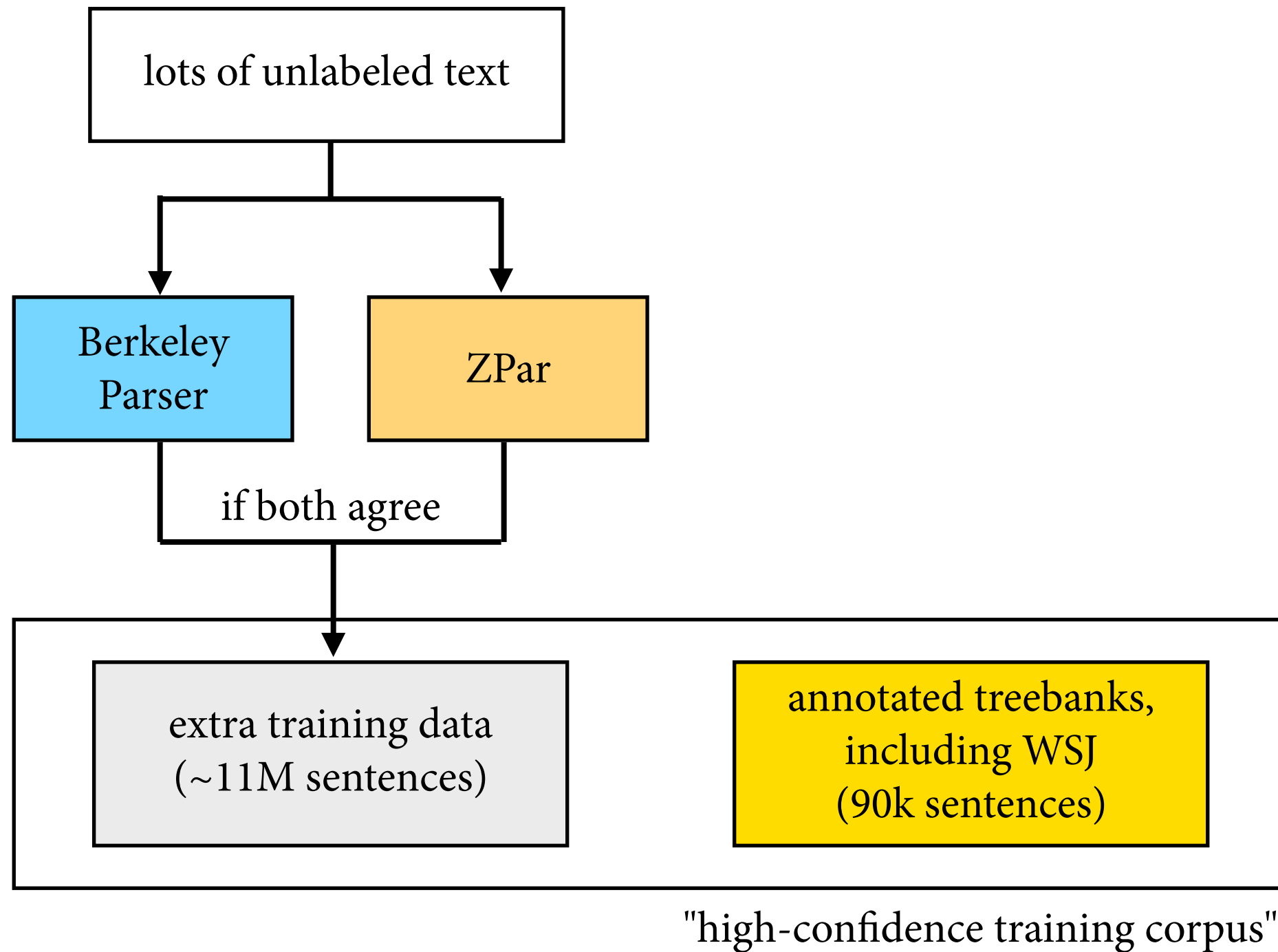
Parser	Training Set	WSJ 22	WSJ 23
baseline LSTM+D	WSJ only	< 70	< 70
LSTM+A+D	WSJ only	88.7	88.3
LSTM+A+D ensemble	WSJ only	90.7	90.5
baseline LSTM	BerkeleyParser corpus	91.0	90.5
LSTM+A	high-confidence corpus	92.8	92.1
Petrov et al. (2006) [12]	WSJ only	91.1	90.4
Zhu et al. (2013) [13]	WSJ only	N/A	90.4
Petrov et al. (2010) ensemble [14]	WSJ only	92.5	91.8
Zhu et al. (2013) [13]	semi-supervised	N/A	91.3
Huang & Harper (2009) [15]	semi-supervised	N/A	91.3
McClosky et al. (2006) [16]	semi-supervised	92.4	92.1

A = with cross-attention

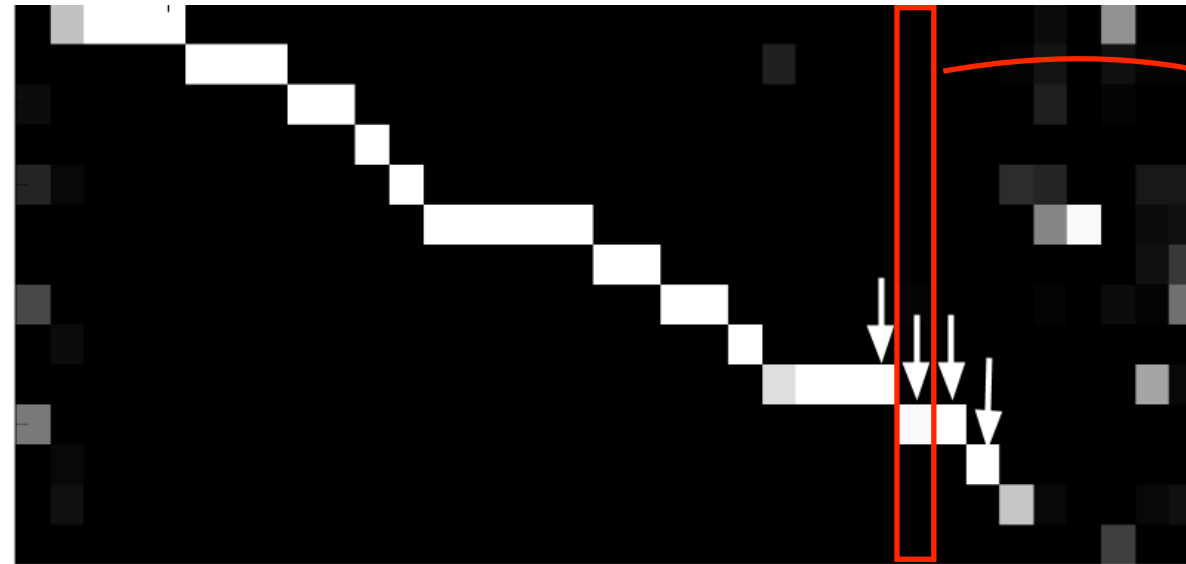
D = with dropout

> 98% of output sequences represent trees

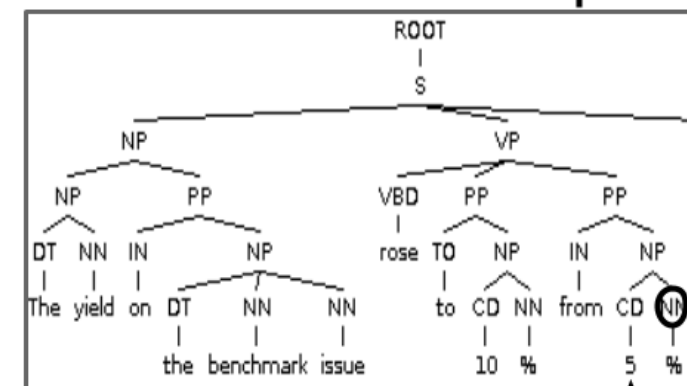
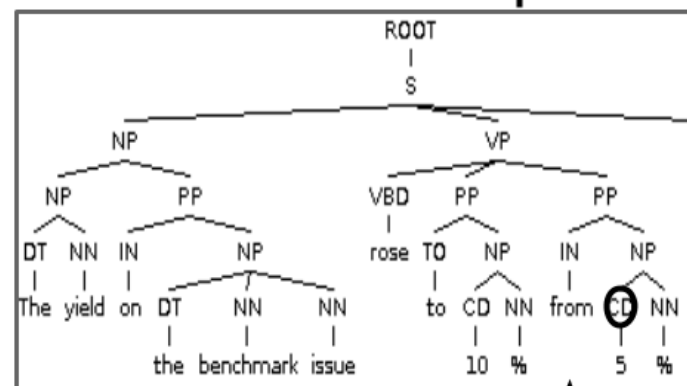
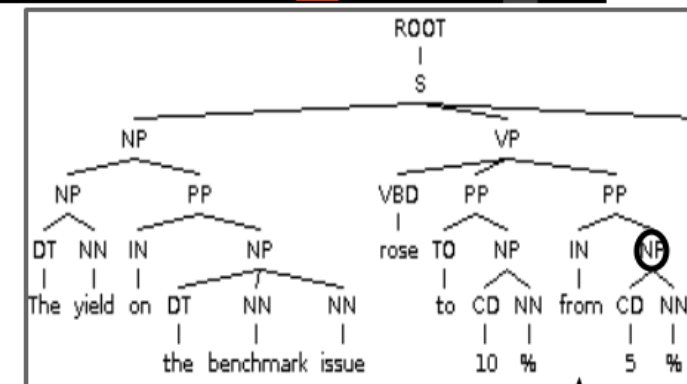
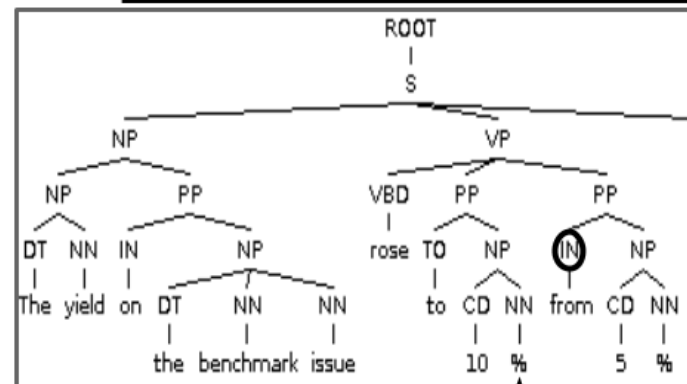
Silver data



Role of attention

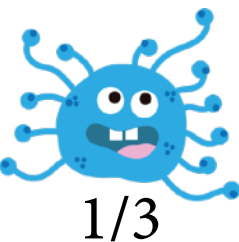


attention over inputs
at this timestep



Observations

- It was shocking to me that a bare seq2seq model would do so well on a structured prediction task.
- Results on high-confidence corpus are excellent, but results on original WSJ training set only okay.
- Can we do better?

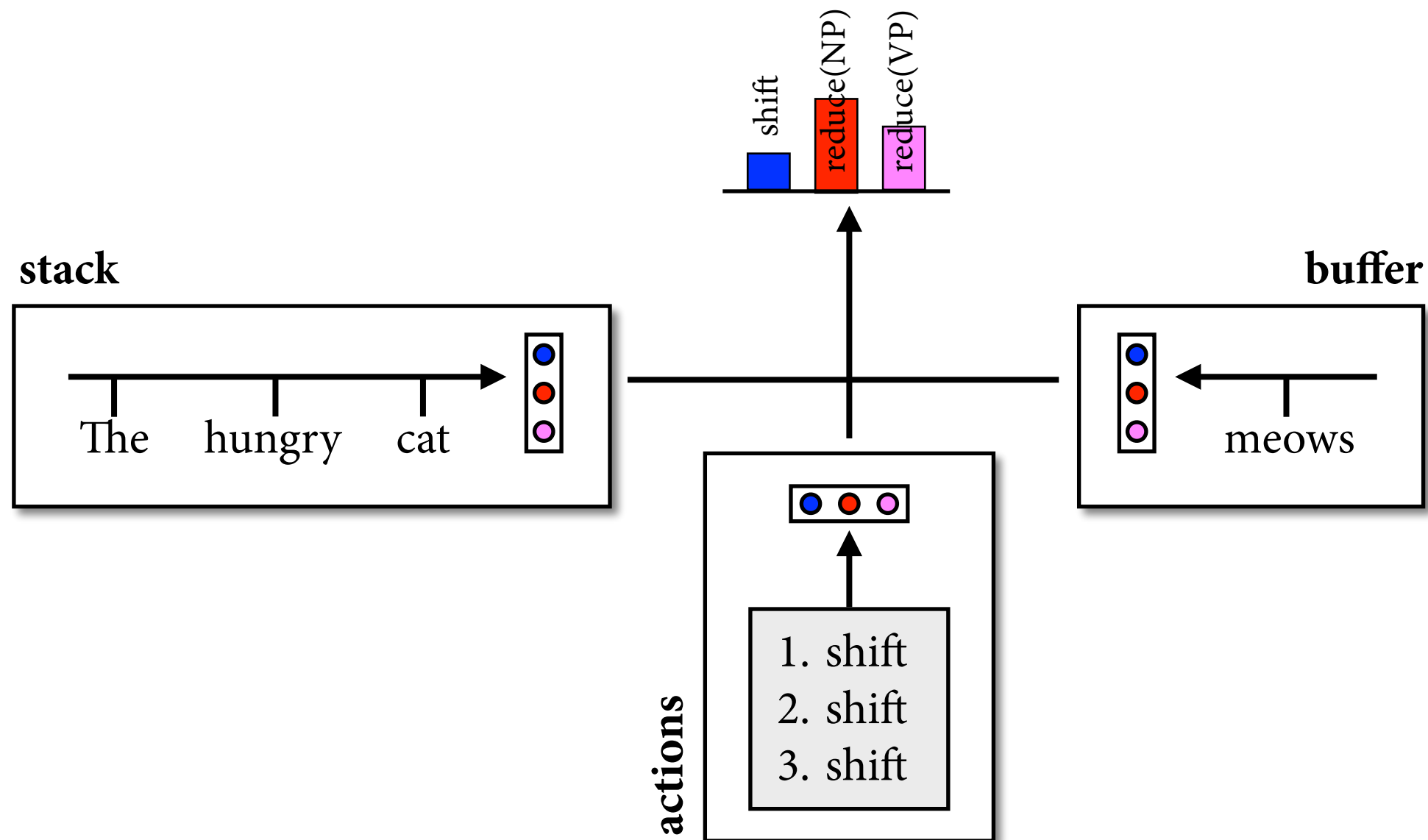


RNN Grammars

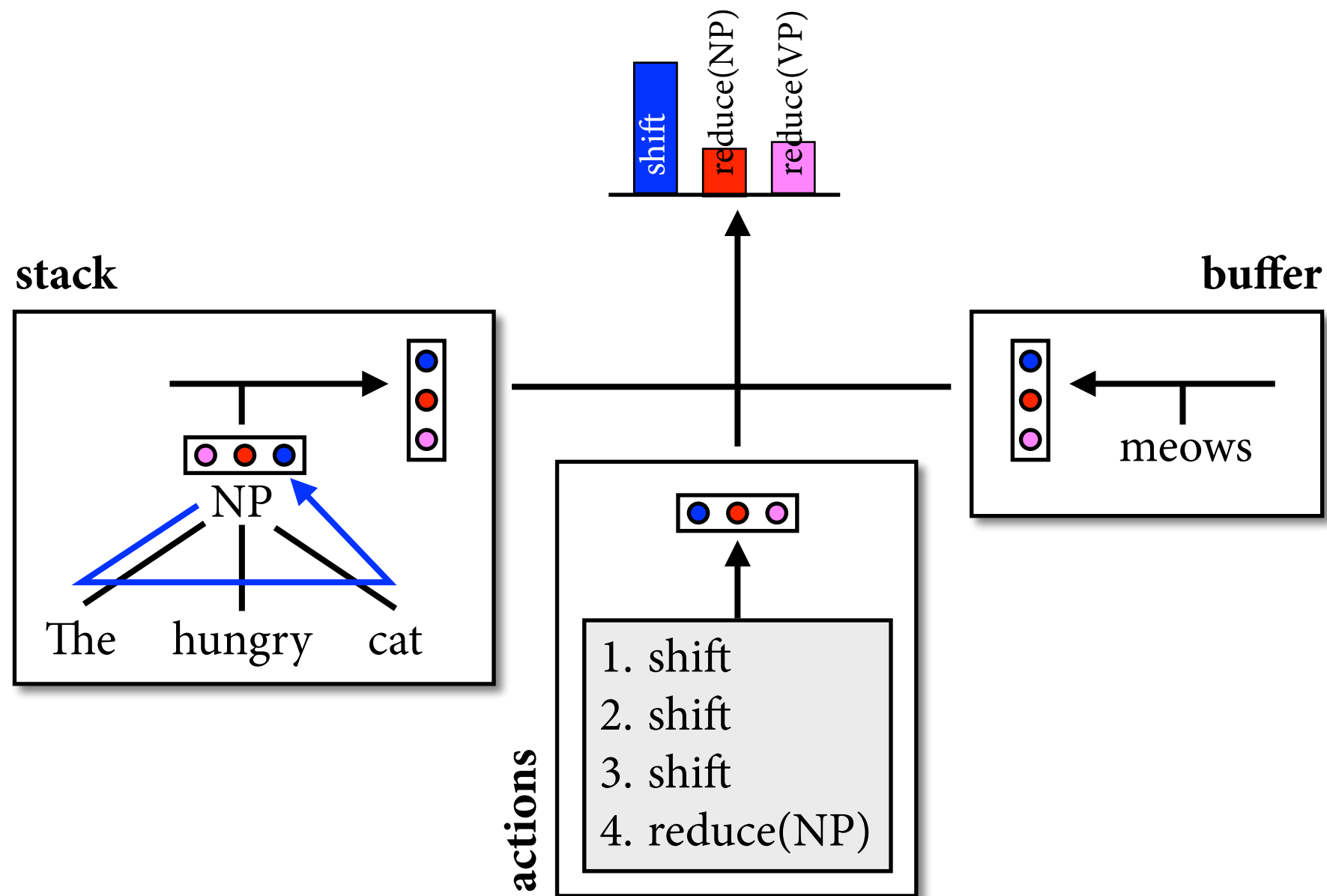


- Can we extend neural shift-reduce parsing into a parser with SOTA accuracy?
- Steps towards Recurrent Neural Network Grammars:
 - ▶ shift-reduce
 - ▶ discriminative RNNs: add top-down context
 - ▶ generative RNNs: better management of conditionals
- This is a *neurosymbolic* and not a purely neural model: It has awareness of trees and uses them to compute representations.

Neural shift-reduce



Neural shift-reduce



Neural shift-reduce

- Advantages:
 - ▶ no grammar needed
 - ▶ linear runtime
- Limitations:
 - ▶ risk of incorrect early parser actions
 - ▶ no top-down context

Discriminative RNNs

- Add top-down context:
 - ▶ add symbols for "open nonterminals" to the stack, e.g. "(S"
 - ▶ add parser action "NT(X)" to add open nonterminal "(X" to the stack.
- Picks up ideas from Earley parser for CFGs.
- Note that Vinyals parser has these "open nonterminals" too.

Example

Input: *The hungry cat meows .*

	Stack	Buffer	Action
0		<i>The hungry cat meows .</i>	NT(S)
1	(S	<i>The hungry cat meows .</i>	NT(NP)
2	(S (NP	<i>The hungry cat meows .</i>	SHIFT
3	(S (NP <i>The</i>	<i>hungry cat meows .</i>	SHIFT
4	(S (NP <i>The hungry</i>	<i>cat meows .</i>	SHIFT
5	(S (NP <i>The hungry cat</i>	<i>meows .</i>	REDUCE
6	(S (NP <i>The hungry cat</i>)	<i>meows .</i>	NT(VP)
7	(S (NP <i>The hungry cat</i>) (VP	<i>meows .</i>	SHIFT
8	(S (NP <i>The hungry cat</i>) (VP <i>meows</i>	<i>.</i>	REDUCE
9	(S (NP <i>The hungry cat</i>) (VP <i>meows</i>)	<i>.</i>	SHIFT
10	(S (NP <i>The hungry cat</i>) (VP <i>meows</i>) .		REDUCE
11	(S (NP <i>The hungry cat</i>) (VP <i>meows</i>) .)		

Training and parsing

- Convert parse trees from the training set into sequences of actions (uniquely).
- Train NN to predict correct parser action in context using MLE.
- At test time, simply predict actions.
- This gives an f-score of 91.2 on the standard WSJ split. Pretty good - can we do better?

Generative RNNs

- Traditional distinction of parsing models:
 - ▶ *Discriminative* models represent prob dist $P(t \mid w)$.
 - ▶ *Generative* models represent prob dist $P(t, w)$.
- PCFGs are generative models and can be parsed efficiently. More complex generative models can't.
- Generative RNNs: Keep NT and REDUCE actions, replace SHIFT with $\text{GEN}(word)$.
 - ▶ Get rid of buffer: we are generating the string, not reading it.

Generative RNNGs

	Stack	Terminals	Action
0			NT(S)
1	(S		NT(NP)
2	(S (NP		GEN(<i>The</i>)
3	(S (NP <i>The</i>	<i>The</i>	GEN(<i>hungry</i>)
4	(S (NP <i>The</i> <i>hungry</i>	<i>The</i> <i>hungry</i>	GEN(<i>cat</i>)
5	(S (NP <i>The</i> <i>hungry</i> <i>cat</i>	<i>The</i> <i>hungry</i> <i>cat</i>	REDUCE
6	(S (NP <i>The hungry cat</i>)	<i>The</i> <i>hungry</i> <i>cat</i>	NT(VP)
7	(S (NP <i>The hungry cat</i>) (VP	<i>The</i> <i>hungry</i> <i>cat</i>	GEN(<i>meows</i>)
8	(S (NP <i>The hungry cat</i>) (VP <i>meows</i>	<i>The</i> <i>hungry</i> <i>cat</i> <i>meows</i>	REDUCE
9	(S (NP <i>The hungry cat</i>) (VP <i>meows</i>)	<i>The</i> <i>hungry</i> <i>cat</i> <i>meows</i>	GEN(.)
10	(S (NP <i>The hungry cat</i>) (VP <i>meows</i>) .	<i>The</i> <i>hungry</i> <i>cat</i> <i>meows</i> .	REDUCE
11	(S (NP <i>The hungry cat</i>) (VP <i>meows</i>) .)	<i>The</i> <i>hungry</i> <i>cat</i> <i>meows</i> .	

Parsing with generative RNNs

- We are looking for

$$\arg \max_{t \in \text{pareses}(w)} P(t, w),$$

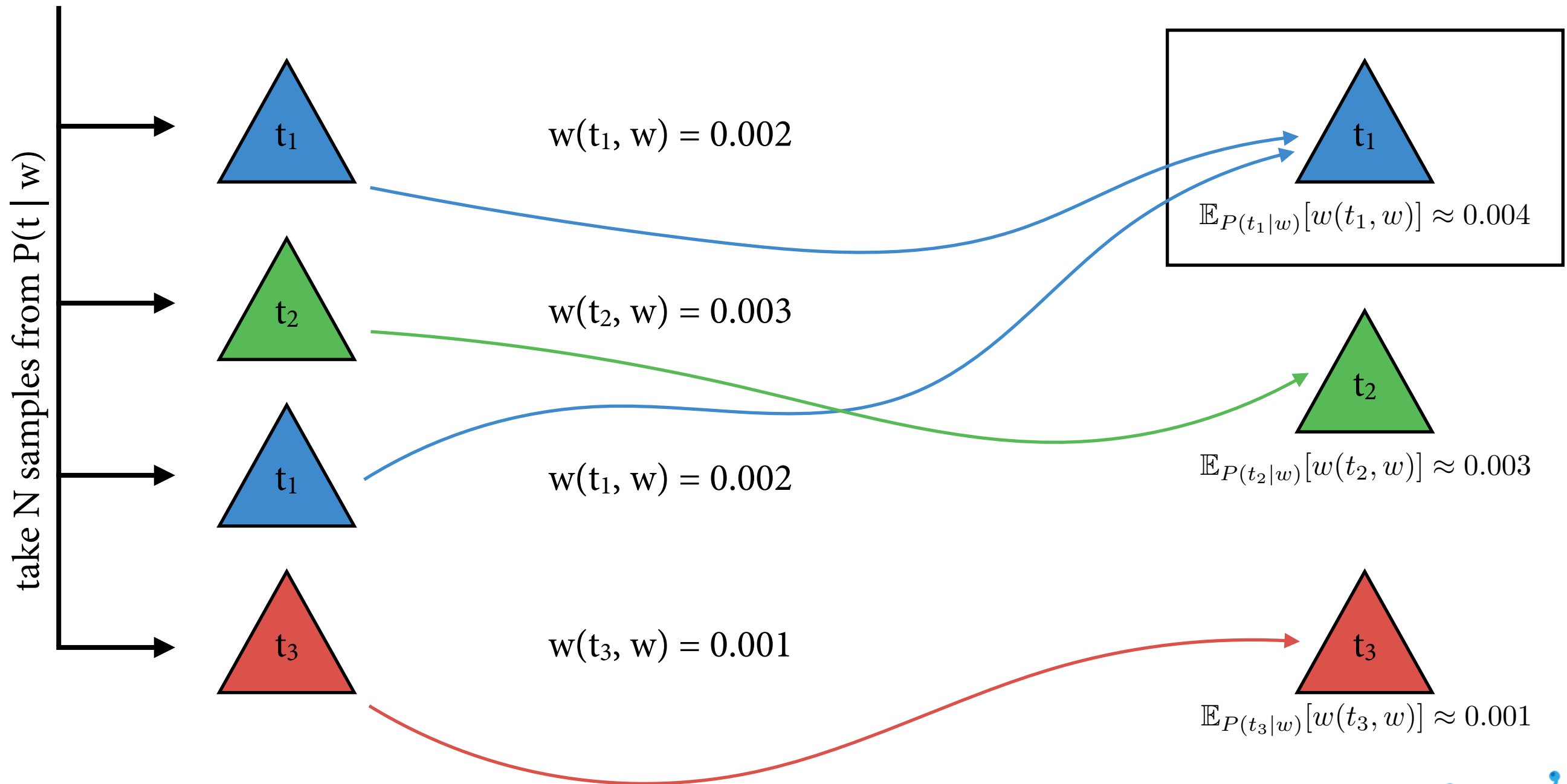
but the generative RNNG doesn't know anything about w , so it is totally unclear how to compute it.

- We *could* get the argmax for $P(t \mid w)$, but this is not the right model.
- We can also sample from $P(t \mid w)$. Let's do it many times and reweight \rightarrow *importance sampling*.

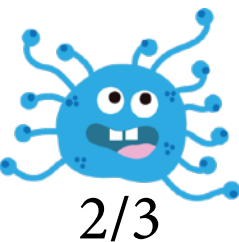
Importance sampling

The hungry cat meows.

$$w(t, w) = P(t, w) / P(t \mid w)$$

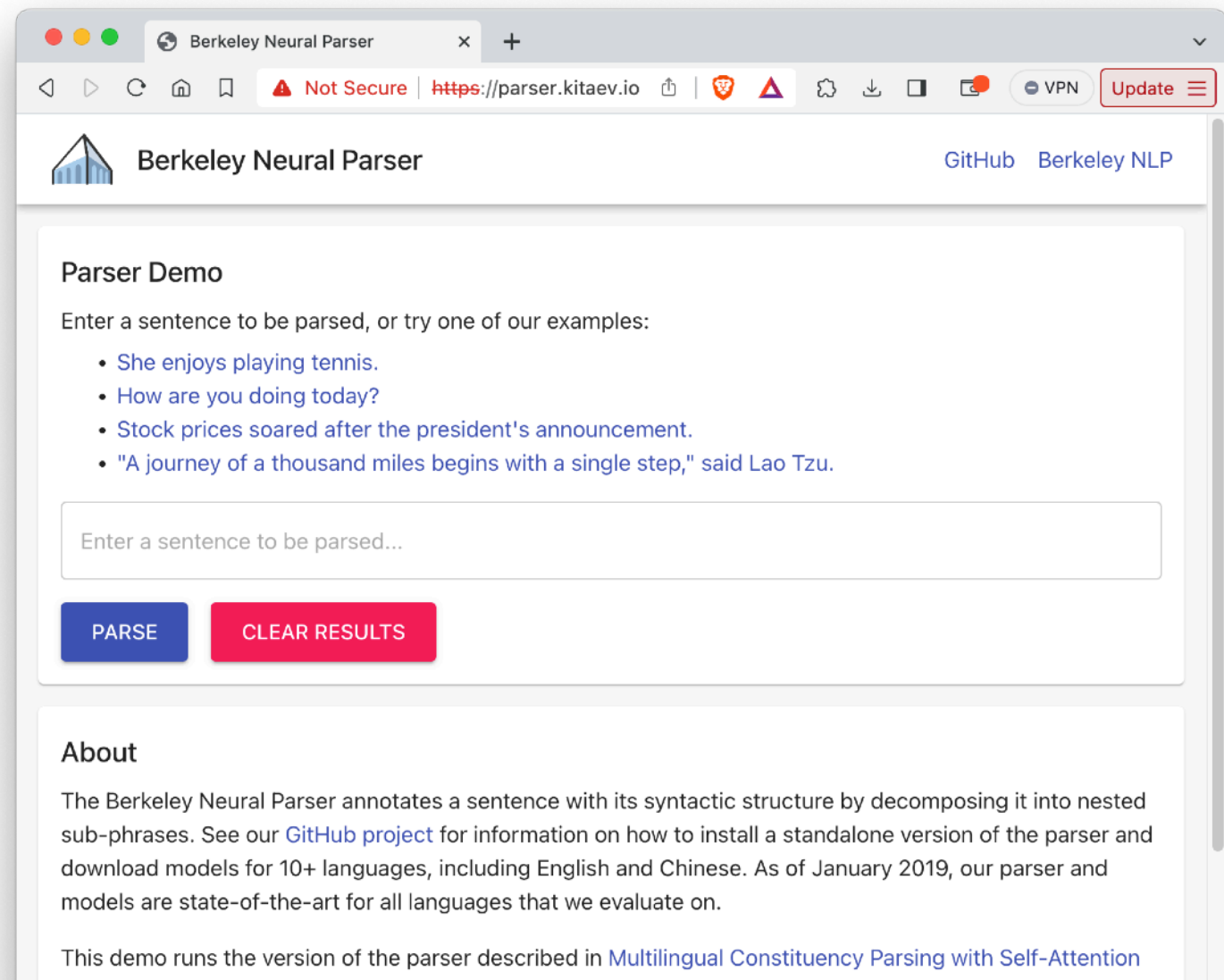


f-score on PTB: 93.3



Neural Berkeley Parser

- RNNs are cool, but they seem complicated.
 - ▶ Parallel training is a nightmare.
- Couldn't we just do CKY parsing with neural scores?



Core of Viterbi-CKY

Viterbi CKY parse chart: $\text{Ch}(i, k) = \{(A, p) \mid p = \max_{d: A \Rightarrow^* w_i \dots w_{k-1}} P(d)\}$

VP: 0.0036	NP: 0.006	N: 0.014	PP: 0.12	... in my pyjamas
VP: 0.06	NP: 0.12	N: 0.3	... elephant	in my pyjamas
	Det: 0.5	... an	elephant	
V: 1.0	... shot	an		shot

Core of Viterbi-CKY

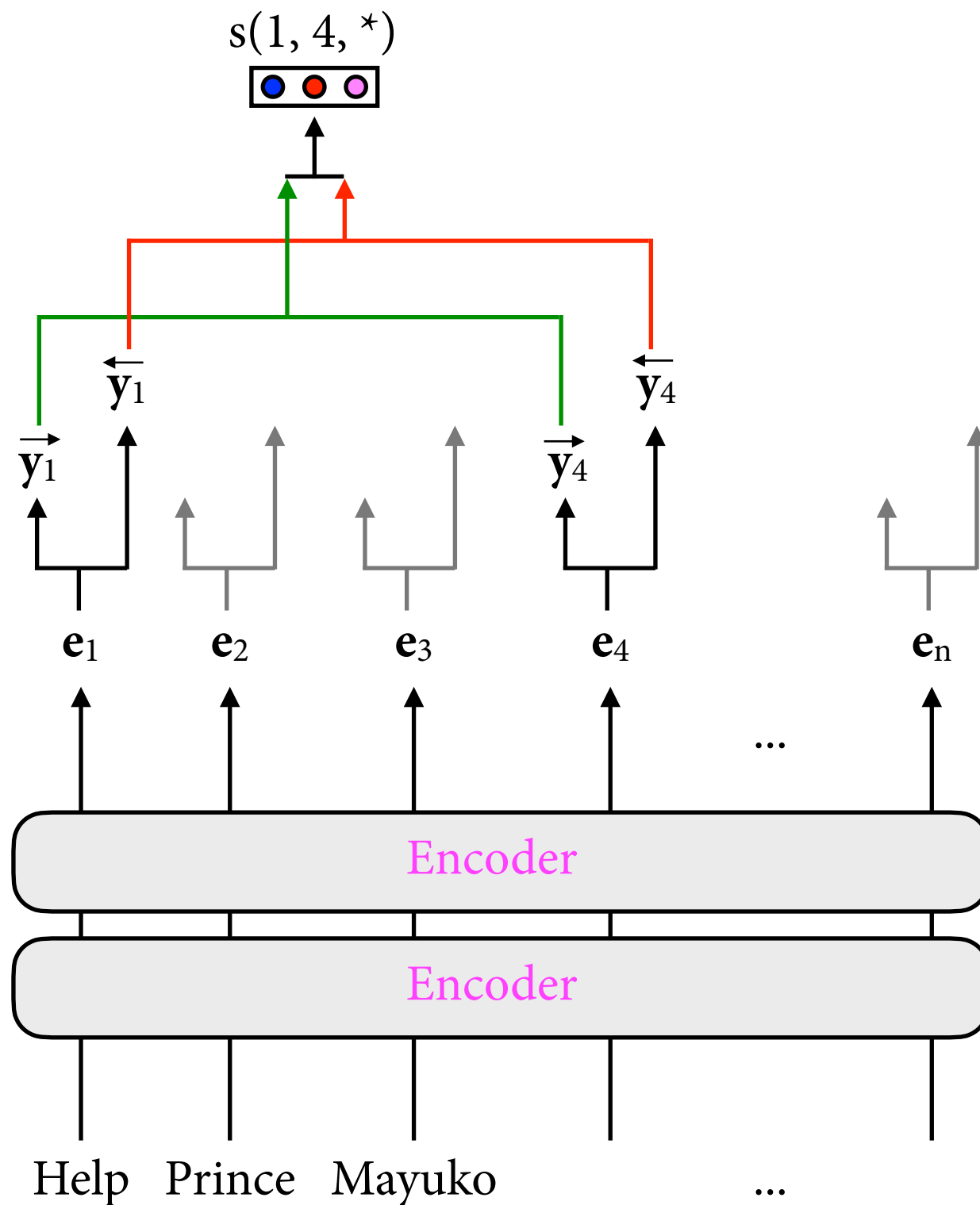
$V(i, j)$ = $\max_{t \text{ for } (i, j)} \sum_{(i', j', \ell) \in t} s(i', j', \ell) = \max_{\ell} s(i, j, \ell) + \max_k (V(i, k) + V(k, j))$

VP: 5 VP/46	NP: 0 NP/26	N: 5 N/16	PP: 8 PP/8	... in my pyjamas
VP: 5 VP/33	NP: 10 X: -20 ... NP/18	N: 3 N/3	... elephant in my pyjamas	
X: -10 X/5	Det: 5 Det/5	... an elephant		
V: 10 V/10 shot	... shot an			

Assume local score $s(i, j, l)$.

Goal: Maximize $\sum_{(i, j, \ell) \in t} s(i, j, \ell)$

Computing local scores



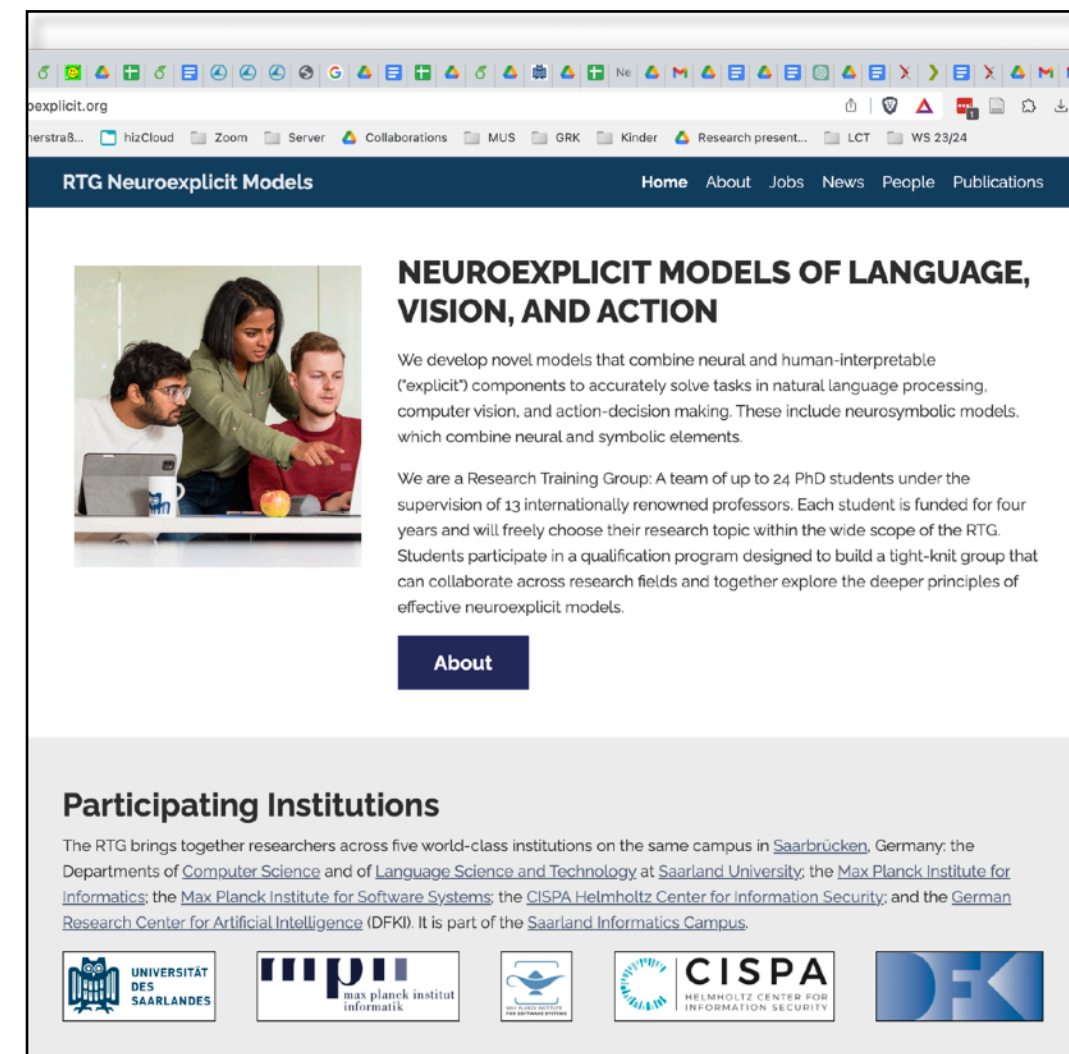
Training uses *hinge loss*: maximize score difference of gold tree to the others.

Evaluation

Encoder Architecture	F1 (dev)	Δ
LSTM (Gaddy et al., 2018)	92.24	-0.43
Self-attentive (Section 2)	92.67	0.00
+ Factored (Section 3)	93.15	0.48
+ CharLSTM (Section 5.1)	93.61	0.94
+ ELMo (Section 5.2)	95.21	2.54

Summary

- Neural models are good at parsing, but neurosymbolic models are better.
 - ▶ RNNGs = shift-reduce plus more complex parsing
 - ▶ Berkeley = Viterbi-CKY plus more complex training
- Neurosymbolic models are really interesting, but need clearer understanding of design principles.



The screenshot shows the website for the RTG Neuroexplicit Models. The browser address bar shows 'neuroexplicit.org'. The website has a dark blue header with the title 'RTG Neuroexplicit Models' and navigation links: Home, About, Jobs, News, People, Publications. Below the header, there is a section titled 'NEUROEXPLICIT MODELS OF LANGUAGE, VISION, AND ACTION'. To the left of this text is a photo of three people (two men and one woman) working together at a table with a laptop. The text describes the group's focus on developing novel models that combine neural and human-interpretable ('explicit') components for tasks in natural language processing, computer vision, and action-decision making. It also describes the group as a Research Training Group with up to 24 PhD students supervised by 13 professors. Below this text is a dark blue button labeled 'About'. At the bottom of the page, there is a section titled 'Participating Institutions' which lists five institutions: Universität des Saarlandes, MPI (Max Planck Institute for Informatics), CISPA (Helmholtz Center for Information Security), and DFKI (German Research Center for Artificial Intelligence). Logos for each of these institutions are displayed at the very bottom.

neuroexplicit.org

herstraß... h1zCloud Zoom Server Collaborations MUS GRK Kinder Research present... LCT WS 23/24

RTG Neuroexplicit Models Home About Jobs News People Publications

NEUROEXPLICIT MODELS OF LANGUAGE, VISION, AND ACTION

We develop novel models that combine neural and human-interpretable ("explicit") components to accurately solve tasks in natural language processing, computer vision, and action-decision making. These include neurosymbolic models, which combine neural and symbolic elements.

We are a Research Training Group: A team of up to 24 PhD students under the supervision of 13 internationally renowned professors. Each student is funded for four years and will freely choose their research topic within the wide scope of the RTG. Students participate in a qualification program designed to build a tight-knit group that can collaborate across research fields and together explore the deeper principles of effective neuroexplicit models.

About

Participating Institutions

The RTG brings together researchers across five world-class institutions on the same campus in [Saarbrücken](#), Germany: the Departments of [Computer Science](#) and of [Language Science and Technology](#) at [Saarland University](#); the [Max Planck Institute for Informatics](#); the [Max Planck Institute for Software Systems](#); the [CISPA Helmholtz Center for Information Security](#); and the [German Research Center for Artificial Intelligence \(DFKI\)](#). It is part of the [Saarland Informatics Campus](#).

UNIVERSITÄT DES SAARLANDES MPI max planck institut informatik CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY DFKI