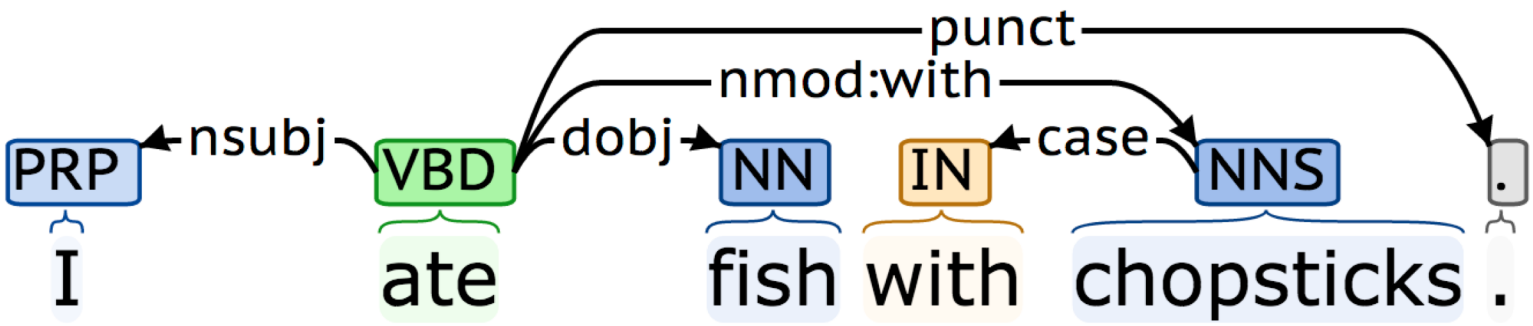
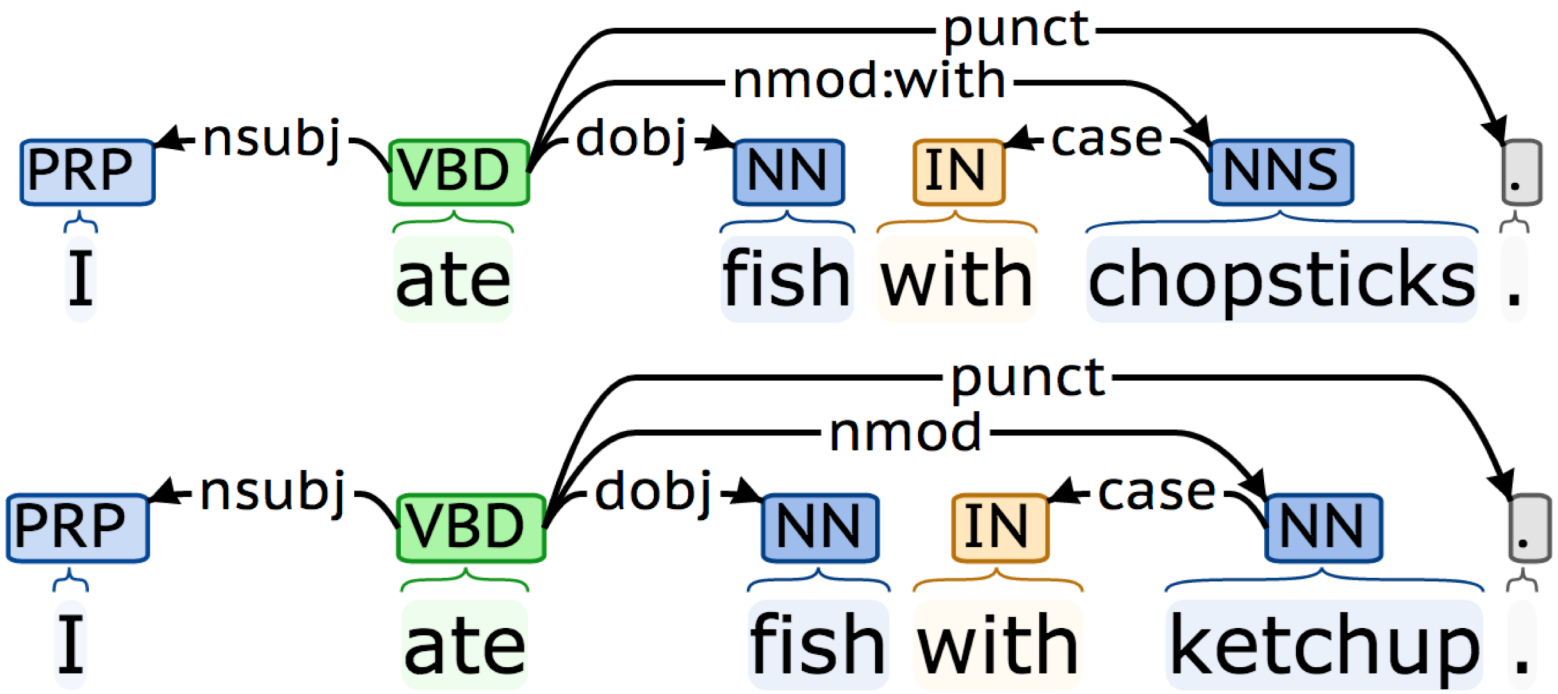
Arc-swift: A Novel Transition System for Dependency Parsing

Peng Qi and Christopher D. Manning
Stanford University

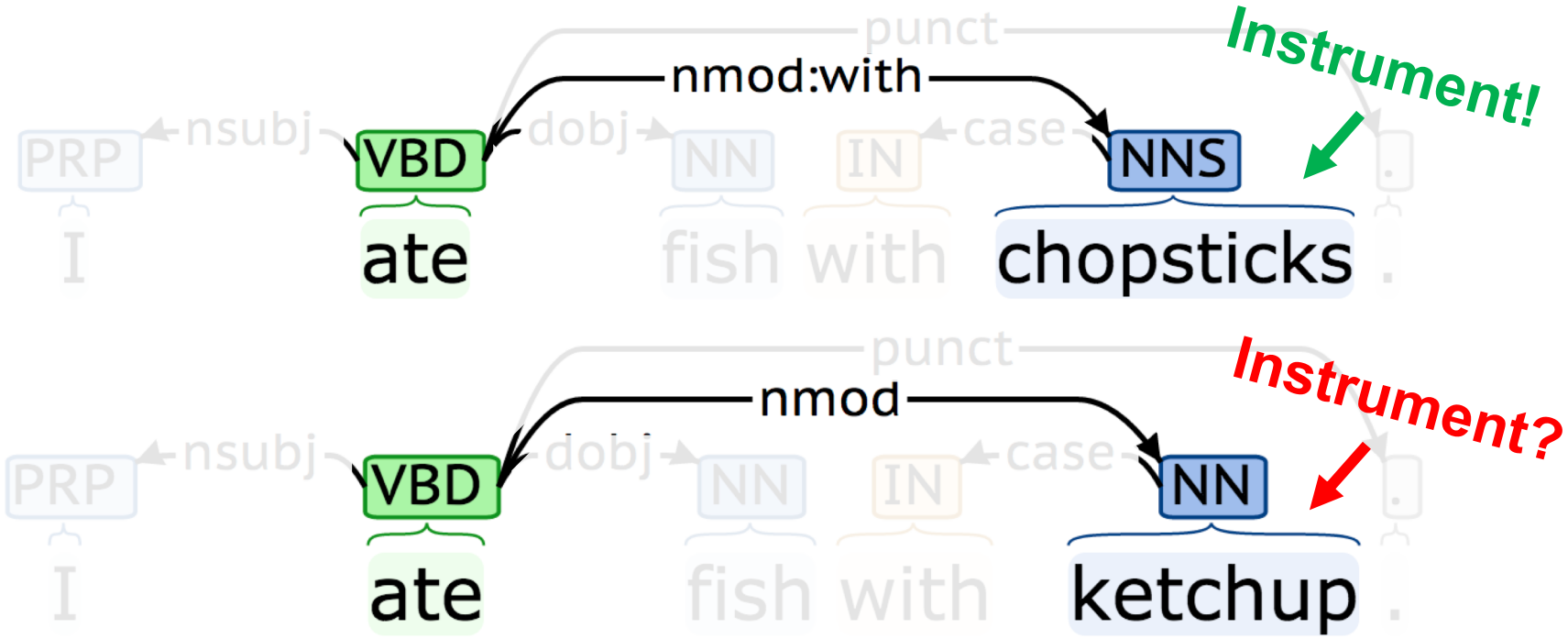
Dependency parsing



Dependency parsing

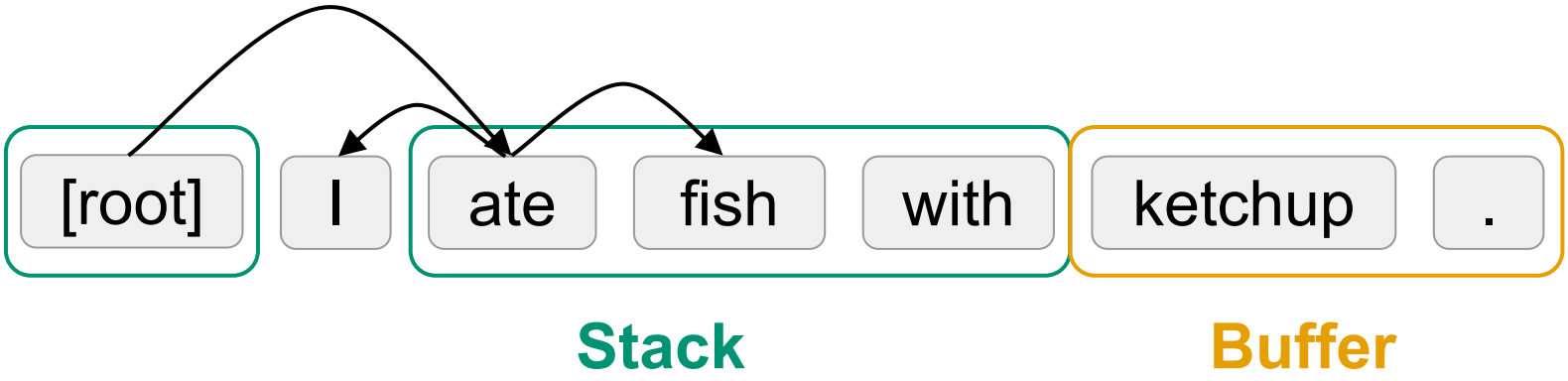


Dependency parsing



http://corenlp.run, 2016/12/21

Transition-based dependency parsing



Transition systems

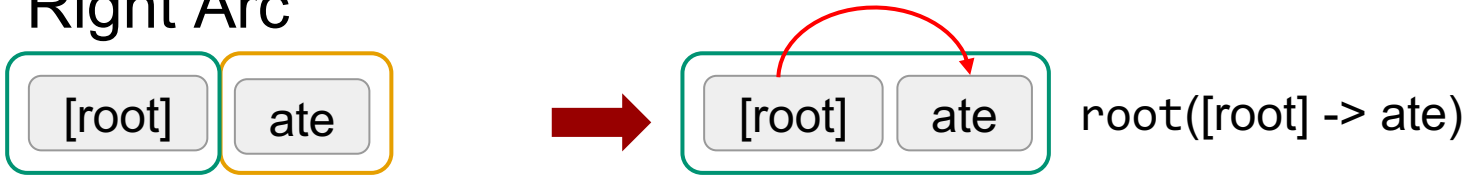
- arc-standard (Nivre, 2004)
- arc-eager (Nivre, 2003; Nivre 2008)
- arc-hybrid (Kuhlmann et al, 2011)

arc-eager

Shift

Left Arc

Right Arc



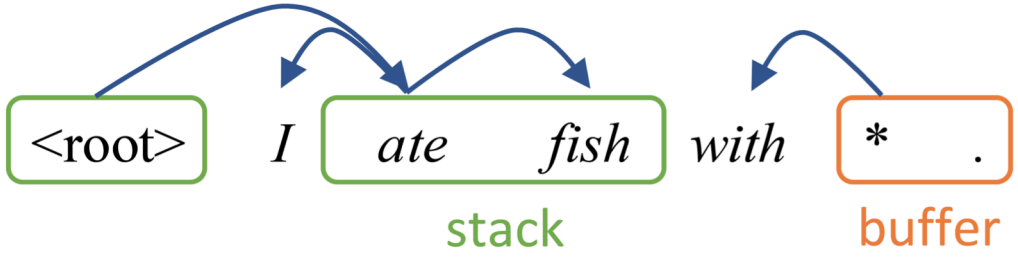
Reduce

Local transitions are difficult (arc-eager)

I ate fish with ketchup .
I ate fish with chopsticks .

Local transitions are difficult (arc-eager)

I ate fish with ketchup .
I ate fish with chopsticks .



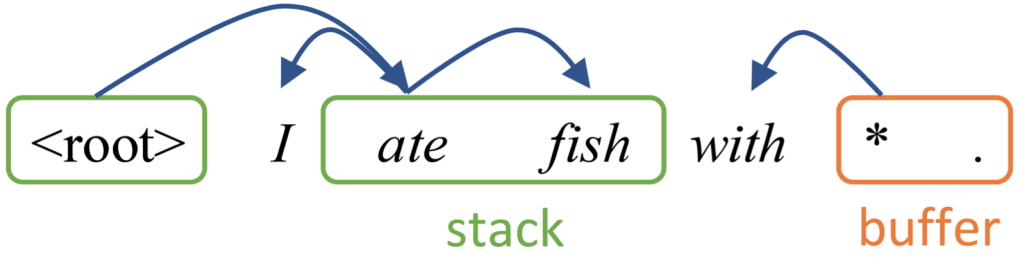
Enhancing local transitions with features

Single-word features (9)
$s_1.w$; $s_1.t$; $s_1.wt$; $s_2.w$; $s_2.t$;
$s_2.wt$; $b_1.w$; $b_1.t$; $b_1.wt$
Word-pair features (8)
$s_1.wt \circ s_2.wt$; $s_1.wt \circ s_2.w$; $s_1.wts_2.t$;
$s_1.w \circ s_2.wt$; $s_1.t \circ s_2.wt$; $s_1.w \circ s_2.w$
$s_1.t \circ s_2.t$; $s_1.t \circ b_1.t$
Three-word feaures (8)
$s_2.t \circ s_1.t \circ b_1.t$; $s_2.t \circ s_1.t \circ lc_1(s_1).t$;
$s_2.t \circ s_1.t \circ rc_1(s_1).t$; $s_2.t \circ s_1.t \circ lc_1(s_2).t$;
$s_2.t \circ s_1.t \circ rc_1(s_2).t$; $s_2.t \circ s_1.w \circ rc_1(s_2).t$;
$s_2.t \circ s_1.w \circ lc_1(s_1).t$; $s_2.t \circ s_1.w \circ b_1.t$

(Chen & Manning, 2014)

Local transitions are difficult (arc-eager)

I ate fish with ketchup .
I ate fish with chopsticks .

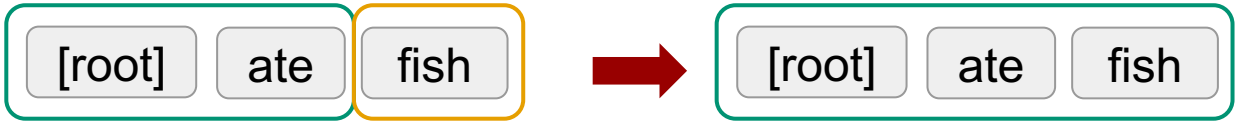


Key observation

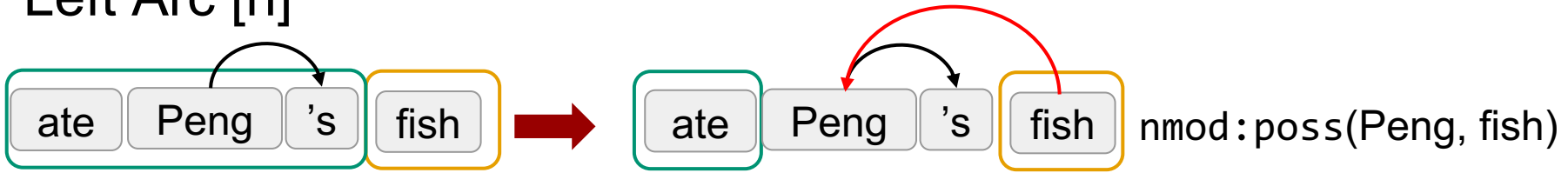
Relatedness of “ate”, “fish”, and “*” determines which arc should be induced.

arc-swift

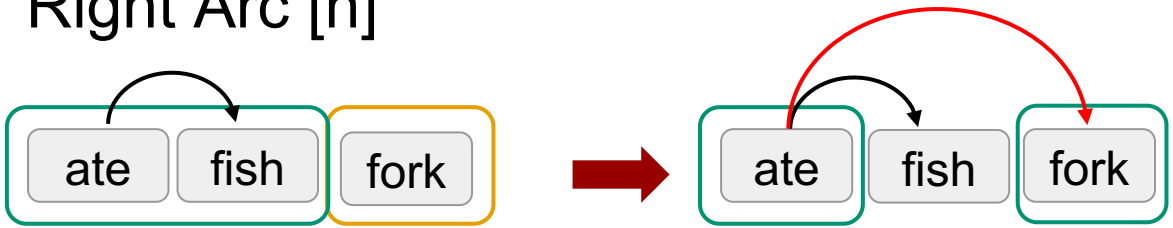
Shift



Left Arc [n]

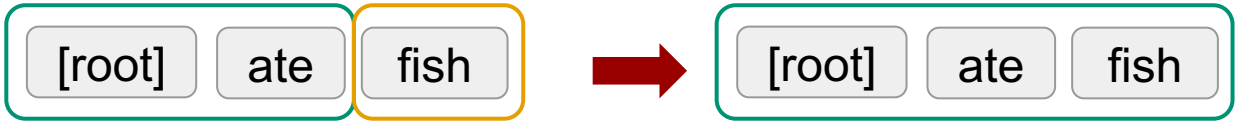


Right Arc [n]

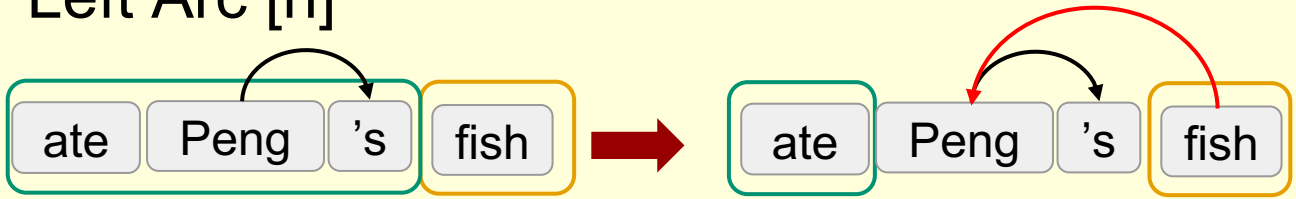


arc-swift

Shift

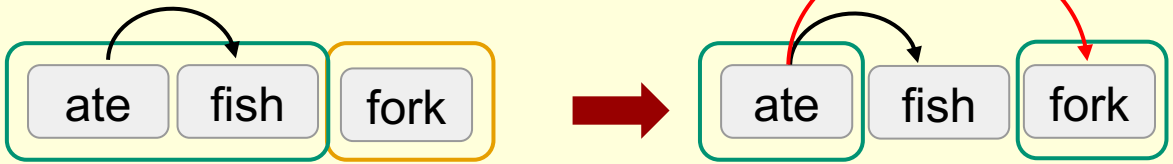


Left Arc [n]



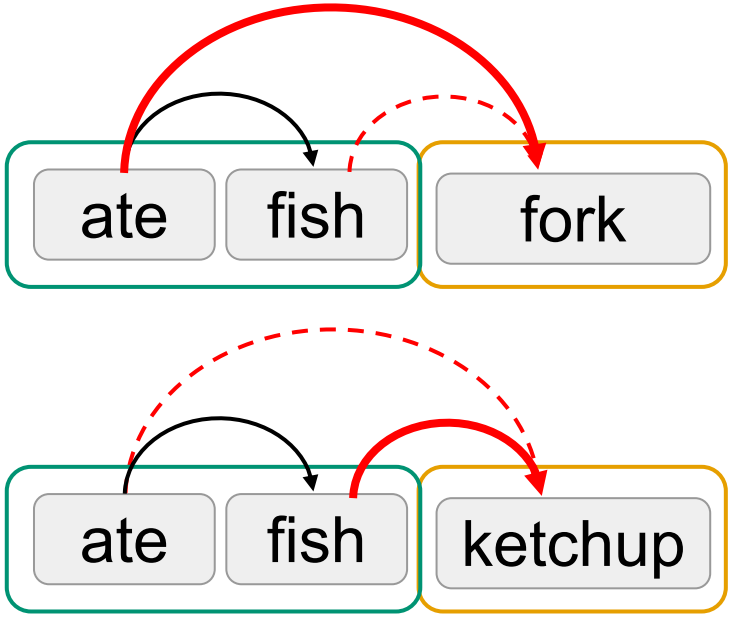
nmod:poss(Peng, fish)

Right Arc [n]

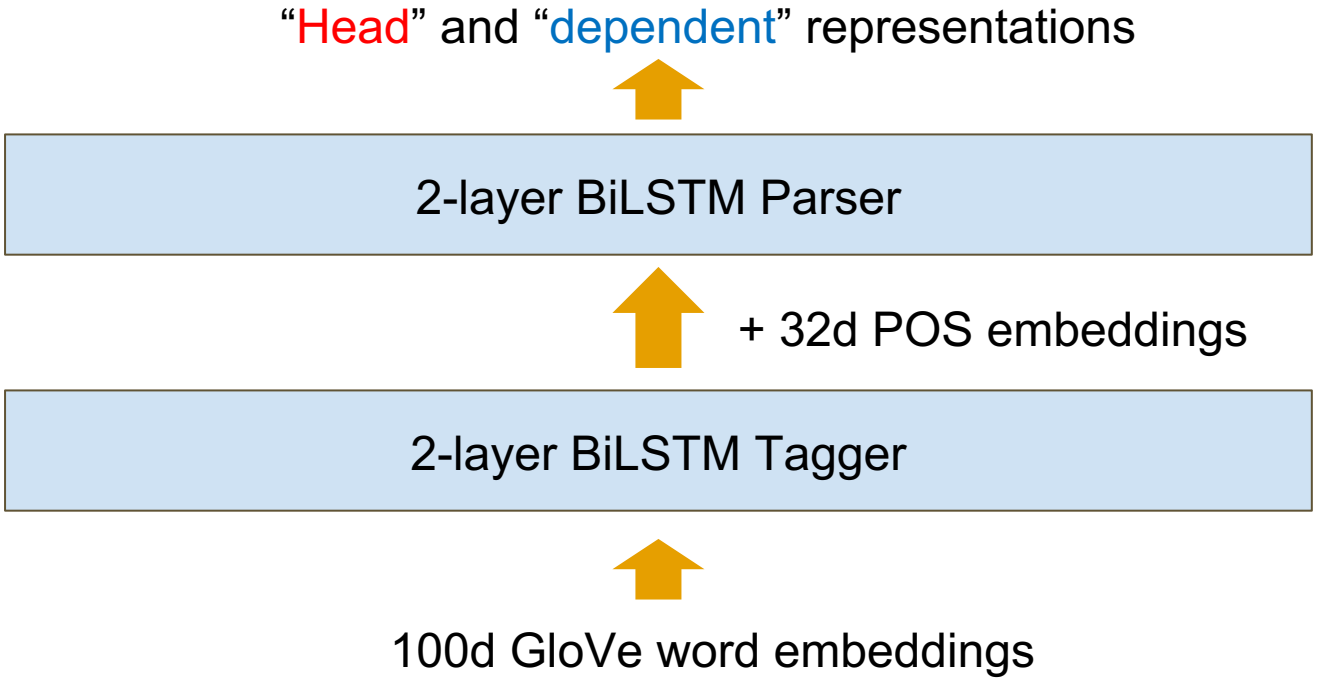


nmod(ate, fork)

Resolving arc confusion with arc-swift

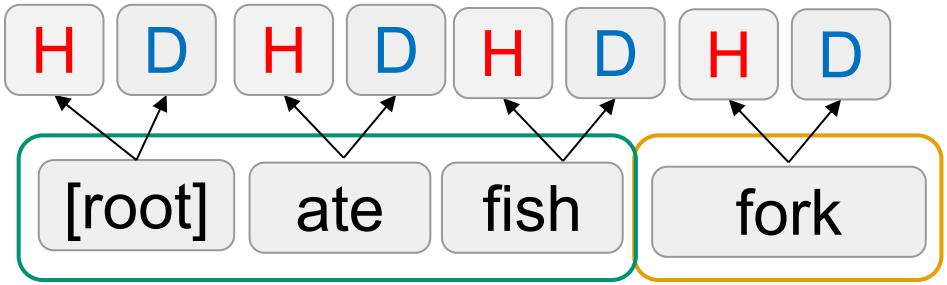


Model



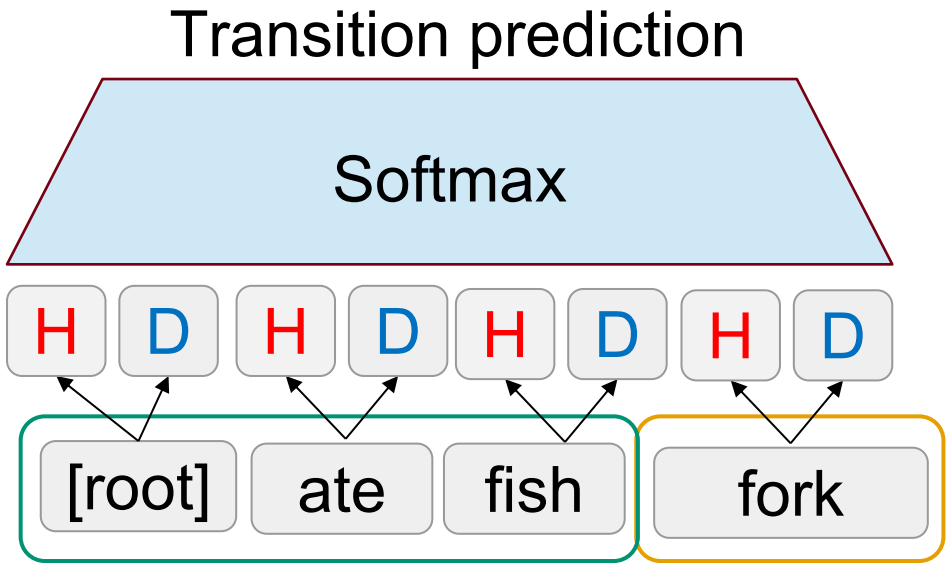
See also: (Kiperwasser & Goldberg, 2016)

Model (cont'd)



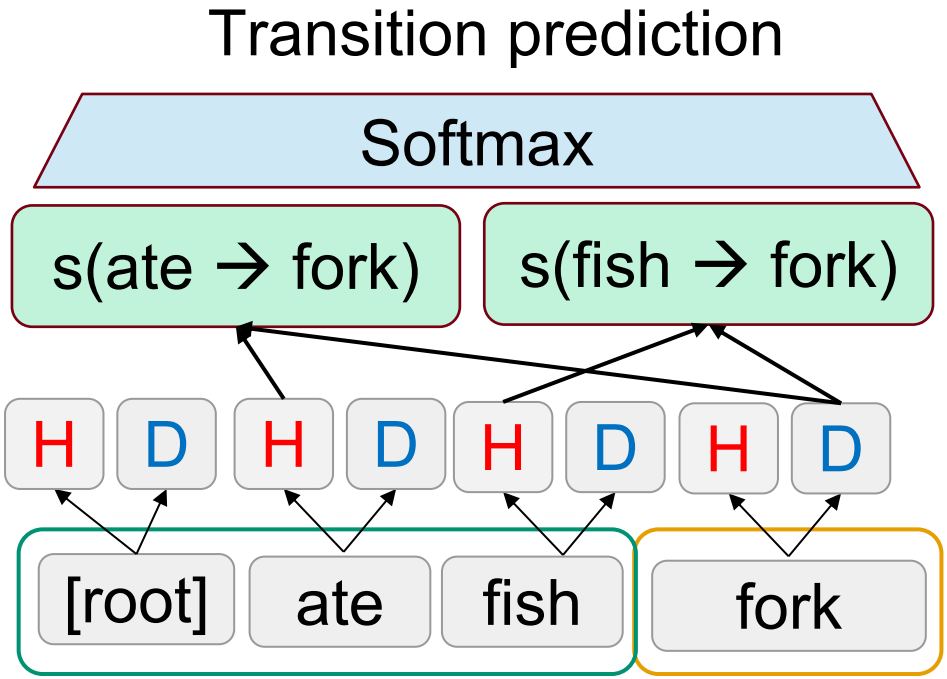
See also: (Kiperwasser & Goldberg, 2016)

Model (cont'd)



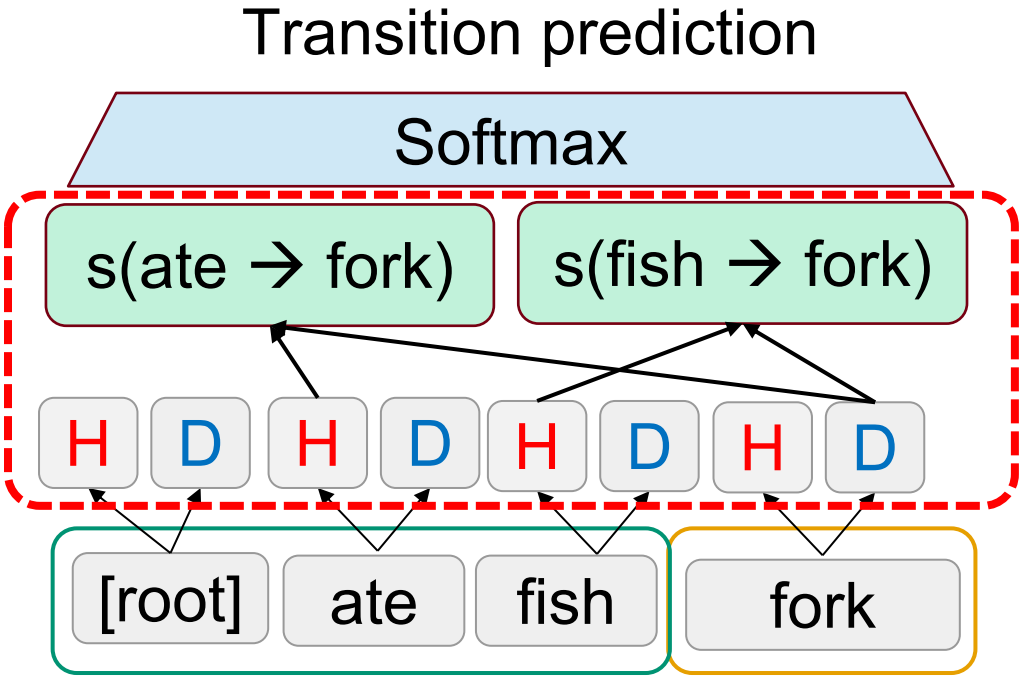
See also: (Kiperwasser & Goldberg, 2016)

Model (cont'd)



See also: (Kiperwasser & Goldberg, 2016)

Model (cont'd)



See also: (Kiperwasser & Goldberg, 2016)

Data & Evaluation

- Data
  - Penn Treebank (Marcus et al., 1999) WSJ portion converted to Stanford Dependencies (de Marneffe and Manning, 2008)
  - Universal Dependencies v1.3 (English) (Nivre et al., 2016)
- Evaluation
  - Unlabeled/labeled attachment score (UAS/LAS)
  - Punctuation removed

Results

Transition System	PTB UAS	PTB LAS	EN-UD UAS	EN-UD LAS
arc-standard	94.0	91.7	85.6	81.5
arc-hybrid	94.0	91.8	85.4	81.4
arc-eager-S	93.8	91.7	85.2	81.2
arc-eager-R	93.9	91.7	85.4	81.3
arc-swift	**94.3**	**92.2**	**86.1**	**82.2**

Results (cont'd)

Implementation	TransSys	Notes	PTB UAS	PTB LAS	EN-UD UAS	EN-UD LAS
Ours	arc-swift		*94.3*	*92.2*	**86.1**	**82.2**
Andor+ 2016	arc-standard	Feedforward, CRF loss, B=32	**94.6**	**92.8**	84.8*	80.4*
K&G 2016	arc-hybrid	Dynamic oracle	93.6	91.5		
Weiss+ 2015	arc-standard	B=8	94.0	92.1		
C&M 2014	arc-standard	Feedforward	91.8	89.6		

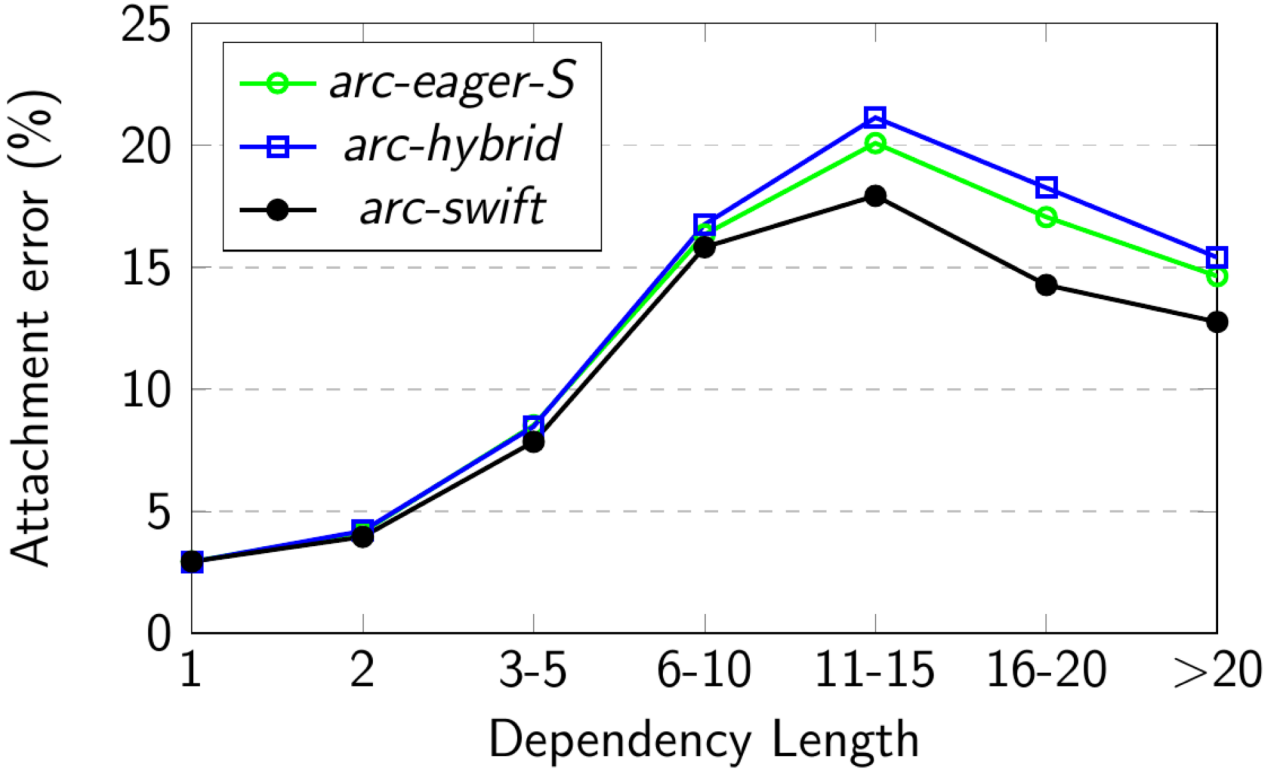
* https://github.com/tensorflow/models/blob/master/syntaxnet/g3doc/universal.md

Significance test (UAS/LAS)

Row > col?	arc-eager-S	arc-standard	arc-hybrid	arc-eager-R
arc-swift	***/***	***/***	***/***	***/***
arc-eager-S		-	-	*/-
arc-standard			-	*/-
arc-hybrid				-

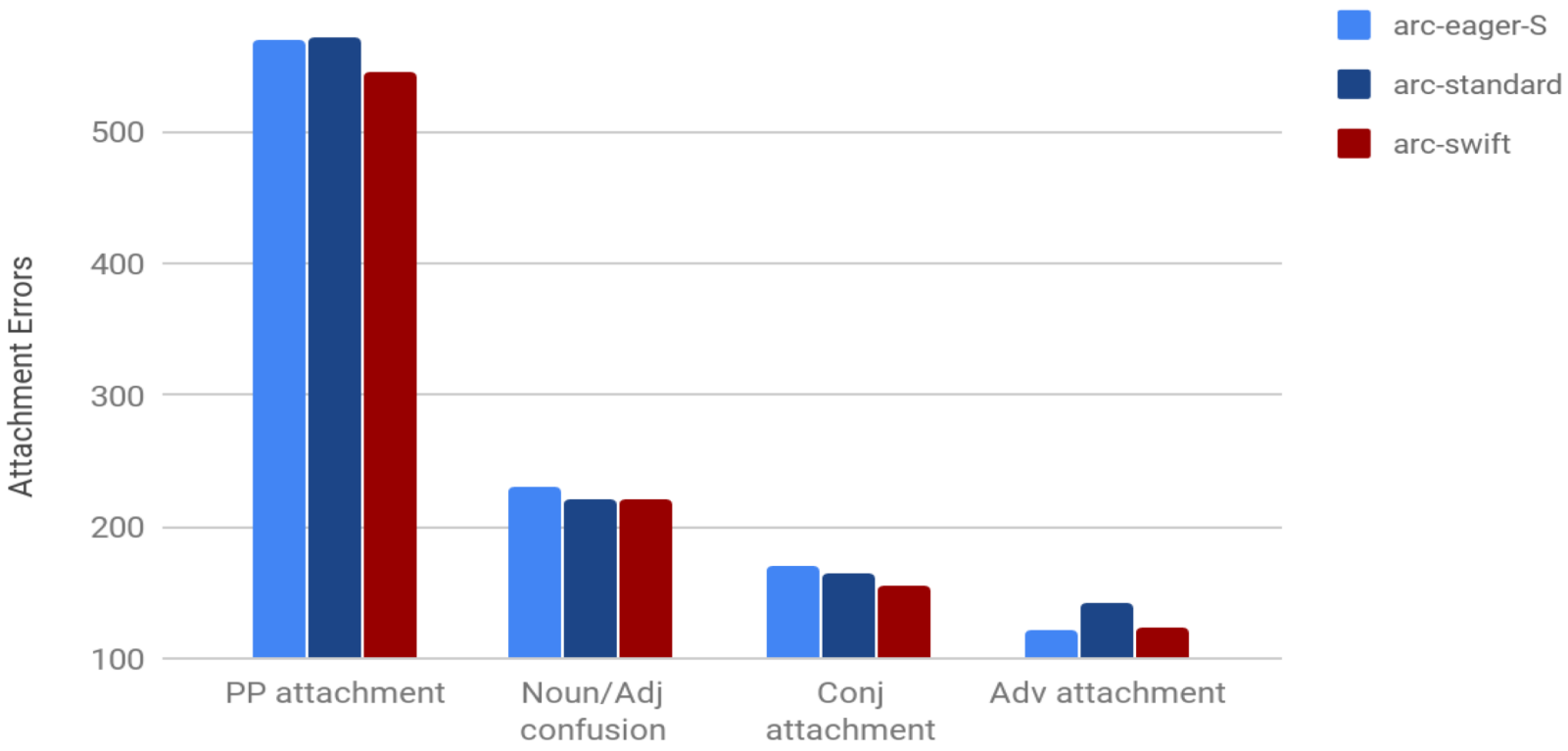
* = p<0.05, ** = p<0.01, *** = p<0.001
In a 10-group Bonferroni-Holm test

Attachment error by dependency length



(Not all baseline transition systems shown)

Error Reduction in Linguistic Categories



Computational Efficiency

	arc-swift	arc-eager beam=2
Number of transitions evaluated per step	124%	400%
Average length of transition sequences	77.5%	100%
UAS	+0.3%	+0.2%
LAS	+0.5%	+0.3%

Thank you!

https://github.com/qipeng/arc-swift

Staying for CoNLL?

Checkout our Shared Task presentation!