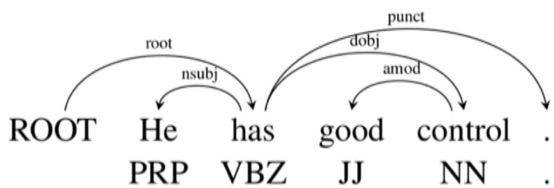
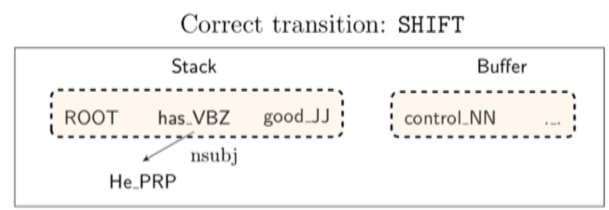
Part 5: Neural Transition-based Methods

Part 5.1: Greedy Parsing

Dependency Parsing

- Neural MaltParser





Transition	Stack	Buffer	A
	[ROOT]	[He has good control .]	$\emptyset$
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-ARC(nsubj)	[ROOT has]	[good control .]	$A\cup$ nsubj(has,He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-ARC(amod)	[ROOT has control]	[.]	$A\cup$amod(control,good)
RIGHT-ARC(dobj)	[ROOT has]	[.]	$A\cup$ dobj(has,control)
...	...	...	...
RIGHT-ARC(root)	[ROOT]	[]	$A\cup$ root(ROOT,has)

Chen, D., & Manning, C. D. (2014). A Fast and Accurate Dependency Parser using Neural Network. ACL.

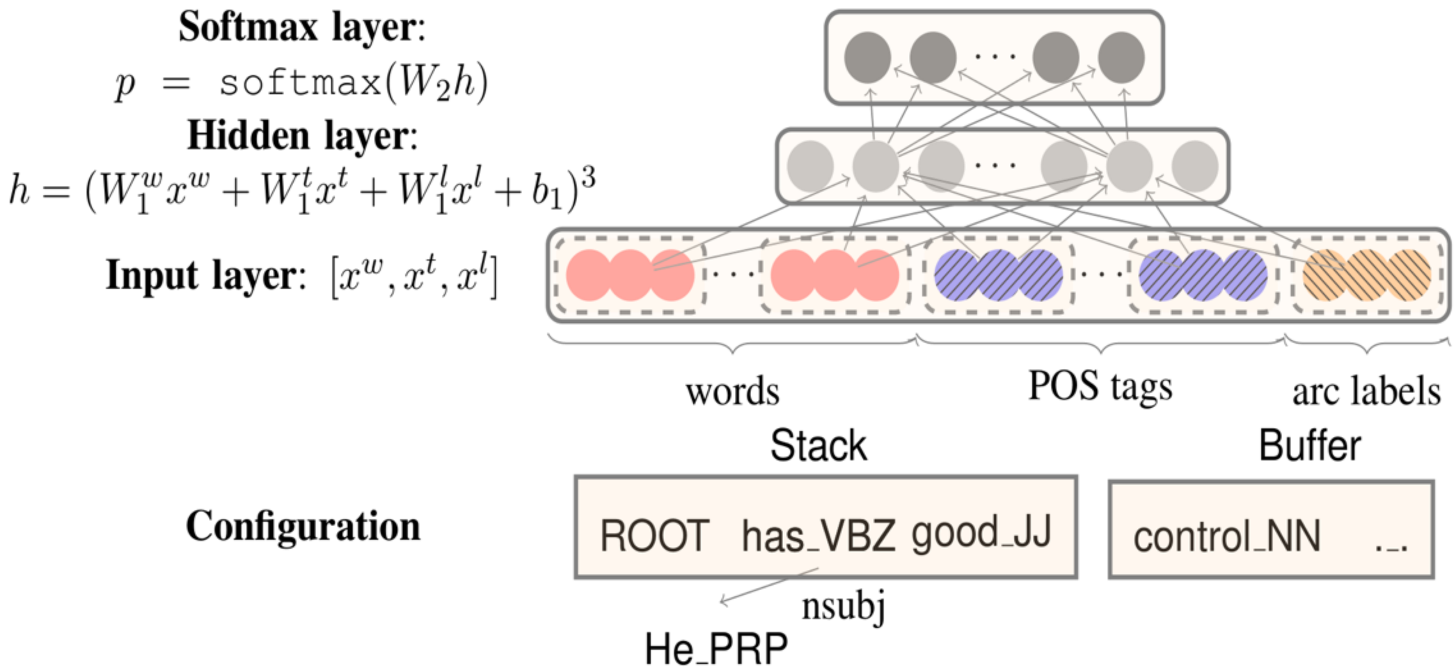
Dependency Parsing

- ZPar 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)	
$s_2.t \circ s_1.t \circ b_1.t$; $s_2.t \circ s_1.t \circ lc_1(s_1).t$;	
$s_2.t \circ s_1.t \circ rc_1(s_1).t$; $s_2.t \circ s_1.t \circ lc_1(s_2).t$;	
$s_2.t \circ s_1.t \circ rc_1(s_2).t$; $s_2.t \circ s_1.w \circ rc_1(s_2).t$;	
$s_2.t \circ s_1.w \circ lc_1(s_1).t$; $s_2.t \circ s_1.w \circ b_1.t$	

Yue Zhang and Joakim Nivre. *Transition-Based Dependency Parsing with Rich Non-Local Features*. In proceedings of ACL 2011, short papers. Portland, USA. June.

Dependency Parsing



Chen, D., & Manning, C. D. (2014). A Fast and Accurate Dependency Parser using Neural Network. ACL.

Dependency Parsing

- Results

Parser	Dev UAS	Dev LAS	Test UAS	Test LAS	Speed (sent/s)
standard	90.2	87.8	89.4	87.3	26
eager	89.8	87.4	89.6	87.4	34
Malt:sp	89.8	87.2	89.3	86.9	469
Malt:eager	89.6	86.9	89.4	86.8	448
MSTParser	91.4	88.1	90.7	87.6	10
Our parser	**92.0**	**89.7**	**91.8**	**89.6**	**654**

PTB (SD)

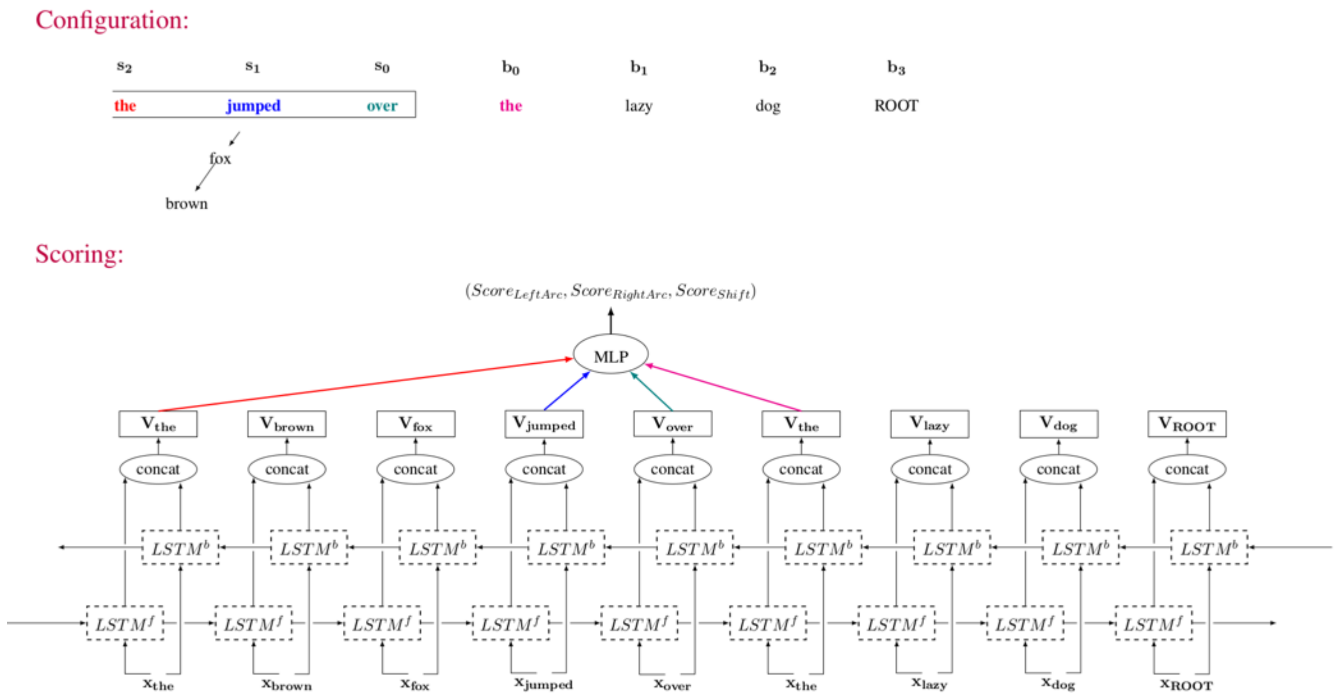
Parser	Dev UAS	Dev LAS	Test UAS	Test LAS	Speed (sent/s)
standard	82.4	80.9	82.7	81.2	72
eager	81.1	79.7	80.3	78.7	80
Malt:sp	82.4	80.5	82.4	80.6	420
Malt:eager	81.2	79.3	80.2	78.4	393
MSTParser	**84.0**	82.1	83.0	81.2	6
Our parser	**84.0**	**82.4**	**83.9**	**82.4**	**936**

CTB (SD)

Chen, D., & Manning, C. D. (2014). A Fast and Accurate Dependency Parser using Neural Network. ACL.

Dependency Parsing

- Chen and Manning with richer features



Keperwasser, E., & Goldberg, Y. (2016). Simple and Accurate Dependency Parsing Using Bidirectional LSTM Feature Representations. TACL.

Dependency Parsing

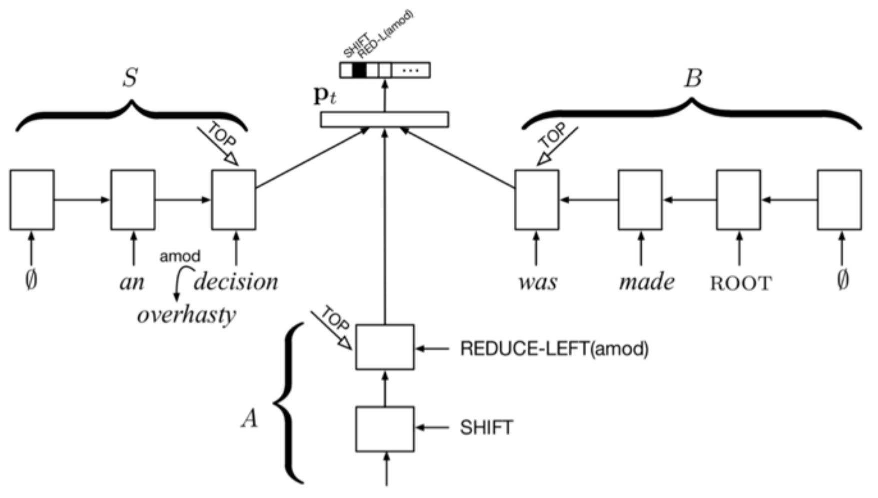
- Results

System	Method	Representation	Emb	PTB-YM	PTB-SD		CTB	
				UAS	UAS	LAS	UAS	LAS
This work	graph, 1st order	2 BiLSTM vectors	–	–	93.1	91.0	**86.6**	**85.1**
This work	transition (greedy, dyn-oracle)	4 BiLSTM vectors	–	–	93.1	91.0	86.2	85.0
This work	transition (greedy, dyn-oracle)	11 BiLSTM vectors	–	–	**93.2**	**91.2**	86.5	84.9
ZhangNivre11	transition (beam)	large feature set (sparse)	–	92.9	–	–	86.0	84.4
Martins13 (TurboParser)	graph, 3rd order+	large feature set (sparse)	–	92.8	93.1	–	–	–
Pei15	graph, 2nd order	large feature set (dense)	–	93.0	–	–	–	–
Dyer15	transition (greedy)	Stack-LSTM + composition	–	–	92.4	90.0	85.7	84.1
Ballesteros16	transition (greedy, dyn-oracle)	Stack-LSTM + composition	–	–	92.7	90.6	86.1	84.5
This work	graph, 1st order	2 BiLSTM vectors	YES	–	93.0	90.9	86.5	84.9
This work	transition (greedy, dyn-oracle)	4 BiLSTM vectors	YES	–	93.6	91.5	87.4	85.9
This work	transition (greedy, dyn-oracle)	11 BiLSTM vectors	YES	–	93.9	91.9	**87.6**	86.1
Weiss15	transition (greedy)	large feature set (dense)	YES	–	93.2	91.2	–	–
Weiss15	transition (beam)	large feature set (dense)	YES	–	**94.0**	**92.0**	–	–
Pei15	graph, 2nd order	large feature set (dense)	YES	93.3	–	–	–	–
Dyer15	transition (greedy)	Stack-LSTM + composition	YES	–	93.1	90.9	87.1	85.5
Ballesteros16	transition (greedy, dyn-oracle)	Stack-LSTM + composition	YES	–	93.6	91.4	**87.6**	**86.2**
LeZuidema14	reranking /blend	inside-outside recursive net	YES	93.1	93.8	91.5	–	–
Zhu15	reranking /blend	recursive conv-net	YES	93.8	–	–	85.7	–

Keperwasser, E., & Goldberg, Y. (2016). Simple and Accurate Dependency Parsing Using Bidirectional LSTM Feature Representations. TACL.

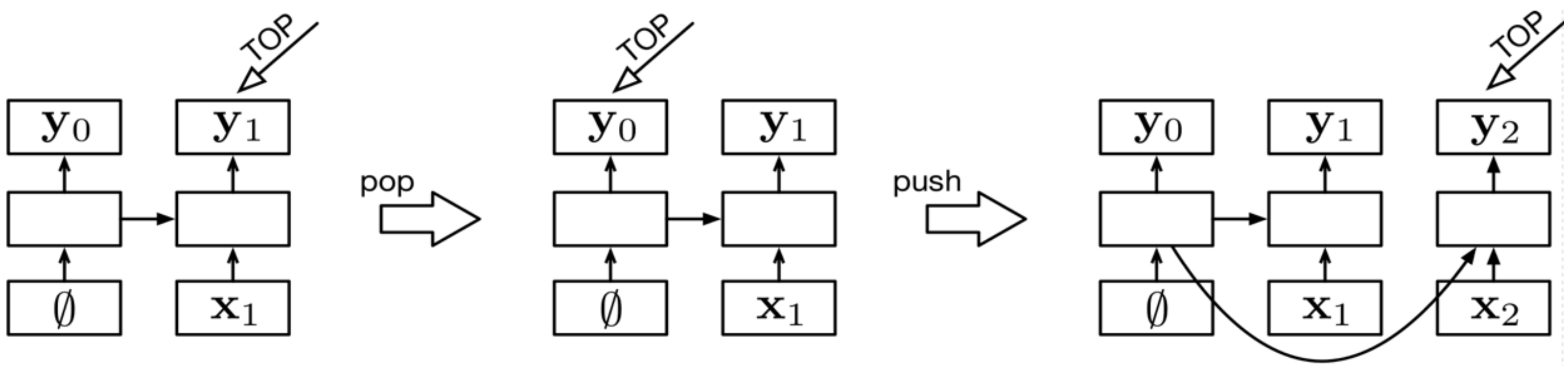
Dependency Parsing

- Chen and Manning with less features



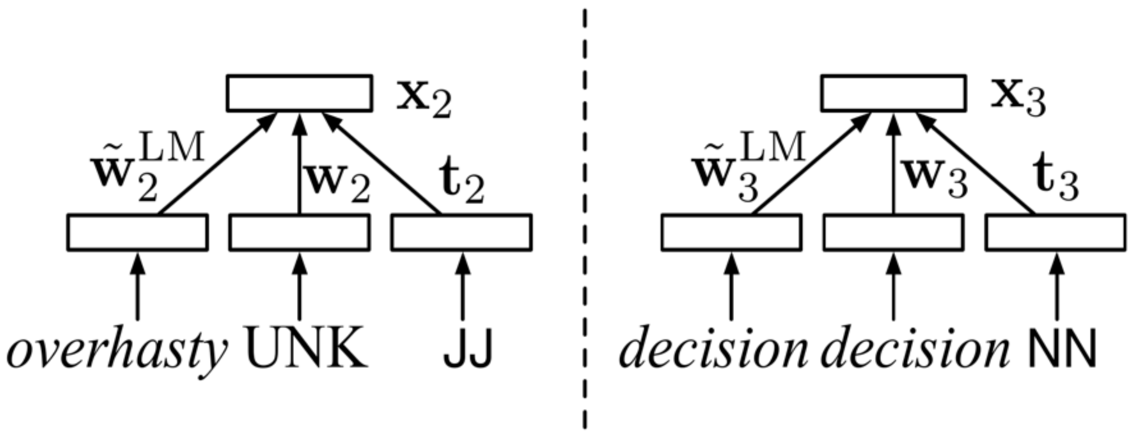
Dyer, C., Ballesteros, M., Ling, W., Matthews, A., & Smith, N. A. (2015). Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ACL.

Dependency Parsing



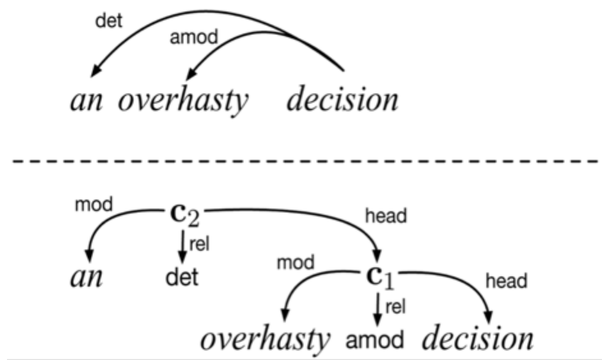
Dyer, C., Ballesteros, M., Ling, W., Matthews, A., & Smith, N. A. (2015). Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ACL.

Dependency Parsing



Dyer, C., Ballesteros, M., Ling, W., Matthews, A., & Smith, N. A. (2015). Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ACL.

Dependency Parsing



Dyer, C., Ballesteros, M., Ling, W., Matthews, A., & Smith, N. A. (2015). Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ACL.

Dependency Parsing

- Results

	Development		Test	
	UAS	LAS	UAS	LAS
S-LSTM	**93.2**	**90.9**	**93.1**	**90.9**
−POS	93.1	90.4	92.7	90.3
−pretraining	92.7	90.4	92.4	90.0
−composition	92.7	89.9	92.2	89.6
S-RNN	92.8	90.4	92.3	90.1
C&M (2014)	92.2	89.7	91.8	89.6

PTB (SD)

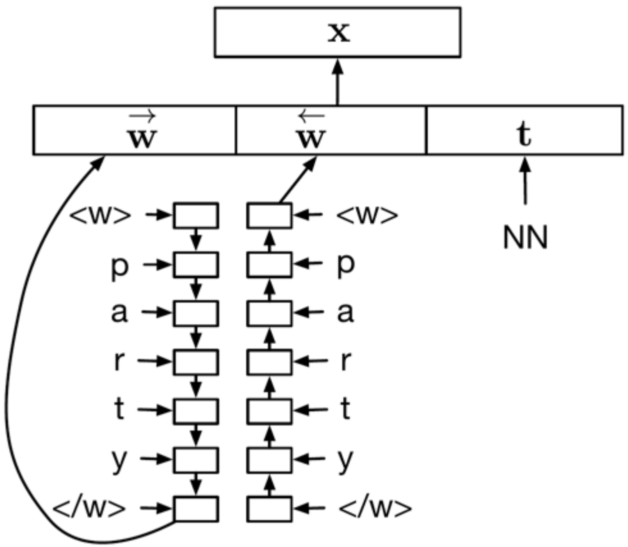
	Development		Test	
	UAS	LAS	UAS	LAS
S-LSTM	**87.2**	**85.9**	**87.2**	**85.7**
−POS	82.8	79.8	82.2	79.1
−pretraining	86.3	84.7	85.7	84.1
−composition	85.8	84.0	85.3	83.6
S-RNN	86.3	84.7	86.1	84.6
C&M (2014)	84.0	82.4	83.9	82.4

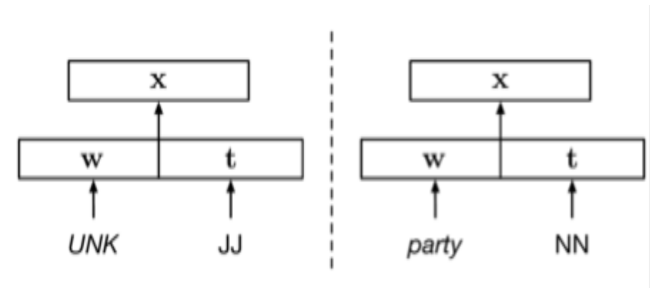
CTB (CTB5)

Dyer, C., Ballesteros, M., Ling, W., Matthews, A., & Smith, N. A. (2015). Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ACL.

Dependency Parsing

- Dyer et al. with character based word vector





Ballesteros, M., Dyer, C., & Smith, N. A. (2015). Improved Transition-Based Parsing by Modeling Characters instead of Words with LSTMs. EMNLP.

Dependency Parsing

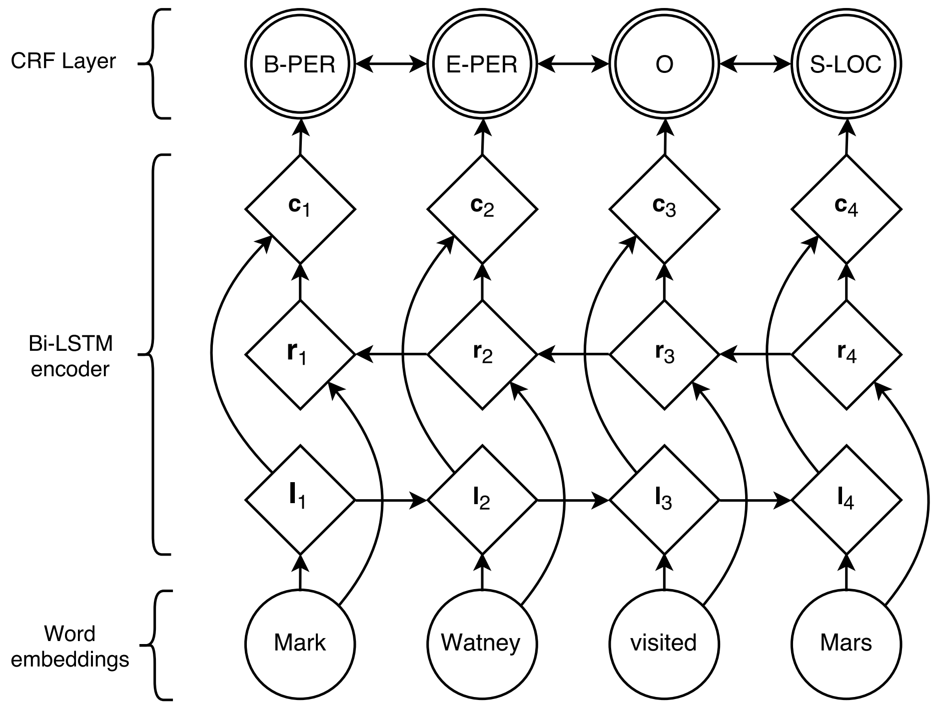
- Results

	This Work			Best Greedy Result			Best Published Result		
Language	**UAS**	**LAS**	**System**	**UAS**	**LAS**	**System**	**UAS**	**LAS**	**System**
Arabic	86.08	83.41	**Chars**	84.57	81.90	B'13	88.32	86.21	B+'13
Basque	85.22	78.61	**Chars + POS**	84.33	78.58	B'13	89.96	85.70	B+'14
French	86.15	82.03	**Words + POS**	83.35	77.98	B'13	89.02	85.66	B+'14
German	87.33	84.62	**Words + POS**	85.38	82.75	B'13	91.64	89.65	B+'13
Hebrew	80.68	72.70	**Words + POS**	79.89	73.01	B'13	87.41	81.65	B+'14
Hungarian	80.92	76.34	**Chars + POS**	83.71	79.63	B'13	89.81	86.13	B+'13
Korean	88.39	86.27	**Chars**	85.72	82.06	B'13	89.10	87.27	B+'14
Polish	87.06	79.83	**Words + POS**	85.80	79.89	B'13	91.75	87.07	B+'13
Swedish	83.43	76.40	**Words + POS**	83.20	75.82	B'13	88.48	82.75	B+'14
Turkish	76.32	64.34	**Chars**	75.82	65.68	N+'06a	77.55	n/a	K+'10
Chinese	85.96	84.40	**Words + POS**	87.20	85.70	D+'15	87.20	85.70	D+'15
English	92.57	90.31	**Words + POS**	93.10	90.90	D+'15	94.08	92.19	W+'15

Ballesteros, M., Dyer, C., & Smith, N. A. (2015). Improved Transition-Based Parsing by Modeling Characters instead of Words with LSTMs. EMNLP.

Named Entity Recognition

- Model



Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, Chris Dyer, Neural Architectures for Named Entity Recognition, In Proceedings of NAACL 2016

Named Entity Recognition

- Model
  - Transitions

$\mathbf{Out}_t$	$\mathbf{Stack}_t$	$\mathbf{Buffer}_t$	**Action**	$\mathbf{Out}_{t+1}$	$\mathbf{Stack}_{t+1}$	$\mathbf{Buffer}_{t+1}$	**Segments**
O	S	$(\mathbf{u}, u), B$	SHIFT	O	$(\mathbf{u}, u), S$	B	—
O	$(\mathbf{u}, u), \ldots, (\mathbf{v}, v), S$	B	REDUCE(y)	$g(\mathbf{u}, \ldots, \mathbf{v}, \mathbf{r}_y), O$	S	B	$(u \ldots v, y)$
O	S	$(\mathbf{u}, u), B$	OUT	$g(\mathbf{u}, \mathbf{r}_\varnothing), O$	S	B	—

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, Chris Dyer, Neural Architectures for Named Entity Recognition, In Proceedings of NAACL 2016

Named Entity Recognition

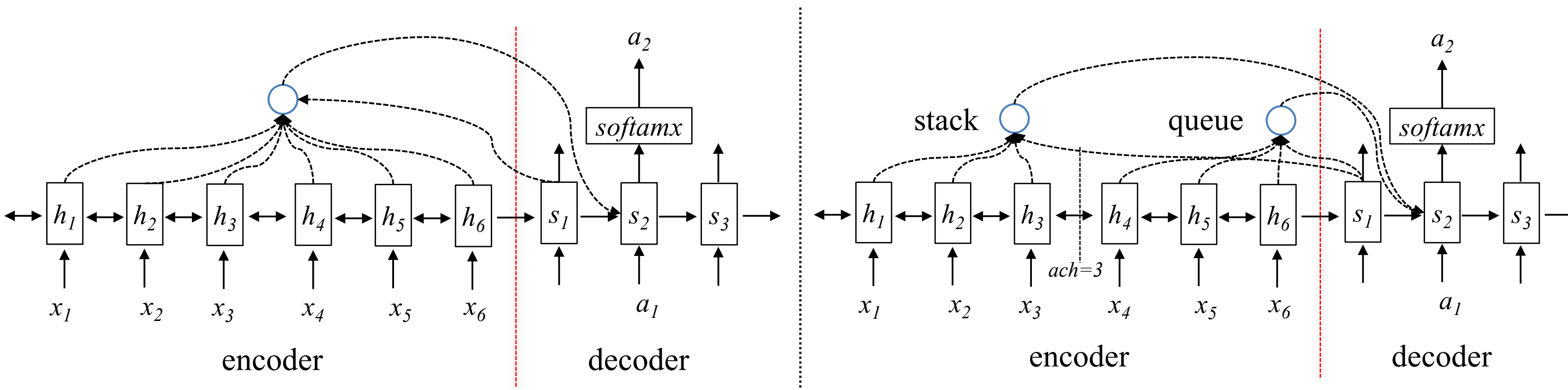
- Results
  - English NER results

Model	F_1
Collobert et al. (2011)*	89.59
Lin and Wu (2009)	83.78
Lin and Wu (2009)*	90.90
Huang et al. (2015)*	90.10
Passos et al. (2014)	90.05
Passos et al. (2014)*	90.90
Luo et al. (2015)* + gaz	89.9
Luo et al. (2015)* + gaz + linking	**91.2**
Chiu and Nichols (2015)	90.69
Chiu and Nichols (2015)*	90.77
LSTM-CRF (no char)	90.20
LSTM-CRF	**90.94**
S-LSTM (no char)	87.96
S-LSTM	90.33

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, Chris Dyer, Neural Architectures for Named Entity Recognition, In Proceedings of NAACL 2016

Dependency and Constituent Parsing

- Decoder



1. Vanilla Decoder

2. Stack-queue Decoder

Jiangming Liu, Yue Zhang, Encoder-Decoder Shift-Reduce Syntactic Parsing. IWPT 2017: 105-114

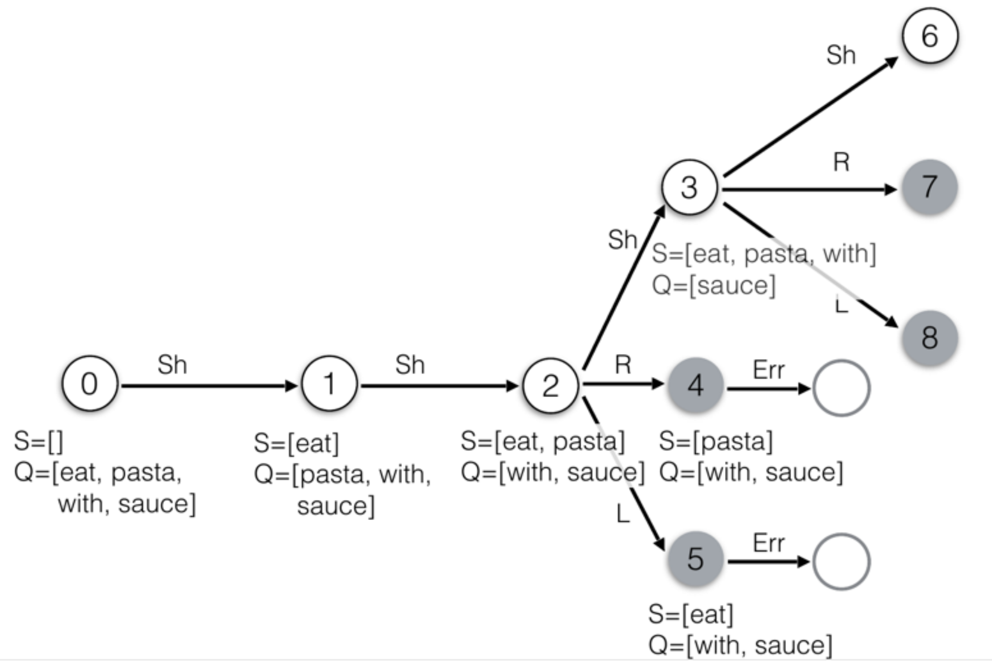
Dependency and Constituent Parsing

- Results

Model	UAS (%)
Dyer et al. (2015)	92.3
Vanilla decoder	88.5
SQ decoder + average pooling	91.9
SQ decoder + attention	92.4
SQ decoder + treeLSTM	92.4

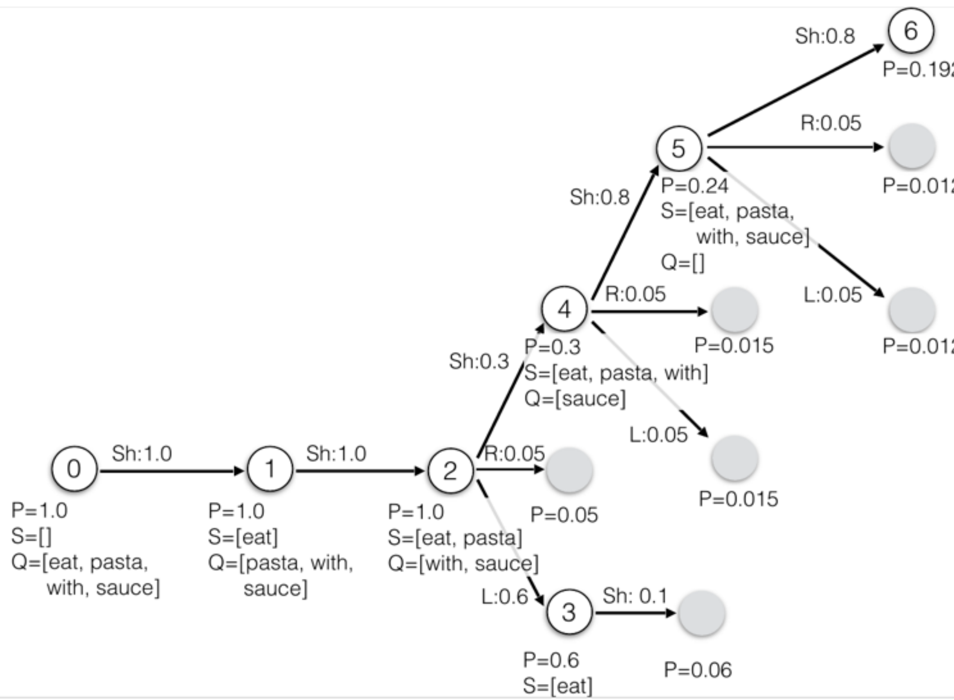
Jiangming Liu, Yue Zhang, Encoder-Decoder Shift-Reduce Syntactic Parsing. IWPT 2017: 105-114

Dependency Parser



Vaswani, A., & Sagae, K. (2016). Efficient Structured Inference for Transition-Based Parsing with Neural Networks and Error States. NAACL.

Dependency Parser



Vaswani, A., & Sagae, K. (2016). Efficient Structured Inference for Transition-Based Parsing with Neural Networks and Error States. NAACL.

Dependency Parser

System	wsj23-S	wsj23-YM
ErrSt–25–rand	**92.17**	**92.16**
ErrSt–25–pre*	**93.61**	**93.21**
Chen & Manning*	91.8	–
Huang & Sagae	–	92.1
Zhang & Nivre	93.5	92.9
Weiss et al.*	93.99	–
Zhang & McDonald	93.71	93.57
Martins et al.	92.82	93.07
Koo et al. (dep2c)*	–	93.16

Vaswani, A., & Sagae, K. (2016). Efficient Structured Inference for Transition-Based Parsing with Neural Networks and Error States. NAACL.

Part 5.2: Dependency Parsing with Beam Search

Dependency Parsing

- Zhang & Nivre (2011)

$$y = \arg\max_{y' \in \mathrm{GEN}(x)} score(y')$$

$$score(y) = \sum_{a \in y} \theta \cdot \Phi(a)$$

Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

Dependency Parsing

- Chen and Manning (2014)

$$h = (W_1 x + b_1)^3$$

$$p = softmax(o)$$

$$o = W_2 h$$

Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

Dependency Parsing

- What does not work

$$s(y) = \sum_{a \in y} \log p_a$$

$$L(\theta) = max(0, \delta - s(y_g) + s(y_p)) + \frac{\lambda}{2} \parallel \theta \parallel^2$$

Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

Dependency Parsing

- Sentence-level log likelihood

$$p(y_i \mid x, \theta) = \frac{e^{f(x,\, \theta)_i}}{\sum_{y_j \in \mathrm{GEN}(x)} e^{f(x,\, \theta)_j}}$$

$$f(x,\ \theta)_i = \sum_{a_k \in y_i} o(x,\ y_i,\ k,\ a_k)$$

Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

Dependency Parsing

- Contrastive Estimation

$$L(\theta) = -\sum_{(x_i, y_i)\in(X,Y)} \log p(y_i \mid x_i, \theta)$$

$$= -\sum_{(x_i, y_i)\in(X,Y)} \log \frac{e^{f(x_i,\theta)_i}}{Z(x_i,\theta)}$$

$$= \sum_{(x_i, y_i)\in(X,Y)} \log Z(x_i,\theta) - f(x_i,\theta)_i$$

$$Z(x,\theta) = \sum_{y_j\in\mathrm{GEN}(x)} e^{f(x,\,\theta)_j}$$

Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

Dependency Parsing

- Contrastive Estimation

$$L'(\theta) = -\sum_{(x_i, y_i) \in (X,Y)} \log p'(y_i \mid x_i, \theta)$$

$$= -\sum_{(x_i, y_i) \in (X,Y)} \log \frac{e^{f(x_i,\theta)_i}}{Z'(x_i,\theta)}$$

$$= \sum_{(x_i, y_i) \in (X,Y)} \log Z'(x_i,\theta) - f(x_i,\theta)_i$$

$$Z'(x,\theta) = \sum_{y_j \in \mathrm{BEAM}(x)} e^{f(x,\,\theta)_j}$$

Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

Dependency Parsing

- Results

Description	UAS	
Baseline	91.63	
	structured	greedy
beam = 1	74.90	91.63
beam = 4	84.64	91.92
beam = 16	91.53	91.90
beam = 64	93.12	91.84
beam = 100	93.23	91.81

Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

Dependency Parsing

- Results

Description	UAS
greedy neural parser	91.47
ranking model	89.08
beam contrastive learning	93.28

Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

Dependency Parsing

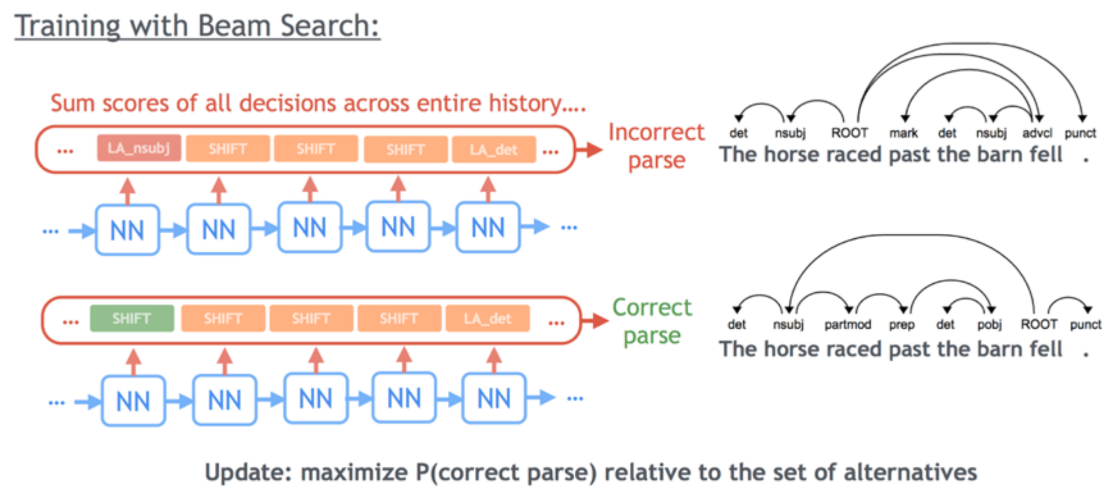
- Results

System		UAS	LAS	Speed
baseline greedy parser		91.47	90.43	0.001
Huang and Sagae (2010)		92.10		0.04
Zhang and Nivre (2011)		92.90	91.80	0.03
Choi and McCallum (2013)		92.96	91.93	0.009
Ma et al. (2014)		93.06		
Bohnet and Nivre (2012)†‡		93.67	92.68	0.4
Suzuki et al. (2009)†		93.79		
Koo et al. (2008)†		93.16		
Chen et al. (2014)†		93.77		
beam size				
training	decoding			
100	100	**93.28**	**92.35**	0.07
100	64	93.20	92.27	0.04
100	16	92.40	91.95	0.01

Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

Google’s SyntaxNet

- Andor et al. follows this method
  - Offers theorem
  - More tasks
  - Better results



Globally Normalized SyntaxNet Architecture (Overview)

Andor, D., Alberti, Chris., Weiss, D., Severyn, A., Presta, A., Ganchev, K., Petrov, S., & Collins, M. (2016). Globally Normalized Transition-Based Neural Networks. ACL.

Google's SyntaxNet

- English Results

Method	WSJ UAS	WSJ LAS	Union-News UAS	Union-News LAS	Union-Web UAS	Union-Web LAS	Union-QTB UAS	Union-QTB LAS
Martins et al. (2013)*	92.89	90.55	93.10	91.13	88.23	85.04	94.21	91.54
Zhang and McDonald (2014)*	93.22	91.02	93.32	91.48	88.65	85.59	93.37	90.69
Weiss et al. (2015)	93.99	92.05	93.91	92.25	89.29	86.44	94.17	92.06
Alberti et al. (2015)	94.23	92.36	94.10	92.55	89.55	86.85	94.74	93.04
Our Local (B=1)	92.95	91.02	93.11	91.46	88.42	85.58	92.49	90.38
Our Local (B=32)	93.59	91.70	93.65	92.03	88.96	86.17	93.22	91.17
Our Global (B=32)	**94.61**	**92.79**	**94.44**	**92.93**	**90.17**	**87.54**	**95.40**	**93.64**
Parsey McParseface (B=8)	-	-	94.15	92.51	89.08	86.29	94.77	93.17

Andor, D., Alberti, Chris., Weiss, D., Severyn, A., Presta, A., Ganchev, K., Petrov, S., & Collins, M. (2016). Globally Normalized Transition-Based Neural Networks. ACL.

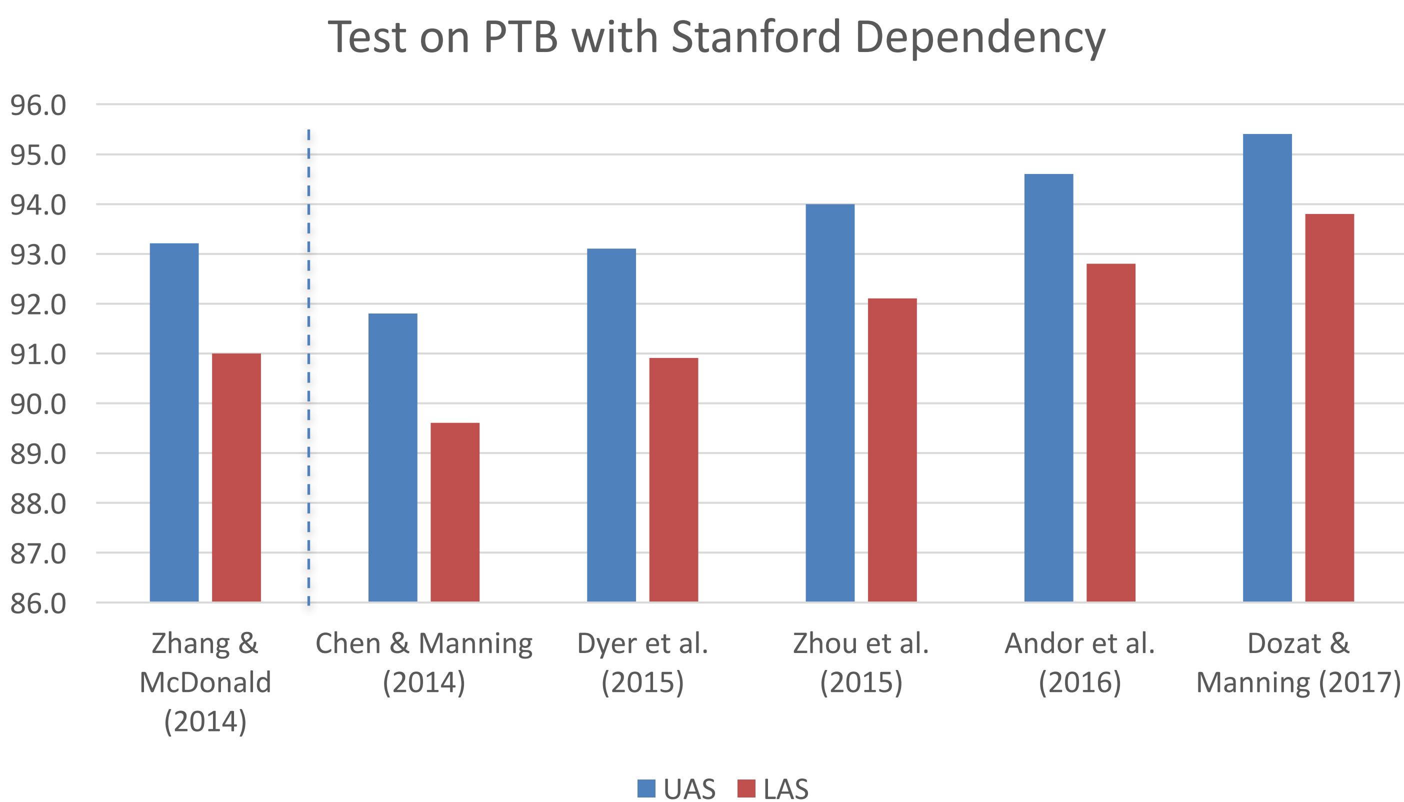
Google's SyntaxNet

- Results for other languages

Method	Catalan UAS	Catalan LAS	Chinese UAS	Chinese LAS	Czech UAS	Czech LAS	English UAS	English LAS	German UAS	German LAS	Japanese UAS	Japanese LAS	Spanish UAS	Spanish LAS
Best Shared Task Result	-	87.86	-	79.17	-	80.38	-	89.88	-	87.48	-	92.57	-	87.64
Ballesteros et al. (2015)	90.22	86.42	80.64	76.52	79.87	73.62	90.56	88.01	88.83	86.10	93.47	92.55	90.38	86.59
Zhang and McDonald (2014)	91.41	87.91	82.87	78.57	86.62	80.59	92.69	90.01	89.88	87.38	92.82	91.87	90.82	87.34
Lei et al. (2014)	91.33	87.22	81.67	76.71	88.76	81.77	92.75	90.00	90.81	87.81	**94.04**	91.84	91.16	87.38
Bohnet and Nivre (2012)	92.44	89.60	82.52	78.51	88.82	83.73	92.87	90.60	**91.37**	**89.38**	93.67	92.63	92.24	89.60
Alberti et al. (2015)	92.31	89.17	83.57	79.90	88.45	83.57	92.70	90.56	90.58	88.20	93.99	**93.10**	92.26	89.33
Our Local (B=1)	91.24	88.21	81.29	77.29	85.78	80.63	91.44	89.29	89.12	86.95	93.71	92.85	91.01	88.14
Our Local (B=16)	91.91	88.93	82.22	78.26	86.25	81.28	92.16	90.05	89.53	87.4	93.61	92.74	91.64	88.88
Our Global (B=16)	**92.67**	**89.83**	**84.72**	**80.85**	**88.94**	**84.56**	**93.22**	**91.23**	90.91	89.15	93.65	92.84	**92.62**	**89.95**

Andor, D., Alberti, Chris., Weiss, D., Severyn, A., Presta, A., Ganchev, K., Petrov, S., & Collins, M. (2016). Globally Normalized Transition-Based Neural Networks. ACL.

Changes of Performance



Part 5.3: Other tasks with beam-search

Word segmentation

step	action	buffer($\cdots w_{-1} w_0$)	queue($c_0 c_1 \cdots$)
0	-	ϕ	中 国 $\cdots$
1	*SEP*	中	国 外 $\cdots$
2	*APP*	中国	外 企 $\cdots$
3	*SEP*	中国 外	企 业 $\cdots$
4	*APP*	中国 外企	业 务 $\cdots$
5	*SEP*	中国 外企 业	务 发 $\cdots$
6	*APP*	中国 外企 业务	发 展 $\cdots$
7	*SEP*	$\cdots$ 业务 发	展 迅 速
8	*APP*	$\cdots$ 业务 发展	迅 速
9	*SEP*	$\cdots$ 发展 迅	速
10	*APP*	$\cdots$ 发展 迅速	ϕ

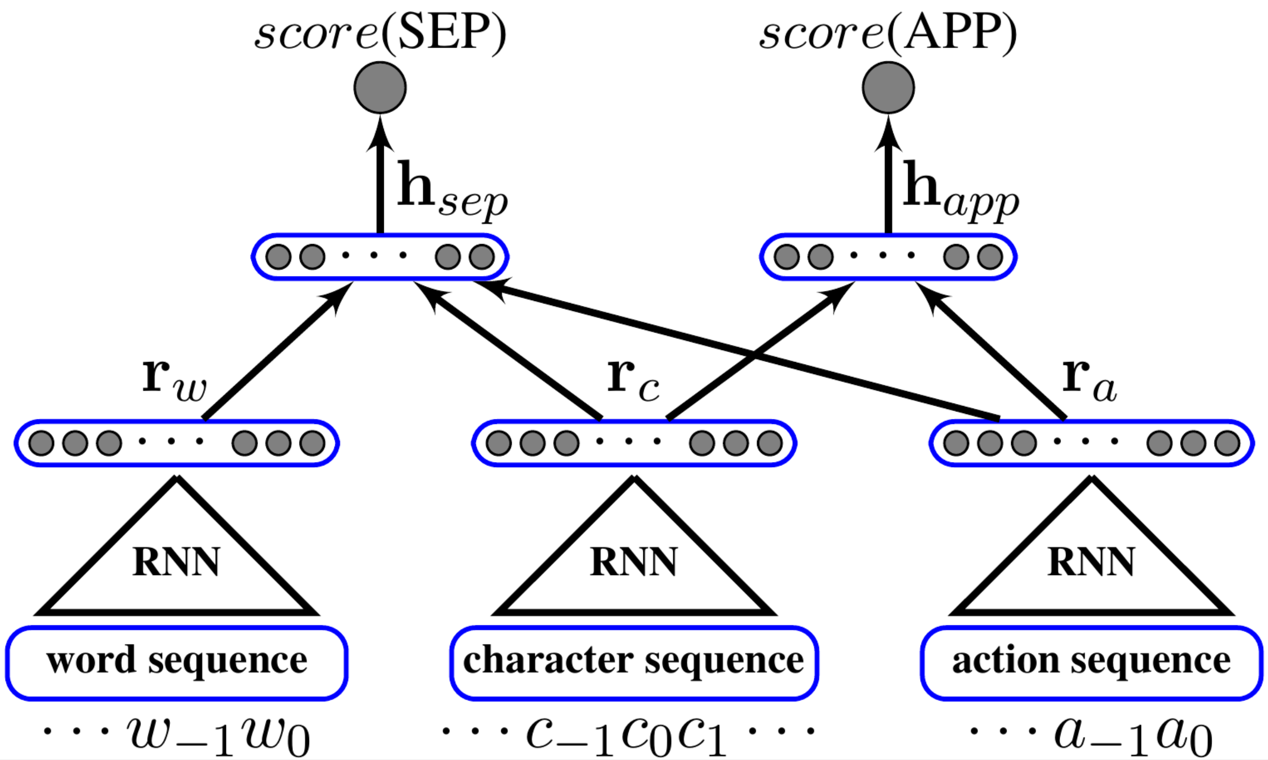
Zhang, M., Zhang, Y., & Fu, G. (2016). Transition-Based Neural Word Segmentation. ACL.

Word segmentation

Feature templates	**Action**
$c_{-1}c_0$	*APP, SEP*
$w_{-1}, w_{-1}w_{-2}, w_{-1}c_0, w_{-2}len(w_{-1})$ $start(w_{-1})c_0, end(w_{-1})c_0$ $start(w_{-1})end(w_{-1}), end(w_{-2})end(w_{-1})$ $w_{-2}len(w_{-1}), len(w_{-2})w_{-1}$ w_{-1}, where $len(w_{-1}) = 1$	*SEP*

Zhang, M., Zhang, Y., & Fu, G. (2016). Transition-Based Neural Word Segmentation. ACL.

Word segmentation



Zhang, M., Zhang, Y., & Fu, G. (2016). Transition-Based Neural Word Segmentation. ACL.

Word segmentation

Models	P	R	F
word-based models			
discrete	95.29	95.26	95.28
neural	95.34	94.69	95.01
combined	**96.11**	**95.79**	**95.95**
character-based models			
discrete	95.38	95.12	95.25
neural	94.59	94.92	94.76
combined	95.63	95.60	95.61
other models			
Zhang et al. (2014)	N/A	N/A	95.71
Wang et al. (2011)	95.83	95.75	95.79
Zhang and Clark (2011)	95.46	94.78	95.13

Main results on CTB60 test dataset

Zhang, M., Zhang, Y., & Fu, G. (2016). Transition-Based Neural Word Segmentation. ACL.

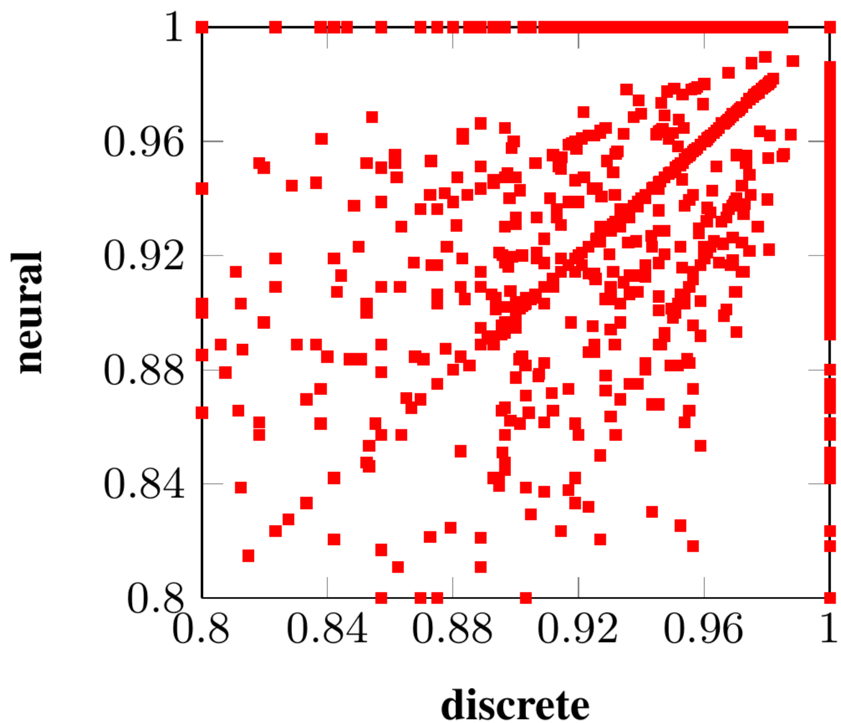
Word segmentation

Models	PKU	MSR
our word-based models		
discrete	95.1	97.3
neural	95.1	97.0
combined	95.7	**97.7**
character-based models		
discrete	94.9	96.8
neural	94.4	97.2
combined	95.4	97.2
other models		
Cai and Zhao (2016)	95.5	96.5
Ma and Hinrichs (2015)	95.1	96.6
Pei et al. (2014)	95.2	97.2
Zhang et al. (2013a)	**96.1**	97.5
Sun et al. (2012)	95.4	97.4
Zhang and Clark (2011)	95.1	97.1
Sun (2010)	95.2	96.9
Sun et al. (2009)	95.2	97.3

Main results on PKU and MSR test dataset

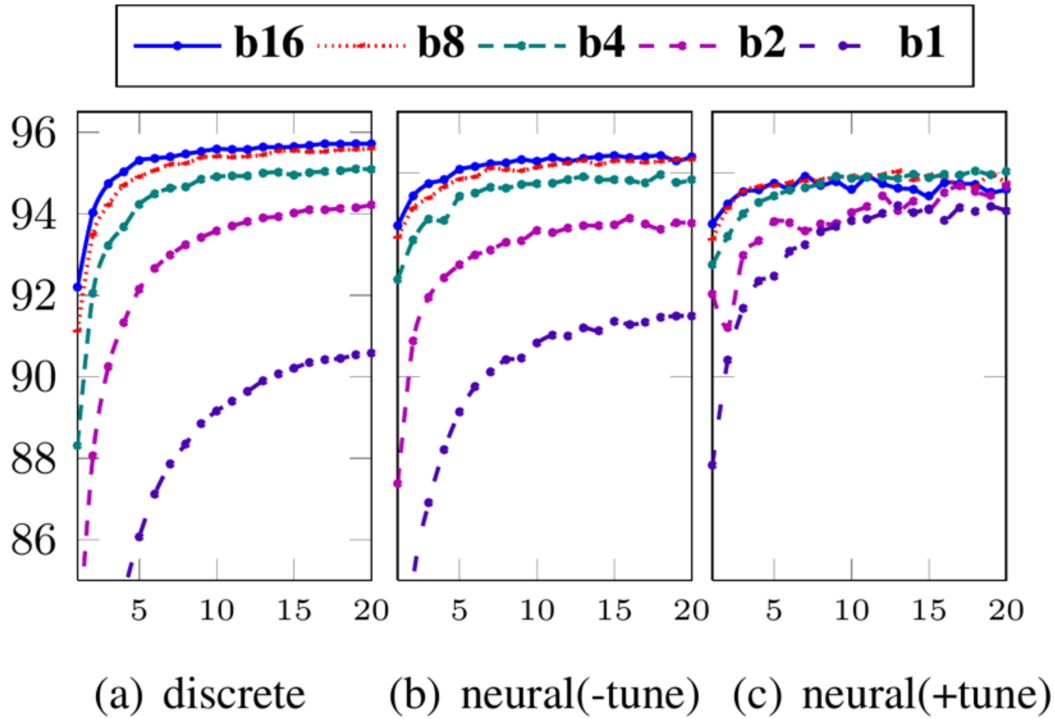
Zhang, M., Zhang, Y., & Fu, G. (2016). Transition-Based Neural Word Segmentation. ACL.

Word segmentation



Zhang, M., Zhang, Y., & Fu, G. (2016). Transition-Based Neural Word Segmentation. ACL.

Word segmentation



Zhang, M., Zhang, Y., & Fu, G. (2016). Transition-Based Neural Word Segmentation. ACL.

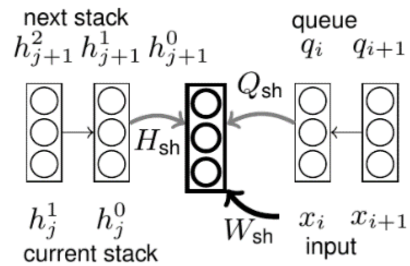
Word segmentation

- Other Methods
  - Cai and Zhao (2016)
  - Yang et al. (2017)

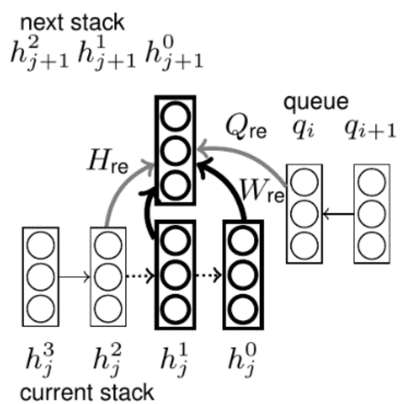
Cai, D., & Zhao, H. (2016). Neural Word Segmentation Learning for Chinese. ACL.
Jie Yang, Yue Zhang, Fei Dong. *Neural Word Segmentation with Rich Pretraining* (ACL). Vancouver, Canada, July.

Constituent parsing

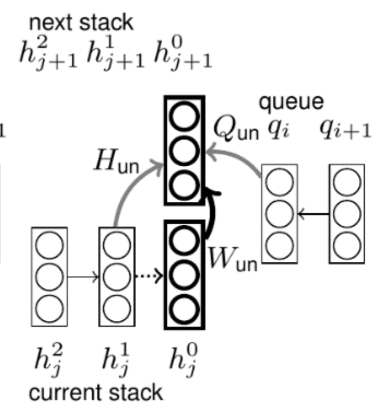
- Model



(a) shift-X action



(b) reduce-X action



(c) unary-X action

Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing

- Update at max-violation

$$j^* = \arg\min_j \left\{ \rho_{\boldsymbol{\theta}}(y_0^j) - \max_{\boldsymbol{d} \in B_j} \rho_{\boldsymbol{\theta}}(\boldsymbol{d}) \right\}$$

- Using expected loss from all violations

$$L(\boldsymbol{w}, \boldsymbol{y}; \boldsymbol{B}, \boldsymbol{\theta}) = \max \left\{ 0, 1 - \rho_{\boldsymbol{\theta}}(y_0^{j^*}) + \mathbb{E}_{\tilde{B}_{j^*}}[\rho_{\boldsymbol{\theta}}] \right\}$$

$$\tilde{B}_{j^*} = \left\{ \boldsymbol{d} \in B_{j^*} \middle| \rho_{\boldsymbol{\theta}}(\boldsymbol{d}) > \rho_{\boldsymbol{\theta}}(y_0^{j^*}) \right\}$$

$$p_{\boldsymbol{\theta}}(\boldsymbol{d}) = \frac{\exp(\rho_{\boldsymbol{\theta}}(\boldsymbol{d}))}{\sum_{\boldsymbol{d}' \in \tilde{B}_{j^*}} \exp(\rho_{\boldsymbol{\theta}}(\boldsymbol{d}'))}$$

$$\mathbb{E}_{\tilde{B}_{j^*}}[\rho_{\boldsymbol{\theta}}] = \sum_{\boldsymbol{d} \in \tilde{B}_{j^*}} p_{\boldsymbol{\theta}}(\boldsymbol{d}) \rho_{\boldsymbol{\theta}}(\boldsymbol{d}).$$

Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing

- English Results

parser	test
Collins (Collins, 1997)	87.8
Berkeley (Petrov and Klein, 2007)	90.1
SSN (Henderson, 2004)	90.1
ZPar (Zhu et al., 2013)	90.4
CVG (Socher et al., 2013)	90.4
Charniak-R (Charniak and Johnson, 2005)	**91.0**
This work: TNCP	90.7

Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing

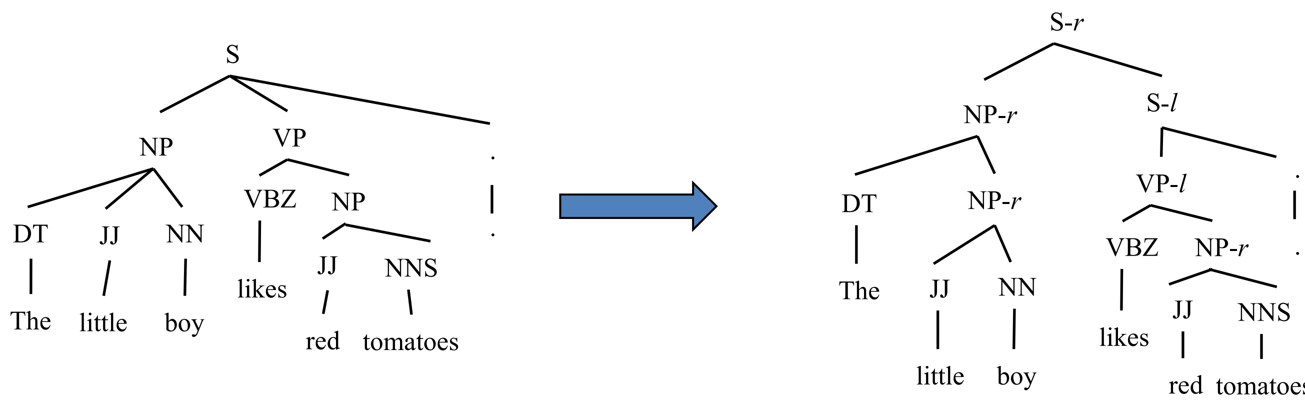
- Chinese Results

parser	test
ZPar (Zhu et al., 2013)	83.2
Berkeley (Petrov and Klein, 2007)	83.3
Joint (Wang and Xue, 2014)	**84.9**
This work: TNCP	84.3

Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

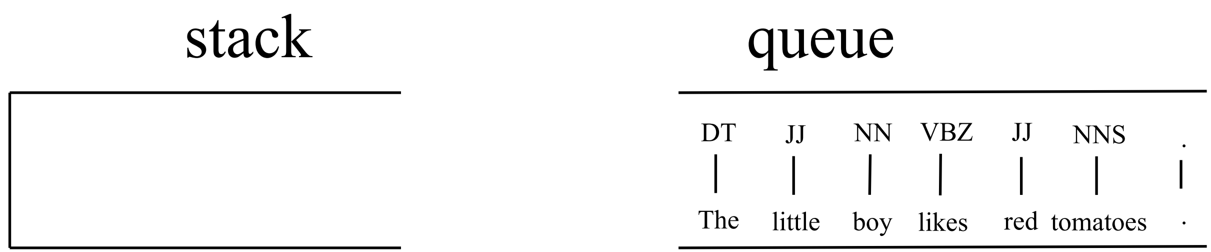
- Binarization



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

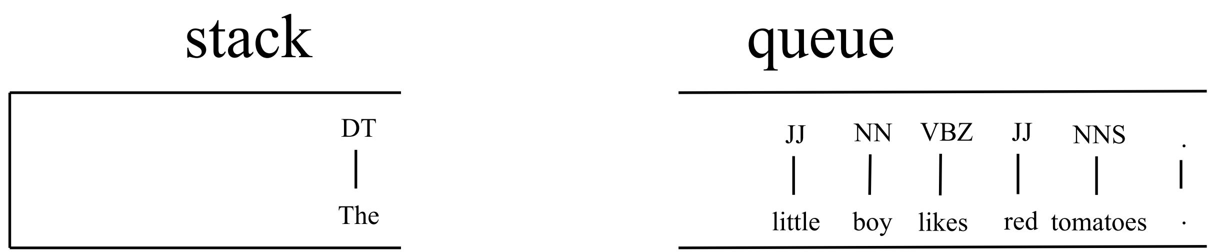
- Actions
  - Shift



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

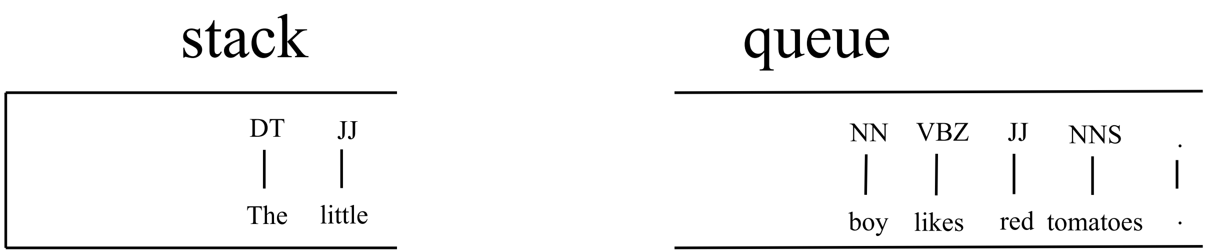
- Actions
  - Shift



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

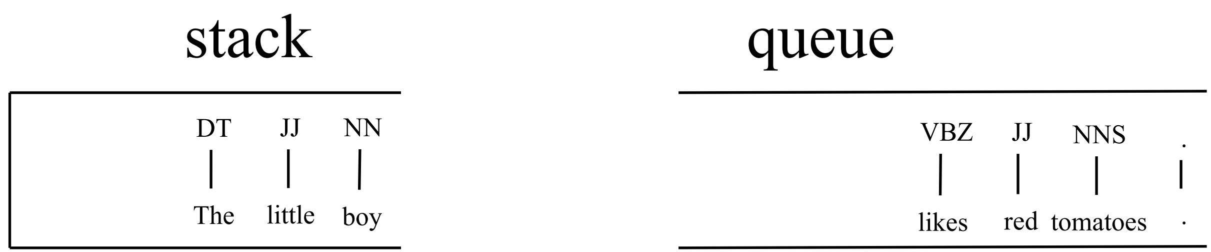
- Actions
  - Shift



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

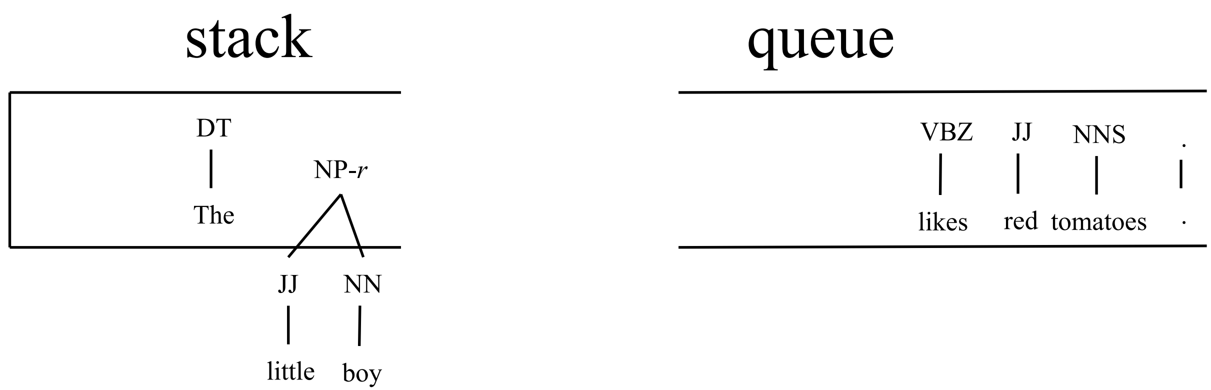
- Actions
  - Reduce-r-NP



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

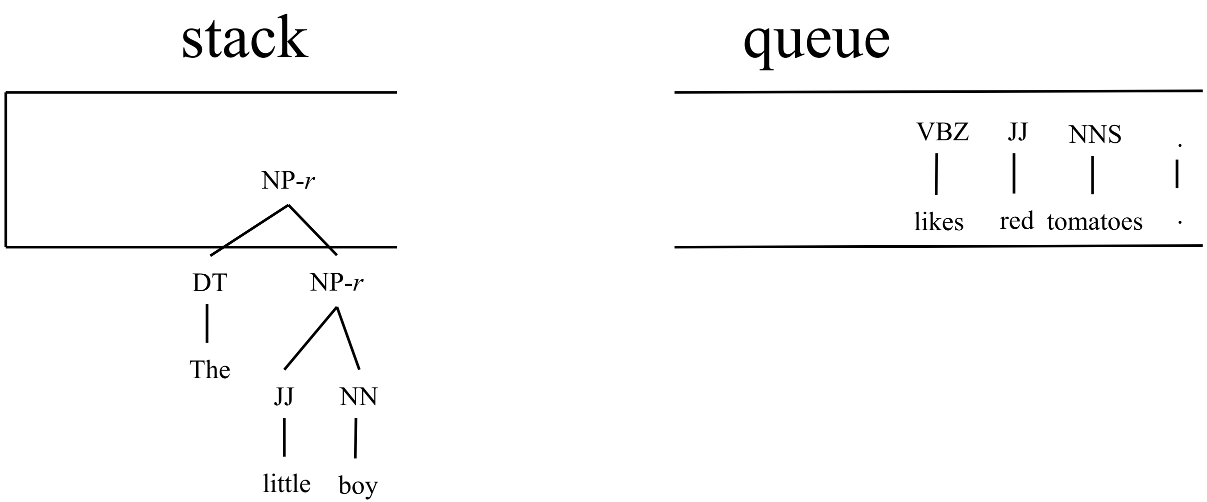
- Actions
  - Reduce-r-NP



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

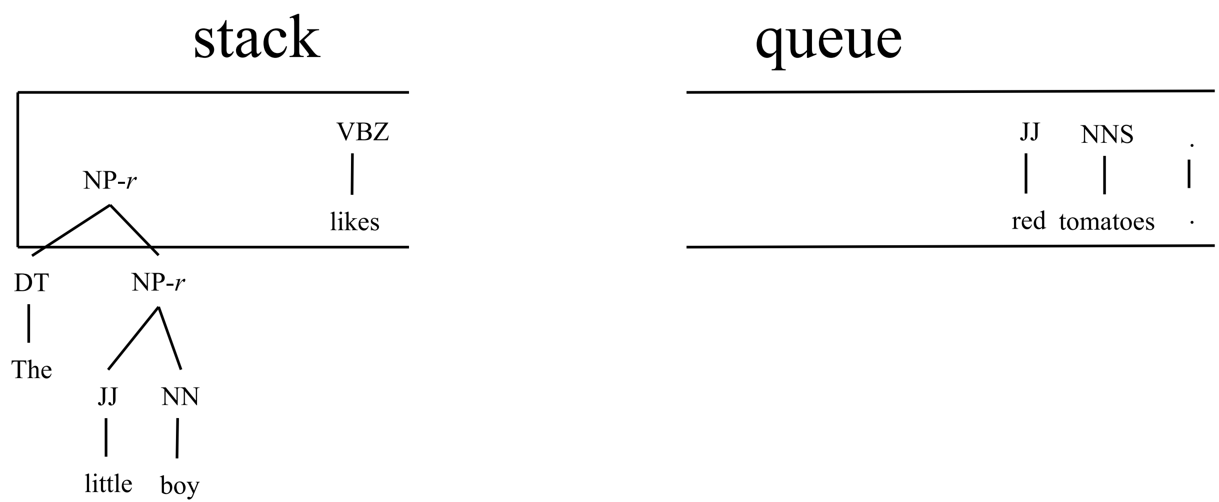
- Actions
  - Shift



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

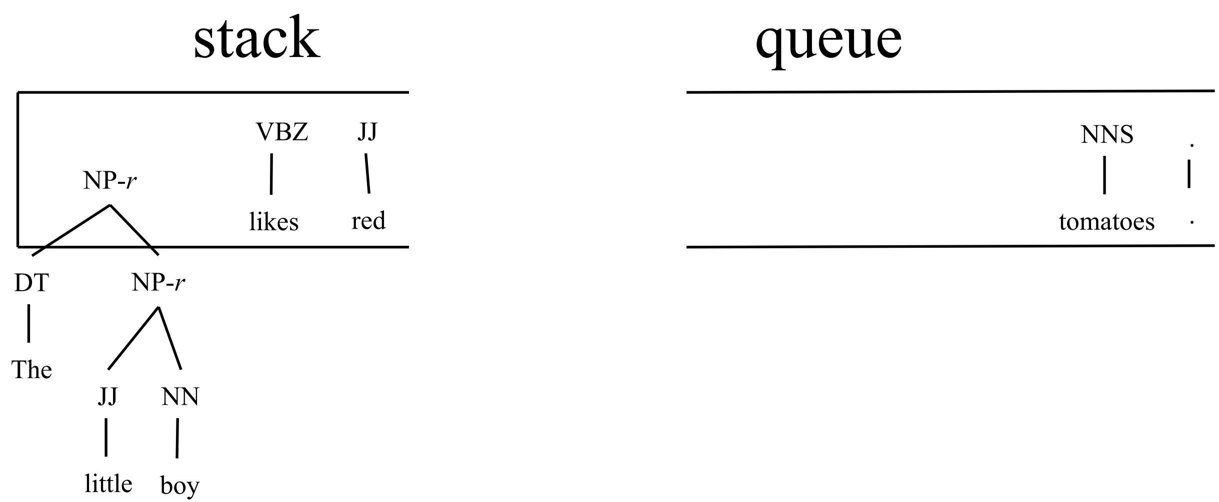
- Actions
  - Shift



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

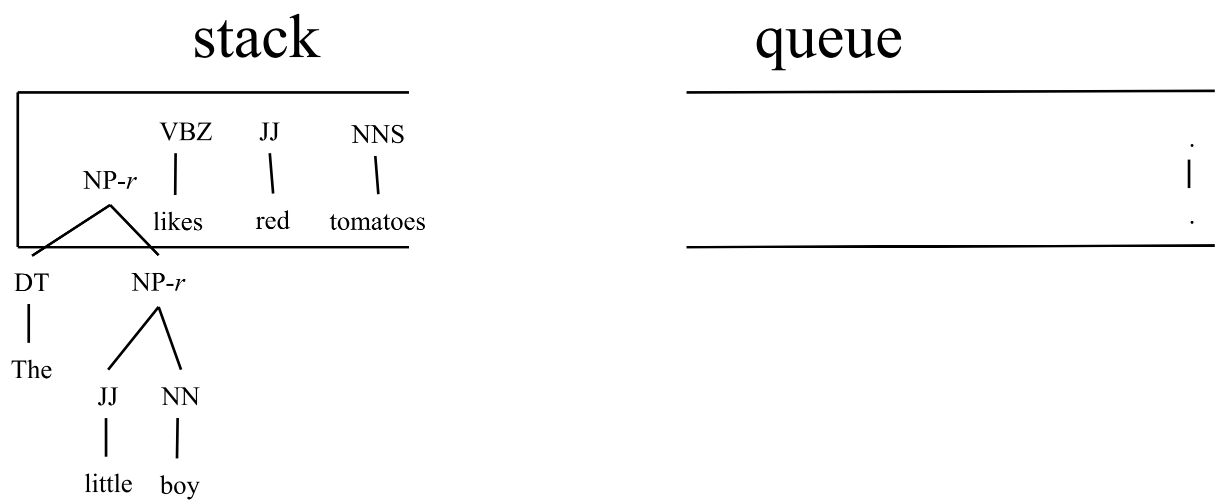
- Actions
  - Shift



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

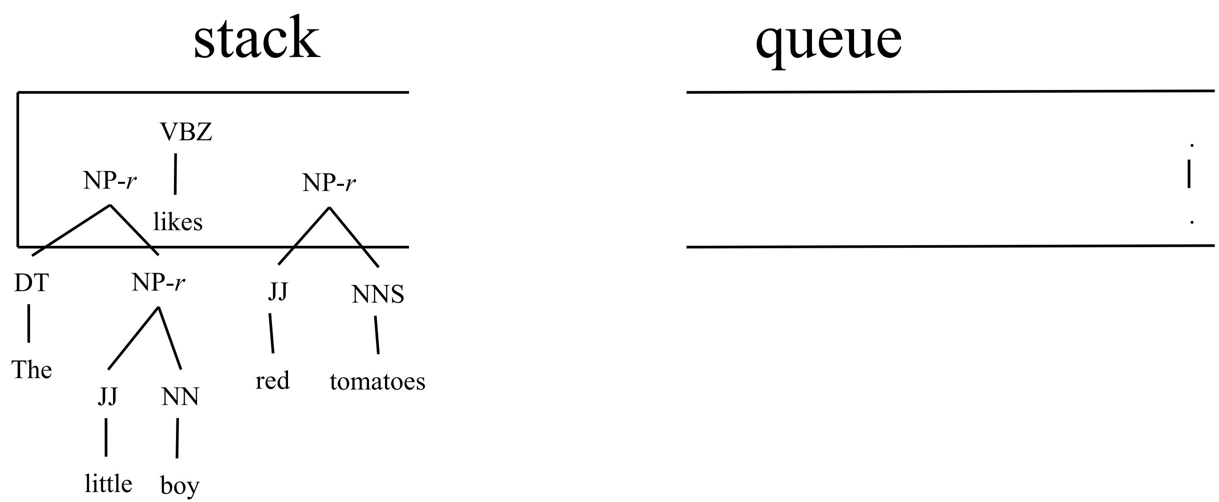
- Actions
  - Reduce-r-NP



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

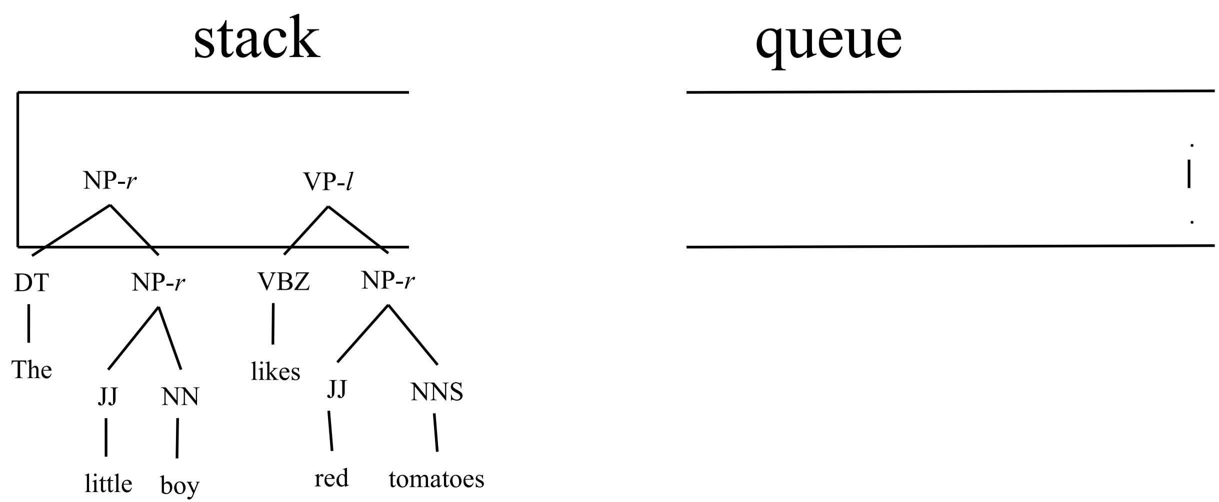
- Actions
  - Reduce-l-VP



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

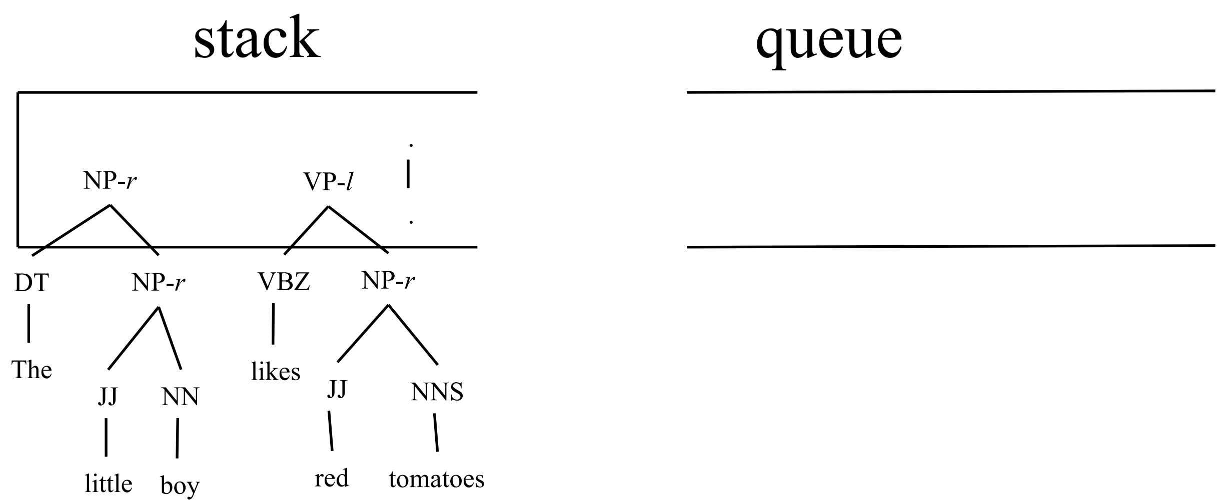
- Actions
  - Shift



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

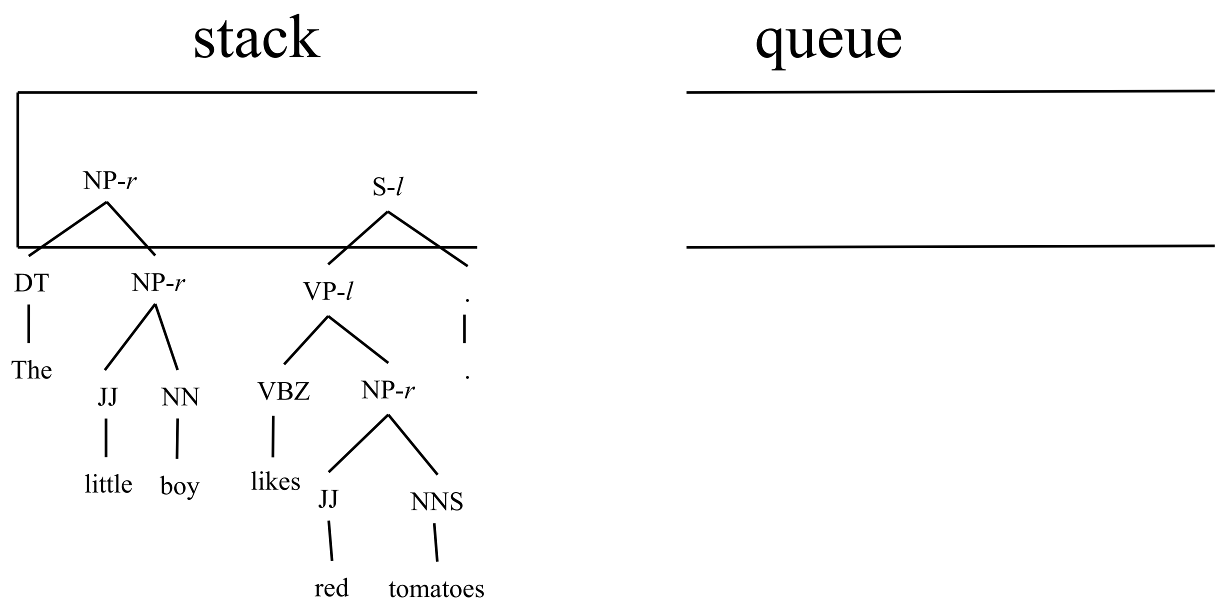
- Actions
  - Reduce-l-S



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

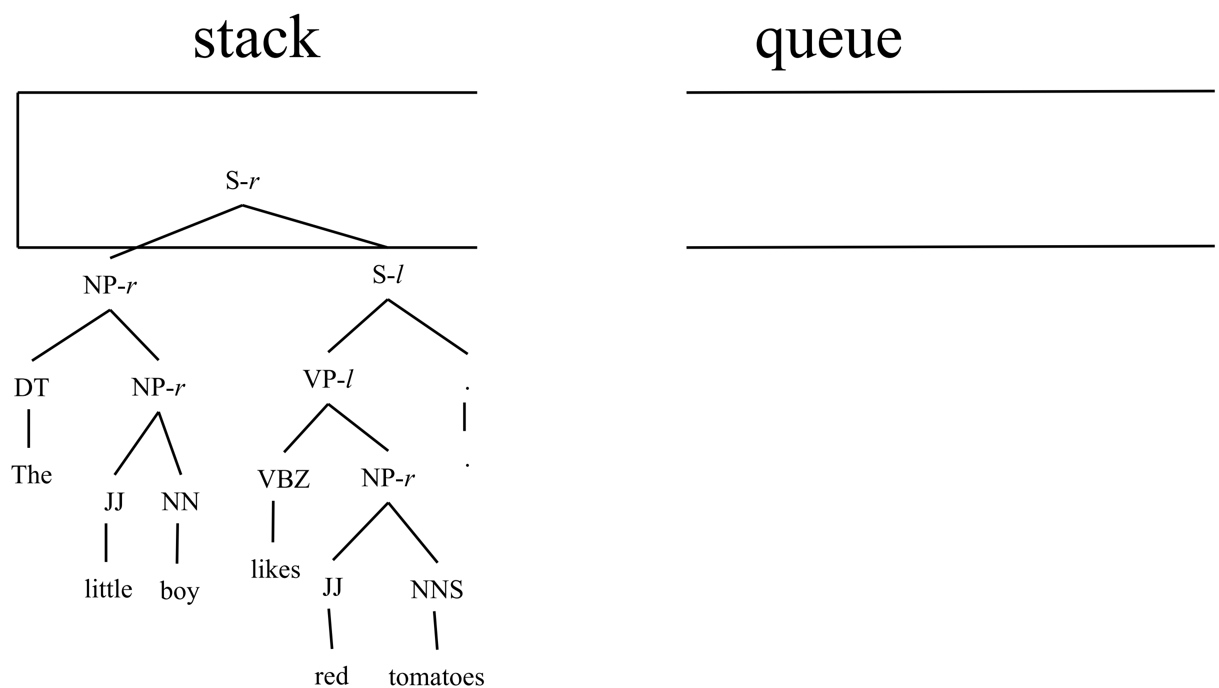
- Actions
  - Reduce-r-S



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

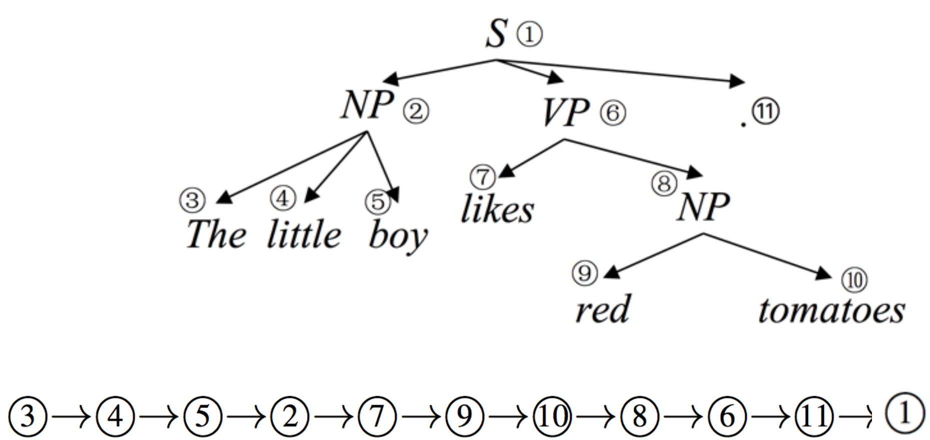
- Actions
  - Terminate



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing (Bottom-up)

- Bottom-up guidance
  - Rich local features from readily constructed trees
  - Lack of the look-ahead guidance
  - Post-order traversal on the tree



Watanabe, Taro, and Eiichiro Sumita. "Transition-based Neural Constituent Parsing." *ACL (1)*. 2015.

Constituent parsing

- Model
  - Parser Transitions (Top-down)
  - Generator Transitions

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing

- Model
  - Parser Transitions (Top-down)
    - NT(X) introduces an “open nonterminal” X onto the top of the stack.
    - SHIFT
    - REDUCE

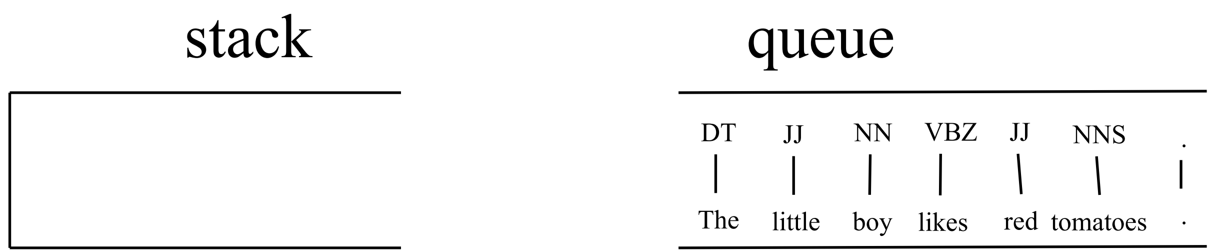
Stack$_t$	**Buffer**$_t$	**Open NTs**$_t$	**Action**	**Stack**$_{t+1}$	**Buffer**$_{t+1}$	**Open NTs**$_{t+1}$
S	B	n	NT(X)	$S \mid (\mathrm{X}$	B	$n+1$
S	$x \mid B$	n	SHIFT	$S \mid x$	B	n
$S \mid (\mathrm{X} \mid \tau_1 \mid \ldots \mid \tau_\ell$	B	n	REDUCE	$S \mid (\mathrm{X}\ \tau_1\ \ldots\ \tau_\ell)$	B	$n-1$

(a) Parser Transitions

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

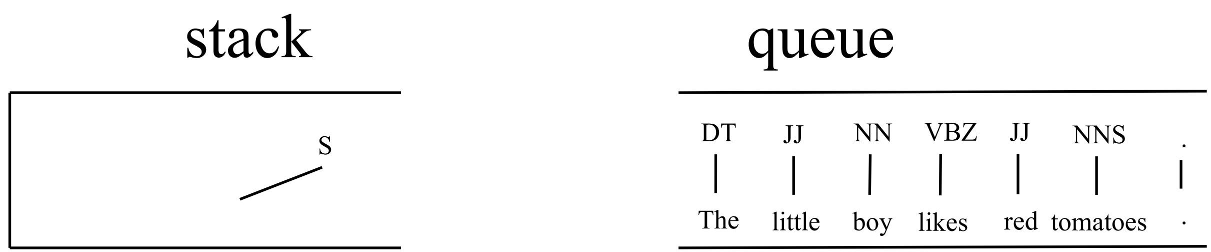
- Actions
  - NT(S)



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

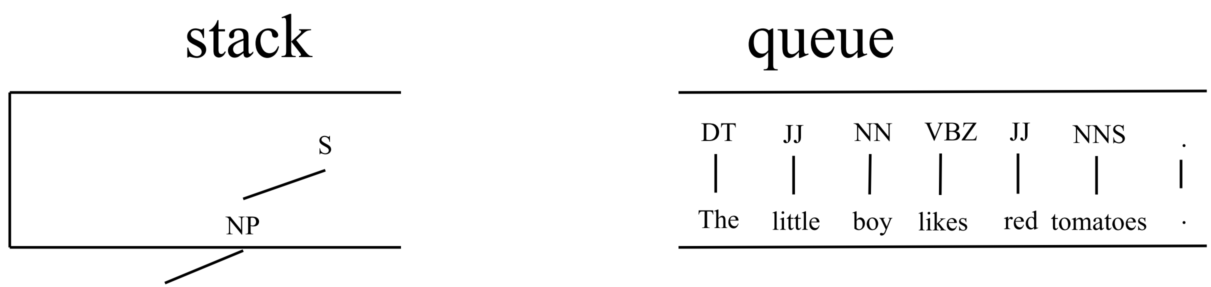
- Actions
  - NT(NP)



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

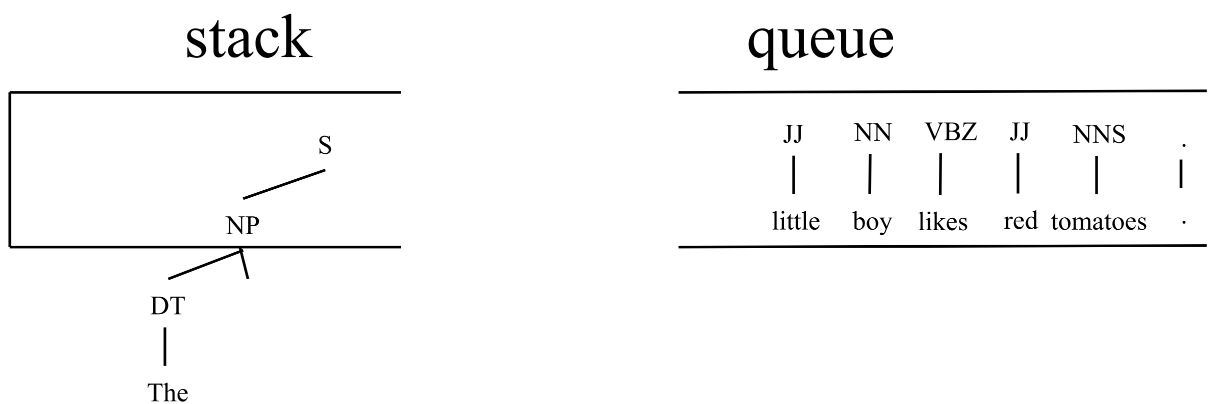
- Actions
  - Shift



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

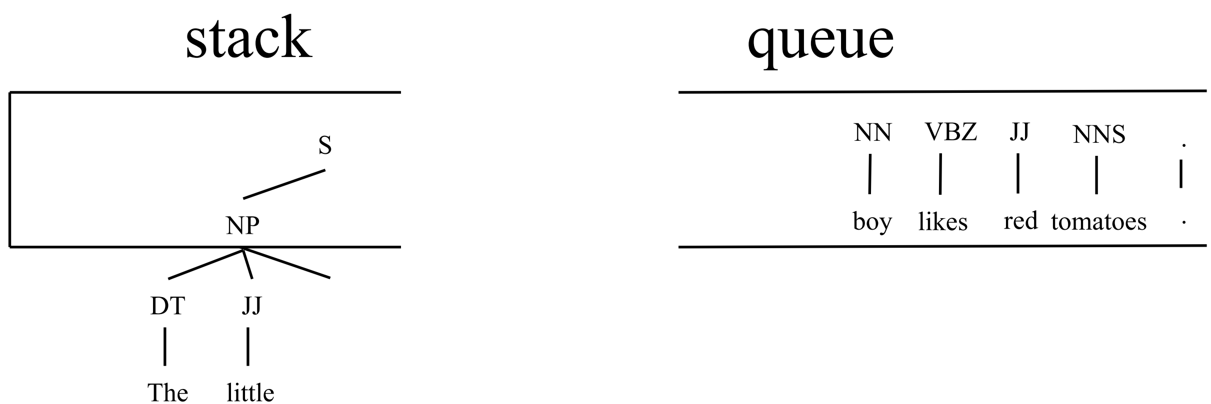
- Actions
  - Shift



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

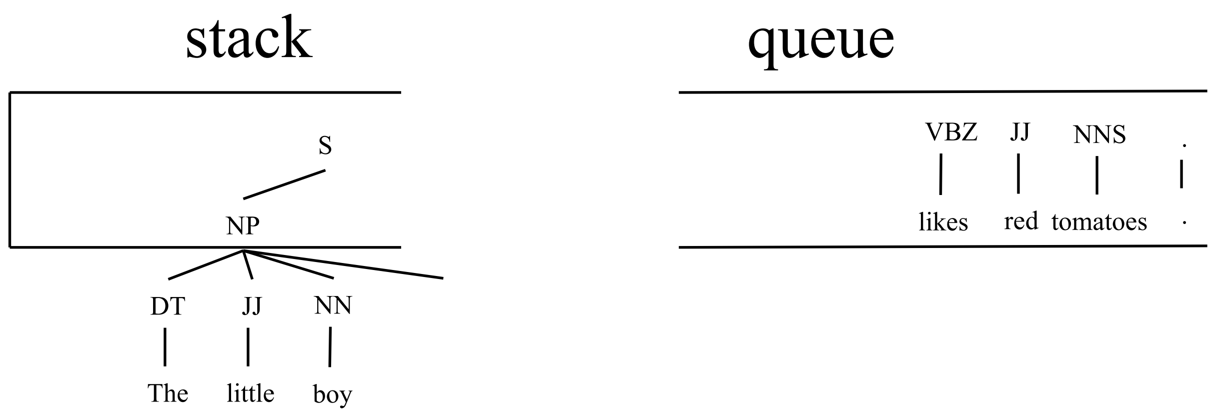
- Actions
  - Shift



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

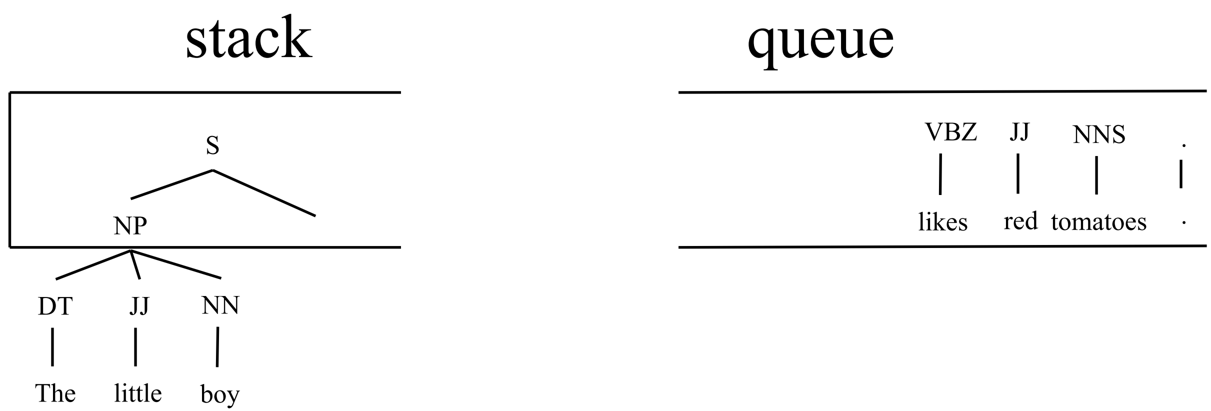
- Actions
  - Reduce



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

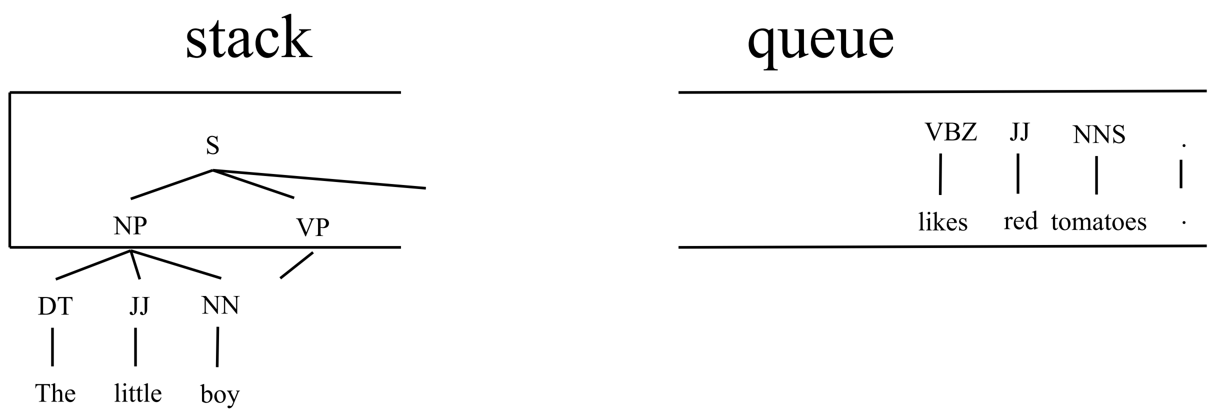
- Actions
  - NT(VP)



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

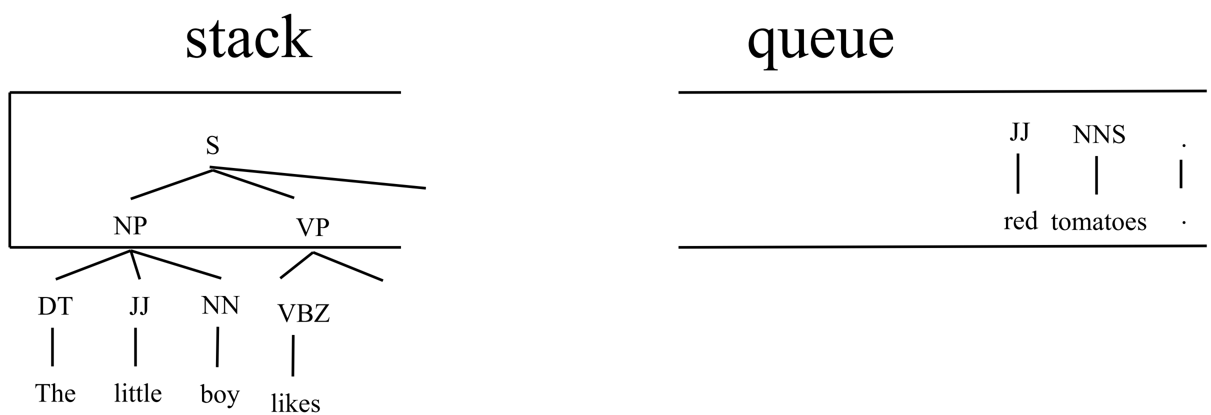
- Actions
  - Shift



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

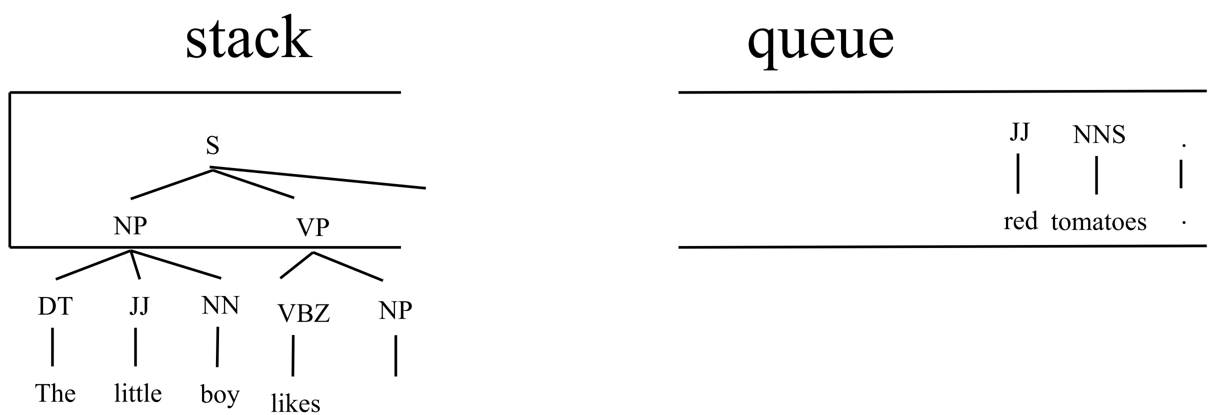
- Actions
  - NT(NP)



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

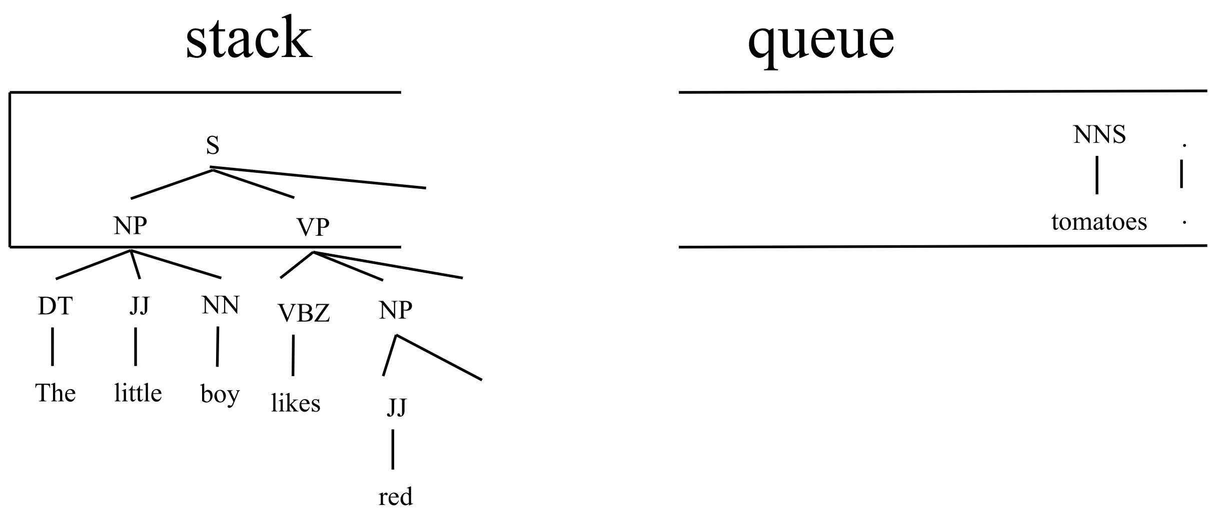
- Actions
  - Shift



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

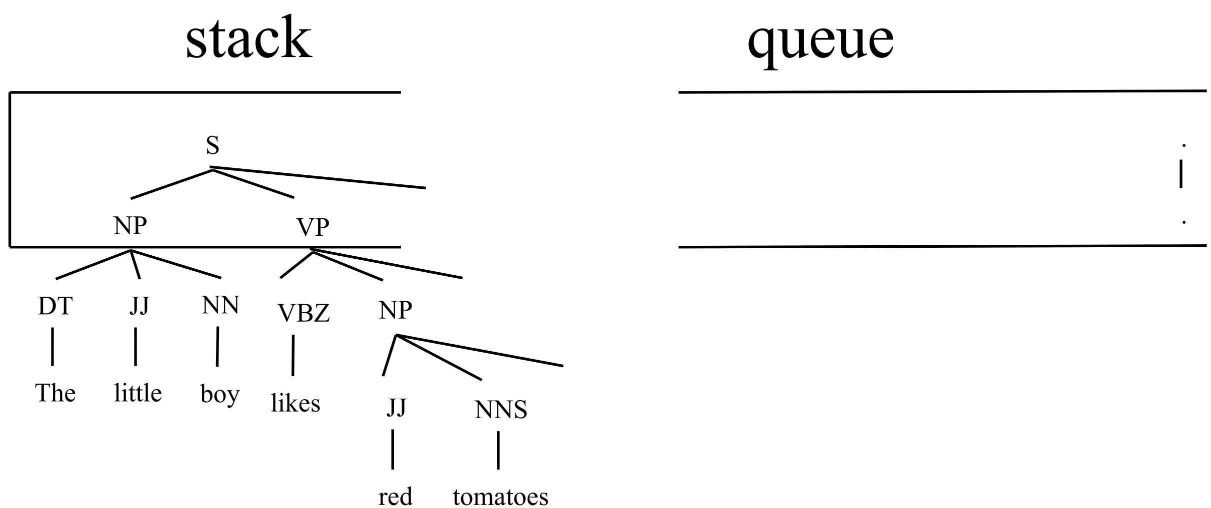
- Actions
  - Shift



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

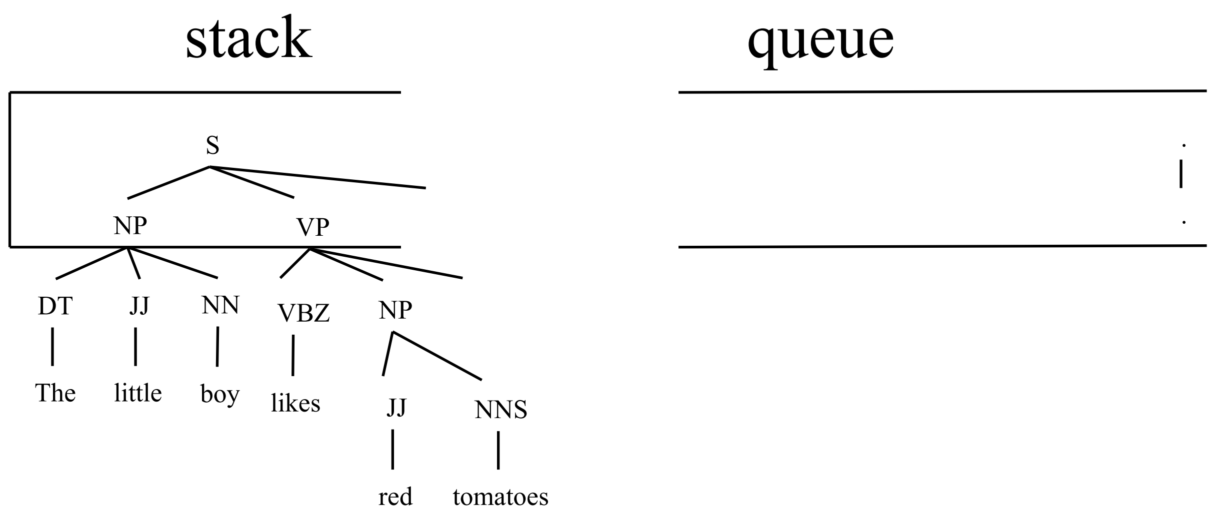
- Actions
  - Reduce



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

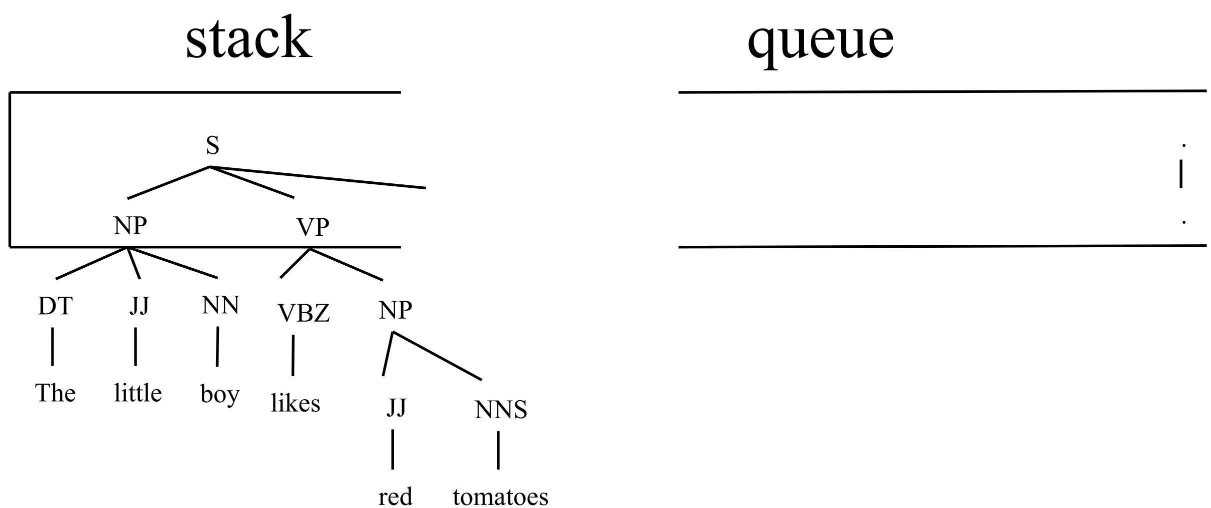
- Actions
  - Reduce



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

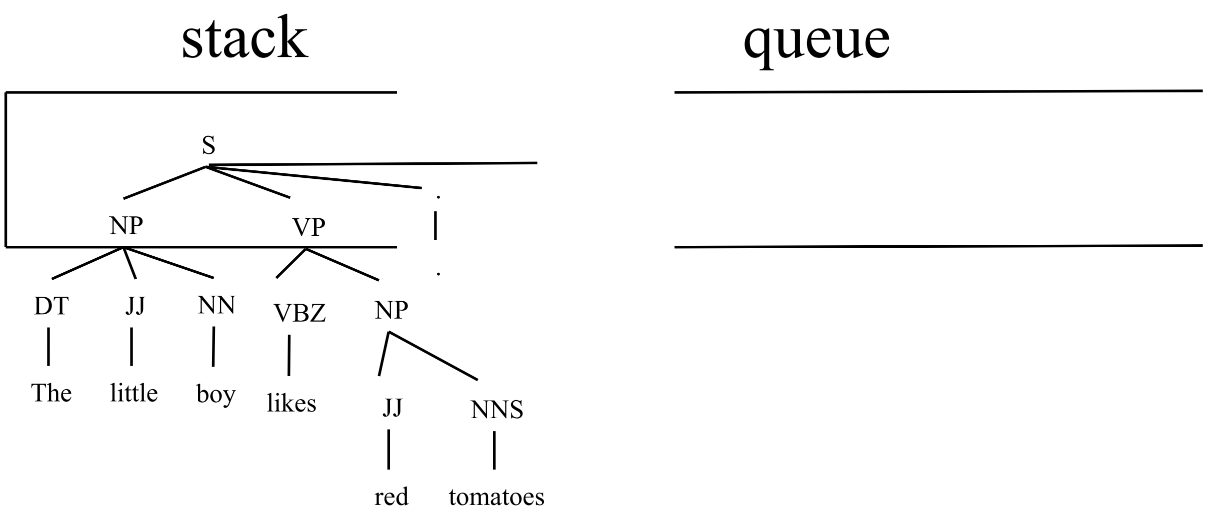
- Actions
  - Shift



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

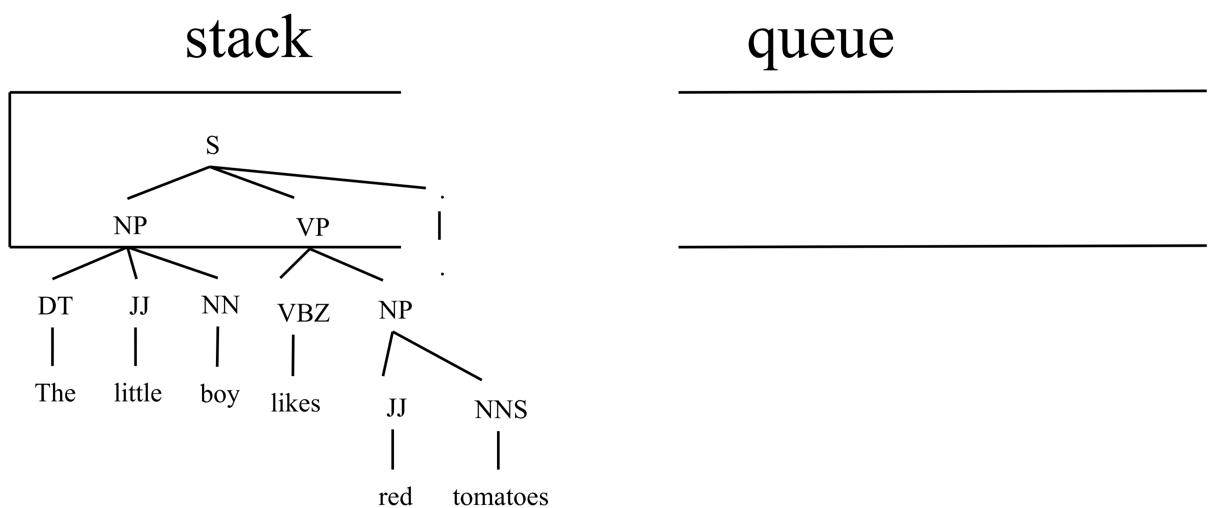
- Actions
  - Reduce



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

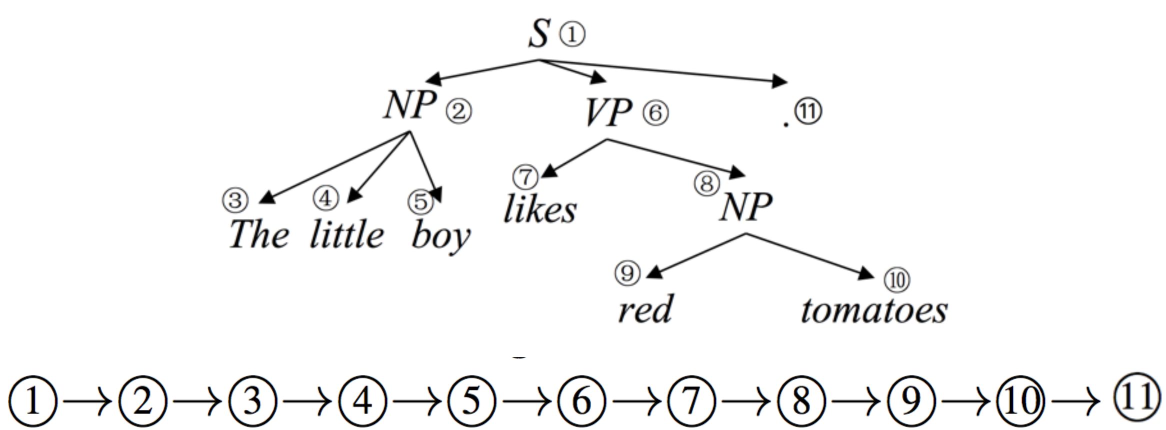
- Actions
  - Terminate



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (Top-down)

- Top-down guidance
  - Non-local information for local decision
  - Strong encoders over the input to predict a constituent hierarchy before its construction
  - Pre-order traversal on the tree



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing

- Generative Model

$$p(\boldsymbol{x}, \boldsymbol{y}) = \prod_{t=1}^{|\boldsymbol{a}(\boldsymbol{x},\boldsymbol{y})|} p(a_t \mid \boldsymbol{a}_{<t})$$

$$= \prod_{t=1}^{|\boldsymbol{a}(\boldsymbol{x},\boldsymbol{y})|} \frac{\exp \mathbf{r}_{a_t}^\top \mathbf{u}_t + b_{a_t}}{\sum_{a' \in \mathcal{A}_G(T_t, S_t, n_t)} \exp \mathbf{r}_{a'}^\top \mathbf{u}_t + b_{a'}},$$

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing

- Model
  - Generator Transitions
    - GEN(x) operations which generate terminal symbol x $\in$ Σ and add it to the top of the stack and the out-put buffer

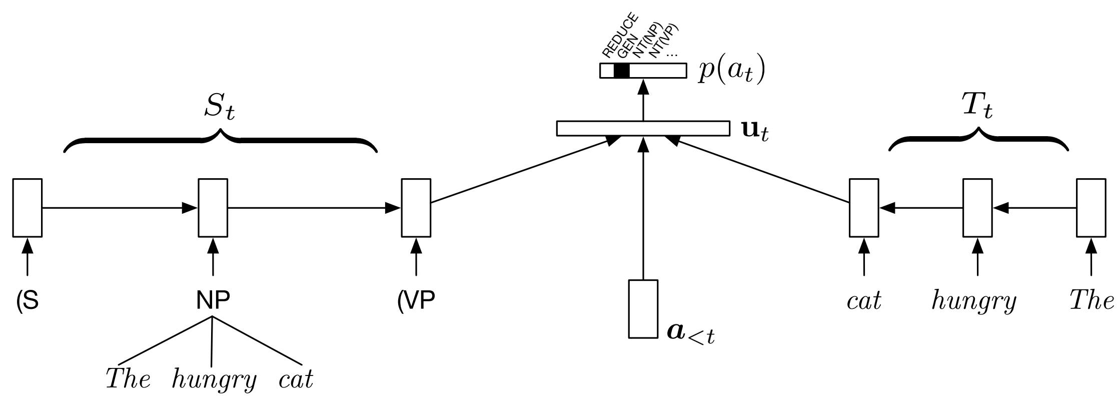
Stack$_t$	**Terms**$_t$	**Open NTs**$_t$	**Action**	**Stack**$_{t+1}$	**Terms**$_{t+1}$	**Open NTs**$_{t+1}$
S	T	n	NT(X)	$S \mid$ (X	T	$n+1$
S	T	n	GEN(x)	$S \mid x$	$T \mid x$	n
$S \mid$ (X $\mid \tau_1 \mid \ldots \mid \tau_\ell$	T	n	REDUCE	$S \mid$ (X $\tau_1 \ldots \tau_\ell$)	T	$n-1$

(b) Generator Transitions

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing

- Generative Model
  - Distribution over stack ($\boldsymbol{S_t}$), output buffer ($\boldsymbol{T_t}$) and history of actions (a < t)



Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing

- English Results

D=discriminative,
G=generative,
S=Semi-supervised

Model	**type**	$\mathbf{F_1}$
Vinyals et al. (2015)* – WSJ only	D	88.3
Henderson (2004)	D	89.4
Socher et al. (2013a)	D	90.4
Zhu et al. (2013)	D	90.4
Petrov and Klein (2007)	G	90.1
Bod (2003)	G	90.7
Shindo et al. (2012) – single	G	91.1
Shindo et al. (2012) – ensemble	G	92.4
Zhu et al. (2013)	S	91.3
McClosky et al. (2006)	S	92.1
Vinyals et al. (2015) – single	S	92.1
Discriminative, $q(\boldsymbol{y} \mid \boldsymbol{x})$	D	89.8
Generative, $\hat{p}(\boldsymbol{y} \mid \boldsymbol{x})$	G	92.4

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing

- Chinese Results

D=discriminative,
G=generative,
S=Semi-supervised

Model	**type**	$\mathbf{F_1}$
Zhu et al. (2013)	D	82.6
Wang et al. (2015)	D	83.2
Huang and Harper (2009)	D	84.2
Charniak (2000)	G	80.8
Bikel (2004)	G	80.6
Petrov and Klein (2007)	G	83.3
Zhu et al. (2013)	S	85.6
Wang and Xue (2014)	S	86.3
Wang et al. (2015)	S	86.6
Discriminative, $q(\boldsymbol{y} \mid \boldsymbol{x})$	D	80.7
Generative, $\hat{p}(\boldsymbol{y} \mid \boldsymbol{x})$	G	82.7

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith. Recurrent neural network grammars[J]. In NAACL, 2016.

Constituent parsing (In-Order)

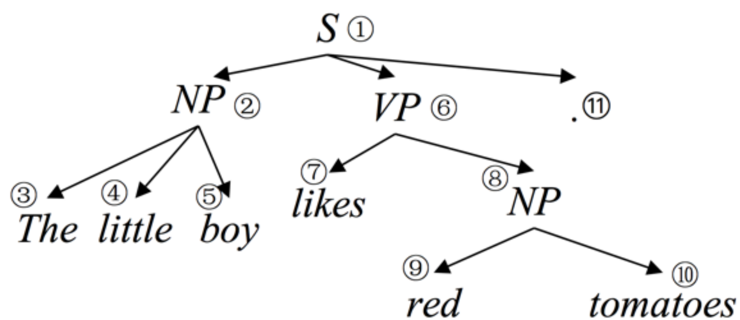
- Trade-off
  - Compromise between bottom-up constituent information and top-down lookahead information

Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

- In-order traversal on the non-binary tree

  – Regard the left-most child as the left branch of the binary tree, and the rest children are traced from left to right.

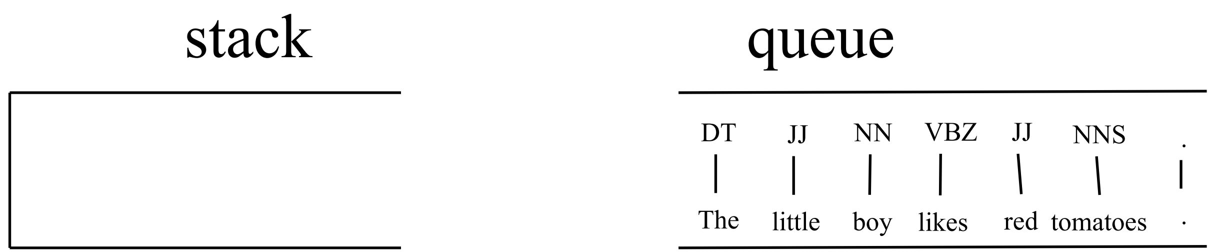


③→②→④→⑤→①→⑦→⑥→⑨→⑧→⑩→⑪

Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

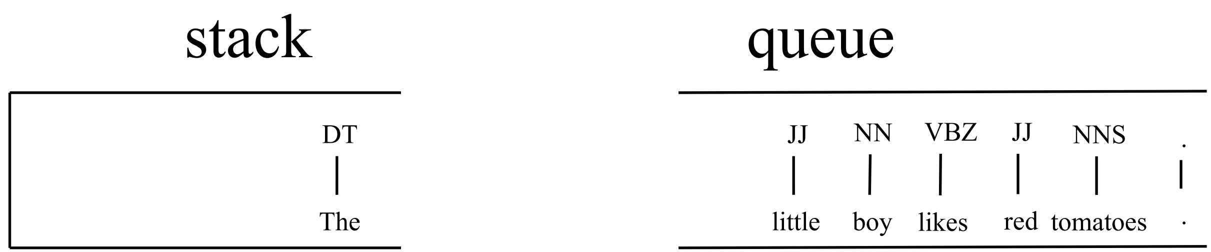
- Actions
  - Shift



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

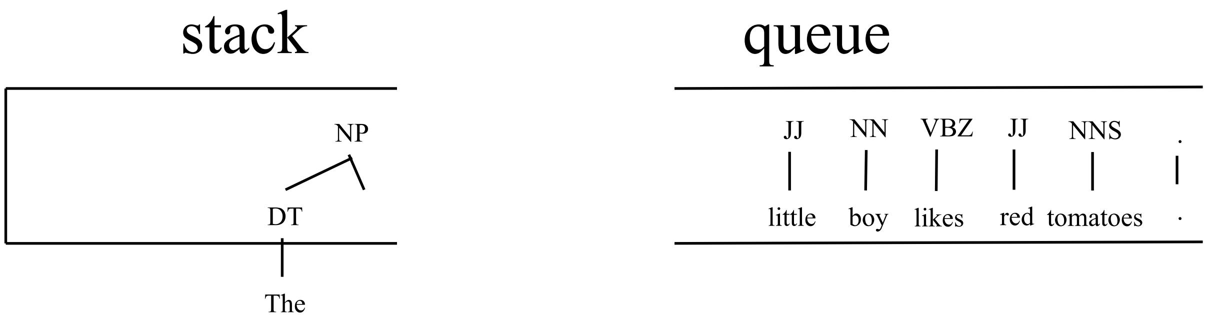
- Actions
  - PJ(NP)



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

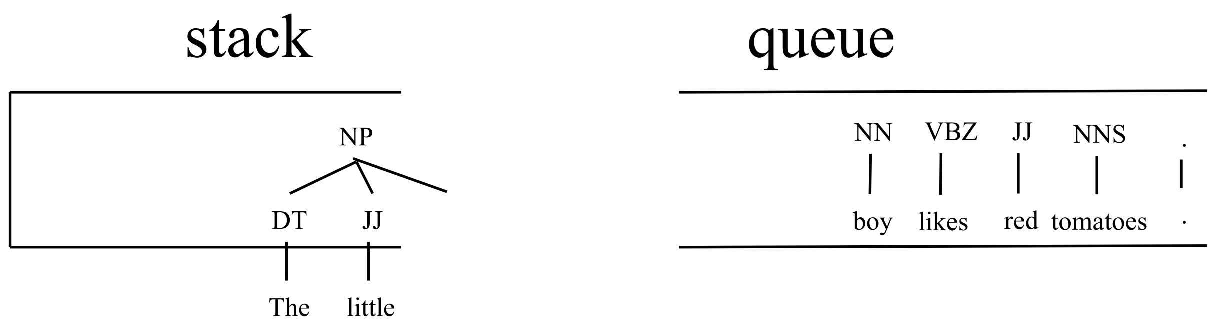
- Actions
  - Shift



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

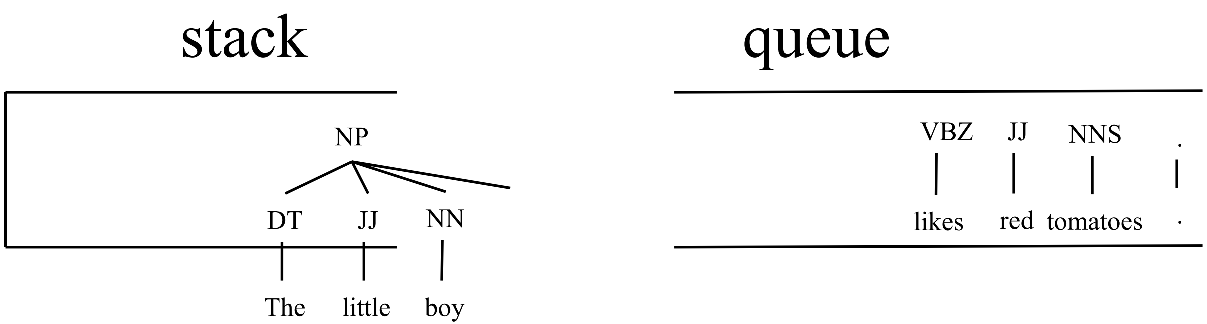
- Actions
  - Shift



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

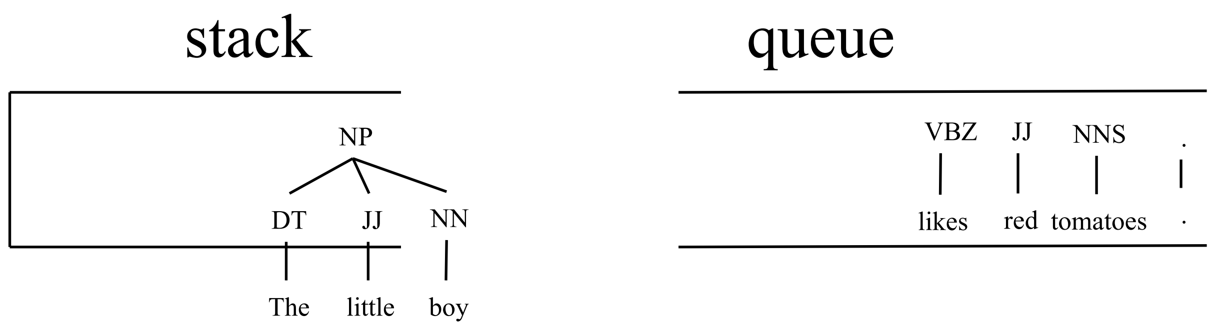
- Actions
  - Reduce



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

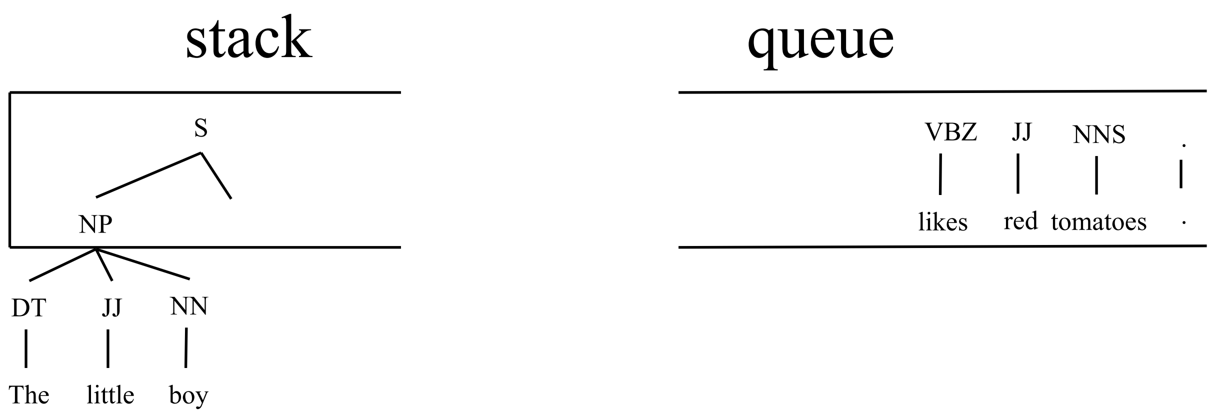
- Actions
  - PJ(S)



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

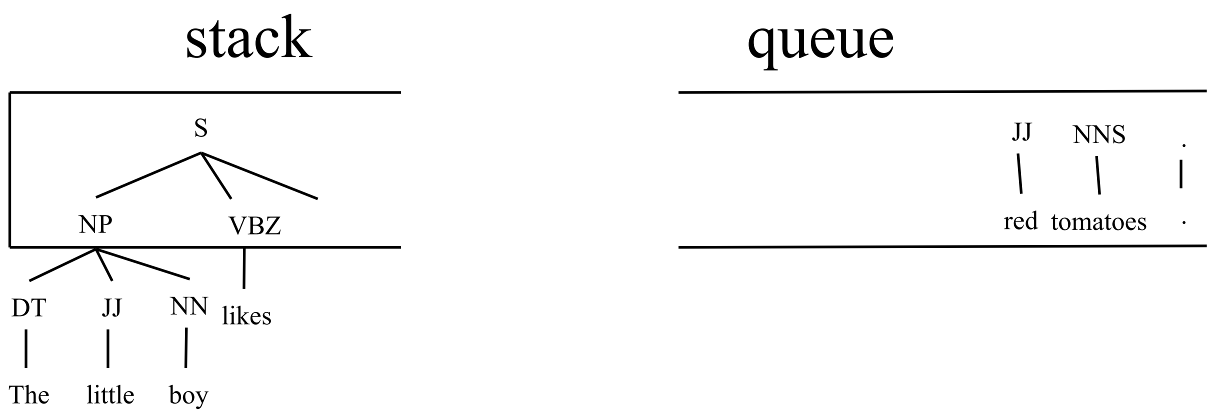
- Actions
  - Shift



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

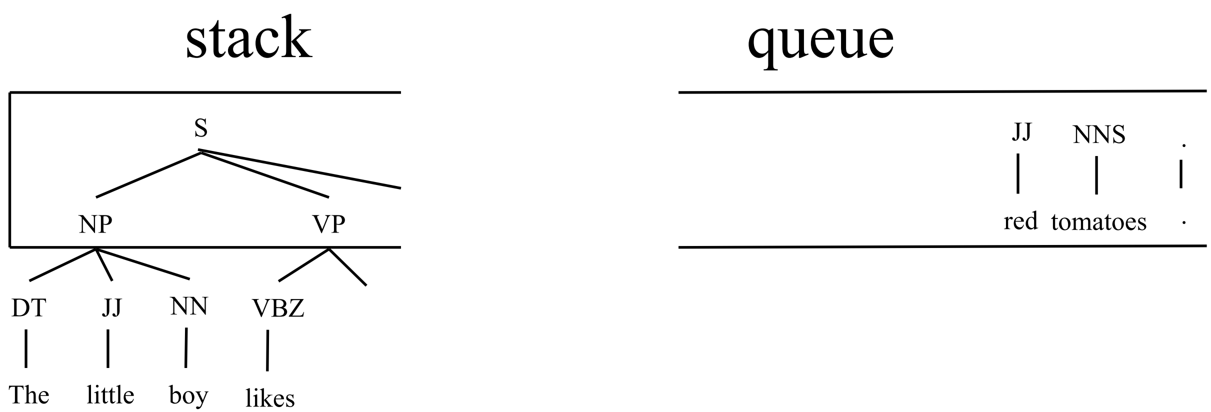
- Actions
  - PJ(VP)



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

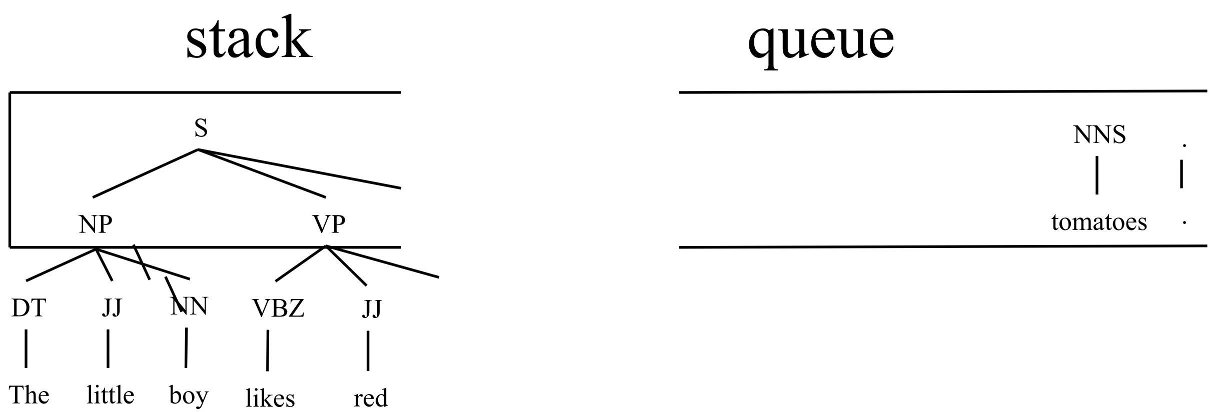
- Actions
  - Shift



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

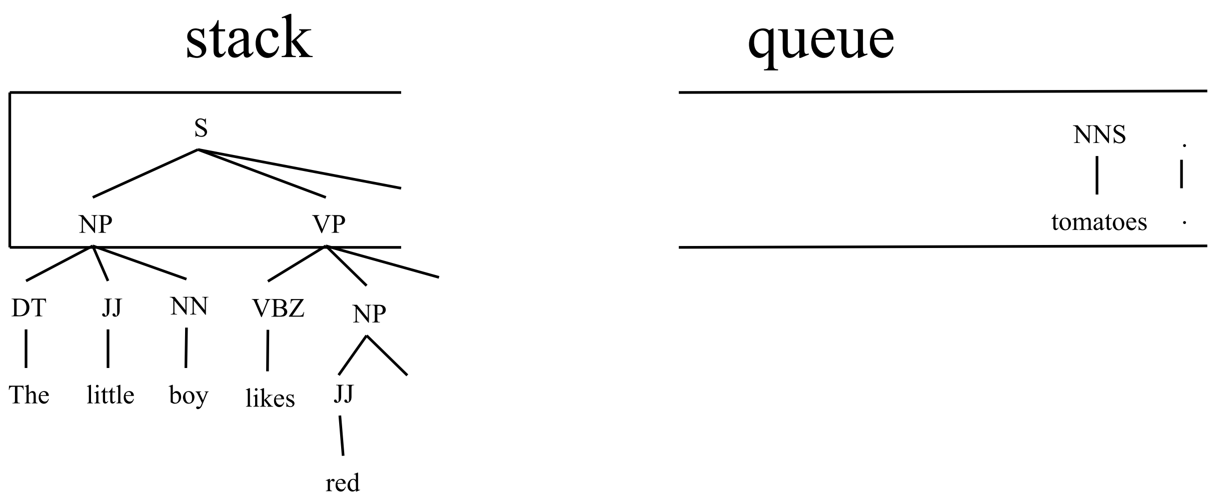
- Actions
  - PJ(NP)



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

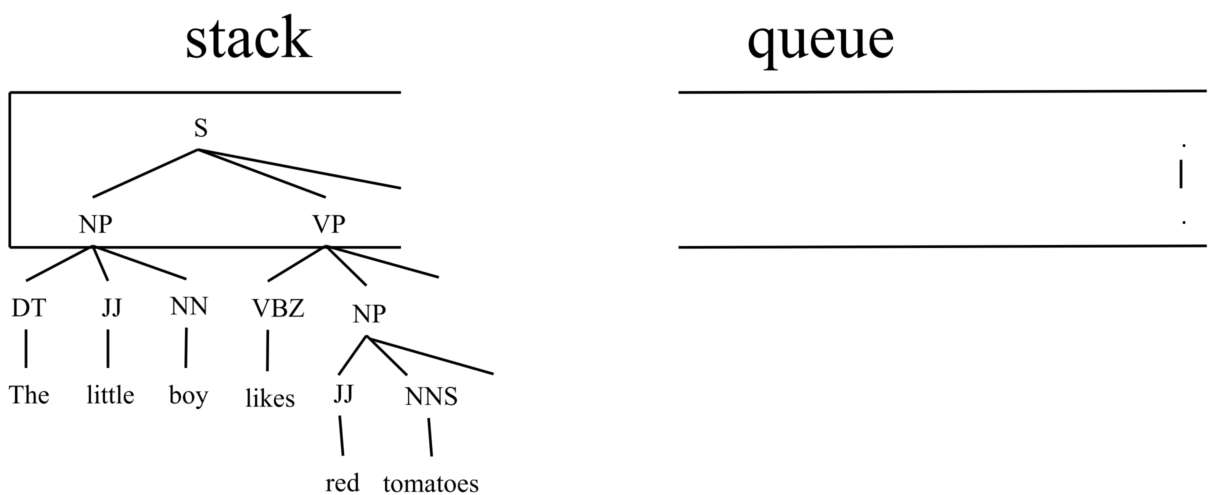
- Actions
  - Shift



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

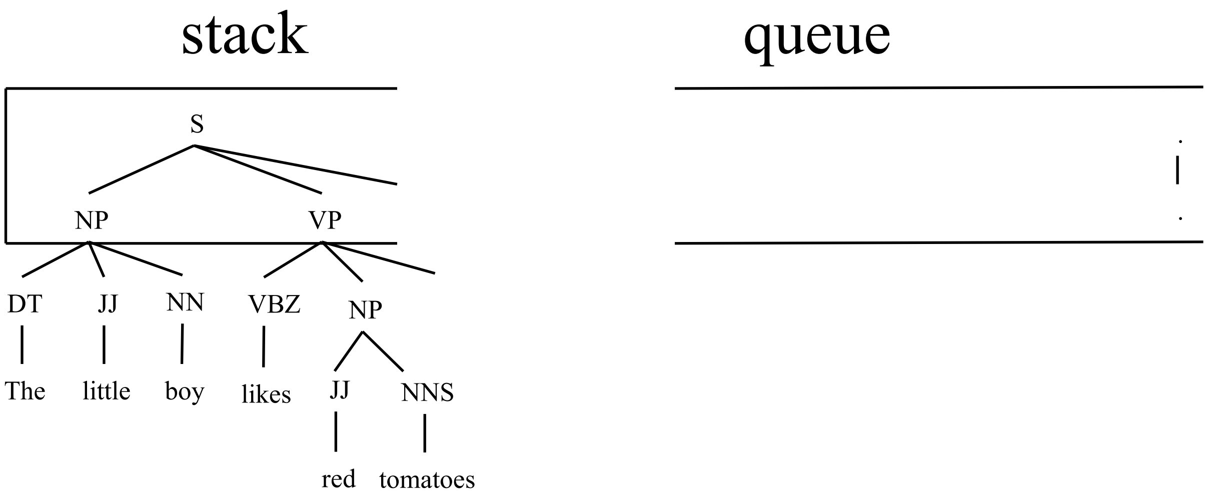
- Actions
  - Reduce



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

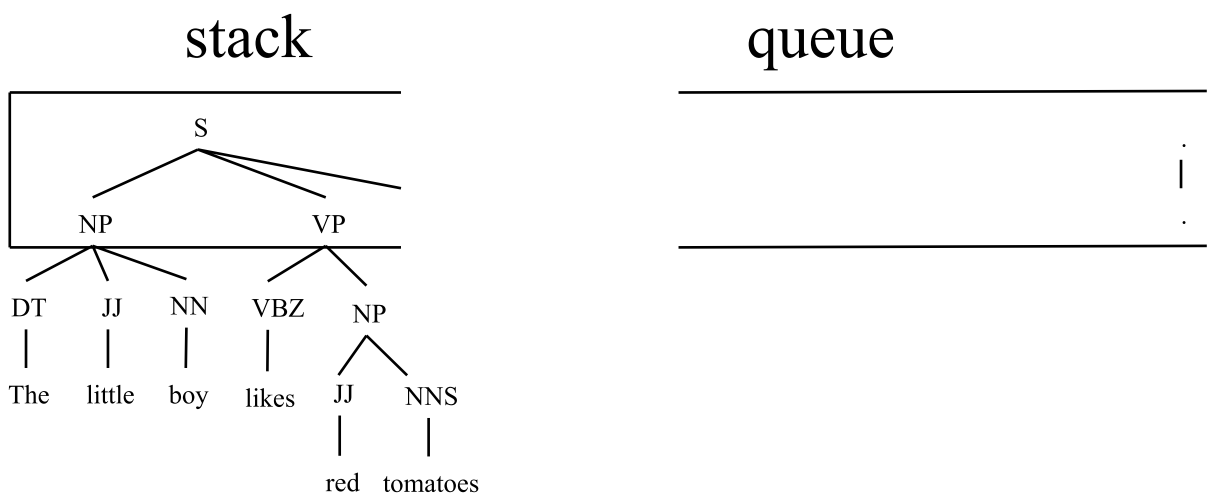
- Actions
  - Reduce



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

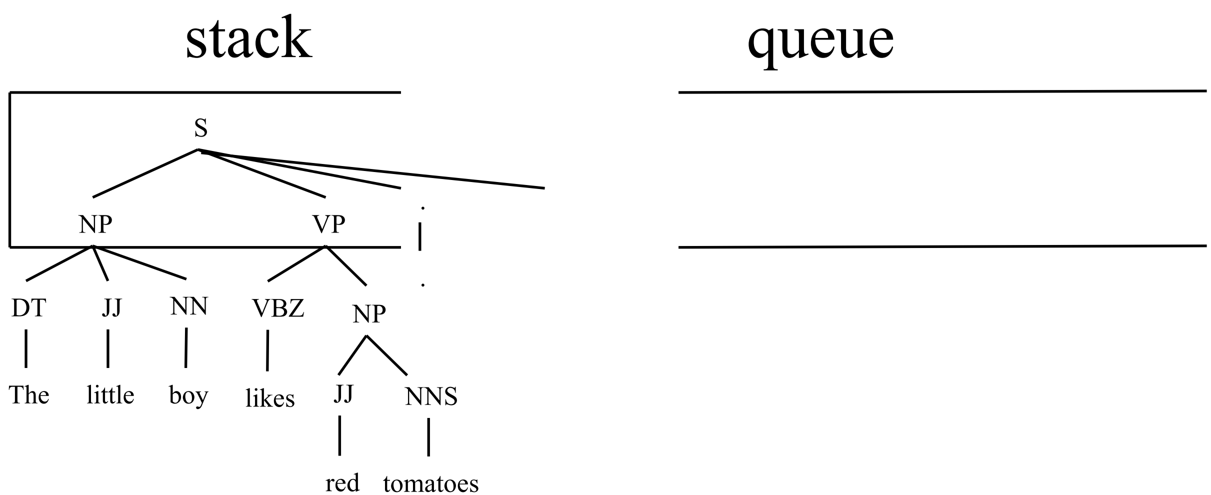
- Actions
  - Shift



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

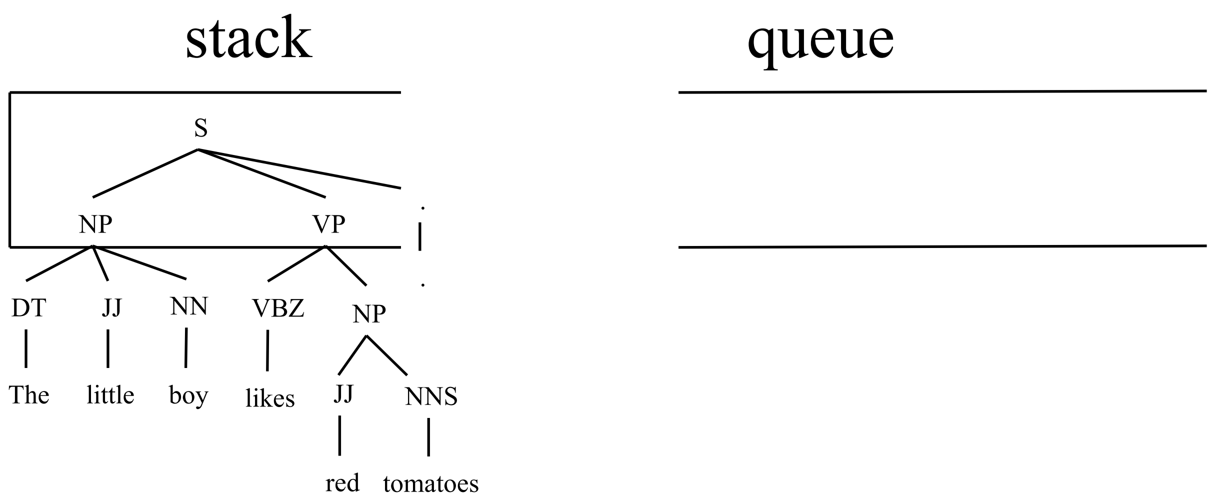
- Actions
  - Reduce



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Constituent parsing (In-Order)

- Actions
  - Terminate



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

In-Order Constituent parsing

- Different Transition Systems

SHIFT $\dfrac{[\sigma, i, false]}{[\sigma|w_i, i+1, false]}$

REDUCE-L/R-X $\dfrac{[\sigma|s_1|s_0, i, false]}{[\sigma|X_{s_1 s_0}, i, false]}$

Unary-X $\dfrac{[\sigma|s_0, i, false]}{[\sigma|X_{s_0}, i, false]}$

FINISH $\dfrac{[\sigma, i, false]}{[\sigma, i, true]}$

(a) bottom-up system

SHIFT $\dfrac{[\sigma, i, /]}{[\sigma|w_i, i+1, /]}$

NT-X $\dfrac{[\sigma, i, /]}{(\sigma|X, i, /]}$

REDUCE $\dfrac{[\sigma|X|s_j|...|s_0, i, /]}{[\sigma|X_{s_j ... s_0}, i, /]}$

(b) top-down system

SHIFT $\dfrac{[\sigma, i, false]}{[\sigma|w_i, i+1, false]}$

PJ-X $\dfrac{[\sigma|s_0, i, false]}{(\sigma|s_0|X, i, false]}$

REDUCE $\dfrac{[\sigma|s_j|X|s_{j-1}|...|s_0, i, false]}{[\sigma|X_{s_j s_{j-1} ... s_0}, i, false]}$

FINISH $\dfrac{[\sigma, i, false]}{[\sigma, i, true]}$

(c) in-order system

Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

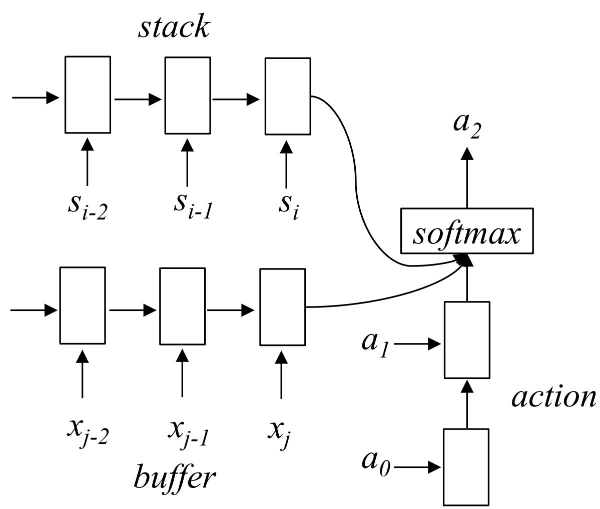
In-Order Constituent parsing

- Generalization
  - In-Order system can be generalized into variants by modifying *k*, the number of leftmost nodes traced before the parent node.
  - If *k = 0*, top-down system is a special case
  - If *k = inf*, bottom-up system is a special case

Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

In-Order Constituent parsing

- Models



Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

In-Order Constituent parsing

- Results
  - English Constituent Results (on WSJ Section 23)

Model	F_1
fully-supervision	
Socher et al. (2013)	90.4
Zhu et al. (2013)	90.4
Vinyals et al. (2015)	90.7
Watanabe and Sumita (2015)	90.7
Shindo et al. (2012)	91.1
Durrett and Klein (2015)	91.1
Dyer et al. (2016)	91.2
Cross and Huang (2016)	91.3
Liu and Zhang (2017)	91.7
Top-down parser	91.2
Bottom-up parser	91.3
In-order parser	**91.8**
reranking	
Huang (2008)	91.7
Charniak and Johnson (2005)	91.5
Choe and Charniak (2016)	92.6
Dyer et al. (2016)	93.3
Kuncoro et al. (2017)	**93.6**
Top-down parser	93.3
Bottom-up parser	93.3
In-order parser	**93.6**
semi-supervised reranking	
Choe and Charniak (2016)	93.8
In-order parser	**94.2**

Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

In-Order Constituent parsing

- Results
  - English Dependency Results(on WSJ Section 23)

Model	UAS	LAS
Kiperwasser and Goldberg (2016)†	93.9	91.9
Cheng et al. (2016) †	94.1	91.5
Andor et al. (2016)	94.6	92.8
Dyer et al. (2016) -re	95.6	94.4
Dozat and Manning (2017)†	95.7	94.0
Kuncoro et al. (2017) -re	95.7	94.5
Choe and Charniak (2016) -sre	95.9	94.1
In-order parser	94.5	93.4
In-order parser -re	95.9	94.9
In-order parser -sre	**96.2**	**95.2**

Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

In-Order Constituent parsing

- Results
  - Chinese Constituent Results (on CTB Test set)

Parser	F_1
fully-supervision	
Zhu et al. (2013)	83.2
Wang et al. (2015)	83.2
Dyer et al. (2016)	84.6
Liu and Zhang (2017)	85.5
Top-down parser	84.6
Bottom-up parser	85.7
In-order parser	**86.1**
rerank	
Charniak and Johnson (2005)	82.3
Dyer et al. (2016)	86.9
Top-down parser	86.9
Bottom-up parser	87.5
In-order parser	**88.0**

Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

In-Order Constituent parsing

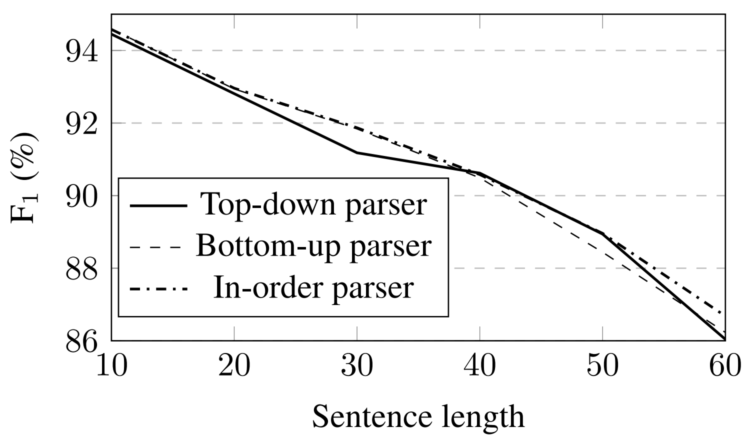
- Results
  - Chinese Dependency Results(on CTB Test set)

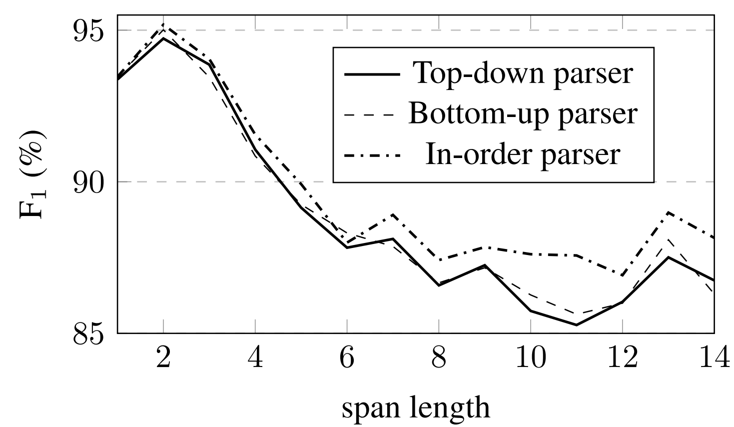
Model	UAS	LAS
Dyer et al. (2016)	85.5	84.0
Ballesteros et al. (2016)	87.7	86.2
Kiperwasser and Goldberg (2016)	87.6	86.1
Cheng et al. (2016) †	88.1	85.7
Dozat and Manning (2017) †	89.3	88.2
In-order parser	87.4	86.4
In-order parser -re	**89.4**	**88.4**

Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

In-Order Constituent parsing

- Results
  - Comparison with other system

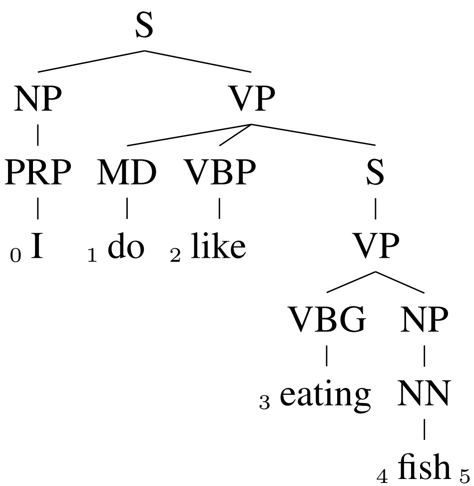




Jiangming Liu, Yue Zhang, In-Order Transition-based Constituent Parsing. In TACL(2017)

Span-Based Constituency Parsing

- Example



(a) gold parse tree

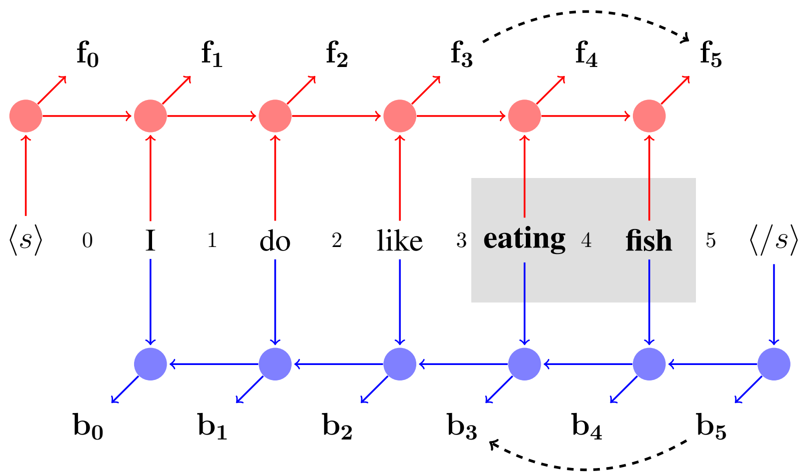
steps	structural action	label action	stack after	bracket
1–2	sh(I/PRP)	label-NP	$_0\triangle_1$	$_0\text{NP}_1$
3–4	sh(do/MD)	nolabel	$_0\triangle_1\triangle_2$	
5–6	sh(like/VBP)	nolabel	$_0\triangle_1\triangle_2\triangle_3$	
7–8	comb	nolabel	$_0\triangle_1\triangle_3$	
9–10	sh(eating/VBG)	nolabel	$_0\triangle_1\triangle_3\triangle_4$	
11–12	sh(fish/NN)	label-NP	$_0\triangle_1\triangle_3\triangle_4\triangle_5$	$_4\text{NP}_5$
13–14	comb	label-S-VP	$_0\triangle_1\triangle_3\triangle_5$	$_3\text{S}_5$, $_3\text{VP}_5$
15–16	comb	label-VP	$_0\triangle_1\triangle_5$	$_1\text{VP}_5$
17–18	comb	label-S	$_0\triangle_5$	$_0\text{S}_5$

(b) static oracle actions

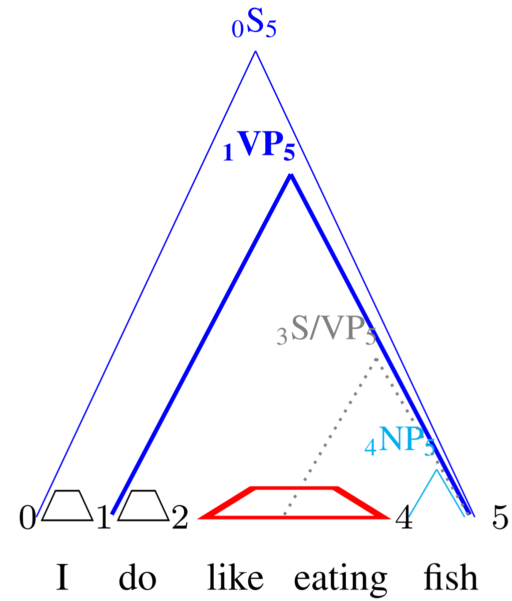
Cross, James, and Liang Huang. "Span-based constituency parsing with a structure-label system and provably optimal dynamic oracles." In *EMNLP* (2016).

Span-Based Constituency Parsing

- Models



(a) Words spans modeling from LSTM output



(b) Dynamic Oracle

Cross, James, and Liang Huang. "Span-based constituency parsing with a structure-label system and provably optimal dynamic oracles." In *EMNLP* (2016).

Span-Based Constituency Parsing

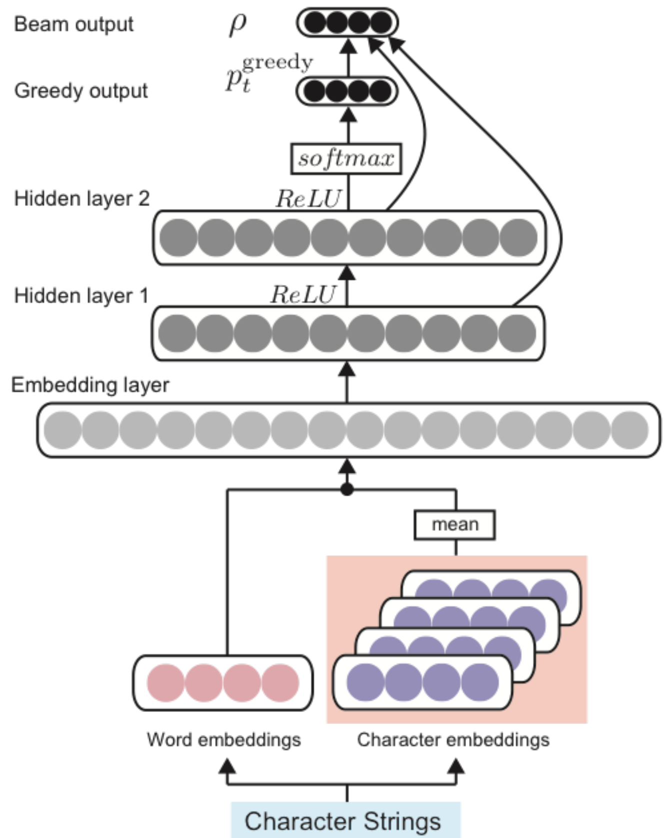
- Results

Closed Training & Single Model	LR	LP	F_1
Sagae and Lavie (2006)	88.1	87.8	87.9
Petrov and Klein (2007)	90.1	90.3	90.2
Carreras et al. (2008)	90.7	91.4	91.1
Shindo et al. (2012)			91.1
†Socher et al. (2013)			90.4
Zhu et al. (2013)	90.2	90.7	90.4
Mi and Huang (2015)	90.7	90.9	90.8
†Watanabe and Sumita (2015)			90.7
Thang et al. (2015) (A*)	90.9	91.2	91.1
†*Dyer et al. (2016) (discrim.)			89.8
†*Cross and Huang (2016)			90.0
†***static oracle**	90.7	91.4	91.0
†***dynamic/exploration**	90.5	92.1	**91.3**

Cross, James, and Liang Huang. "Span-based constituency parsing with a structure-label system and provably optimal dynamic oracles." In *EMNLP* (2016).

Neural Joint Model

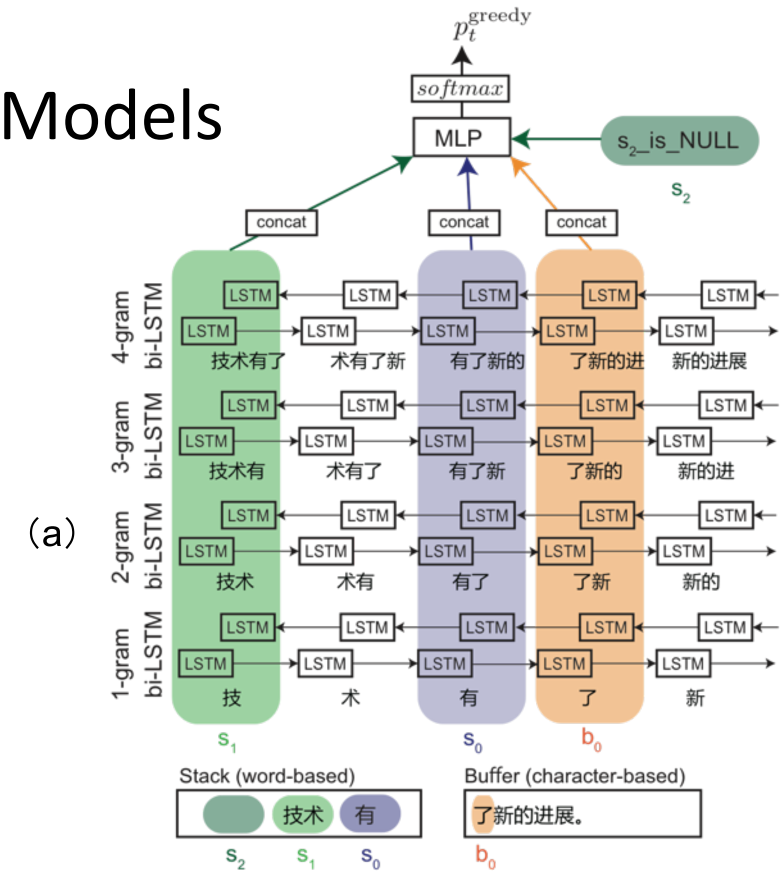
- Models

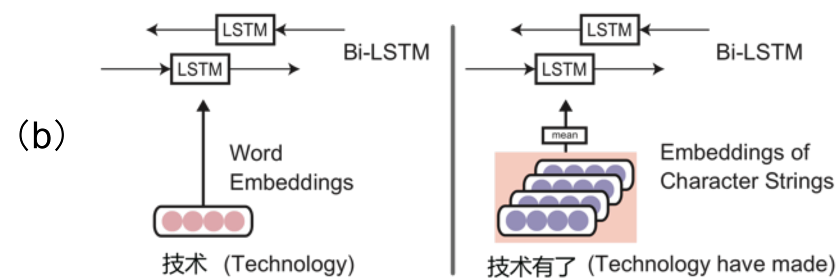


Kurita, Shuhei, Daisuke Kawahara, and Sadao Kurohashi. "Neural Joint Model for Transition-based Chinese Syntactic Analysis." *In ACL* Vol. 1. 2017.

Neural Joint Model

- Models





Kurita, Shuhei, Daisuke Kawahara, and Sadao Kurohashi. "Neural Joint Model for Transition-based Chinese Syntactic Analysis." *In ACL* Vol. 1. 2017.

Neural Joint Model

- Results
  - Joint Segmentation and POS tagging

Model	Seg	POS
Hatori+12 SegTag	97.66	93.61
Hatori+12 SegTag(d)	98.18	94.08
Hatori+12 SegTagDep	97.73	94.46
Hatori+12 SegTagDep(d)	98.26	94.64
M. Zhang+14 EAG	97.76	94.36
Y. Zhang+15	98.04	94.47
SegTag(g)	98.41	**94.84**
SegTag	**98.60**	94.76

Kurita, Shuhei, Daisuke Kawahara, and Sadao Kurohashi. "Neural Joint Model for Transition-based Chinese Syntactic Analysis." *In ACL* Vol. 1. 2017.

Neural Joint Model

- Results
  - Joint Segmentation, POS tagging and Dependency Parsing

Model	Seg	POS	Dep
Hatori+12	97.75	94.33	81.56
M. Zhang+14 EAG	97.76	94.36	**81.70**
SegTagDep(g)	98.24	94.49	80.15
SegTagDep	**98.37**	**94.83**	81.42

Kurita, Shuhei, Daisuke Kawahara, and Sadao Kurohashi. "Neural Joint Model for Transition-based Chinese Syntactic Analysis." *In ACL* Vol. 1. 2017.

Dependency Parsing(with feature set)

- Positional Features

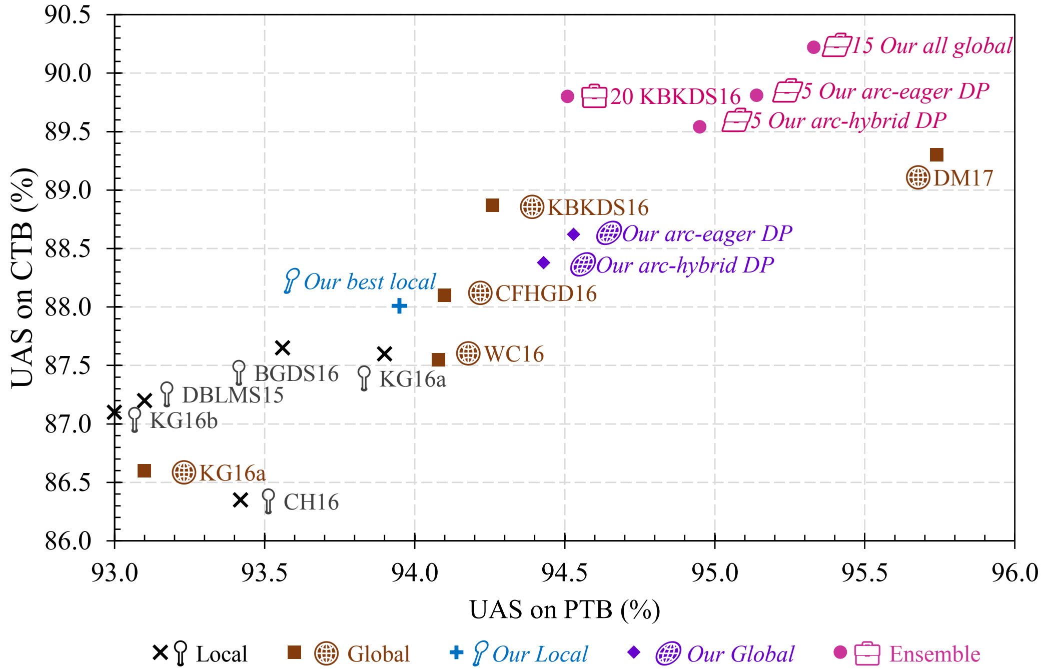
Features	Arc-standard	Arc-hybrid	Arc-eager
$\{\overset{\rightarrow\leftarrow}{s}_2, \overset{\rightarrow\leftarrow}{s}_1, \overset{\rightarrow\leftarrow}{s}_0, \overset{\rightarrow\leftarrow}{b}_0\}$	$93.95_{\pm 0.12}$	$94.08_{\pm 0.13}$	$93.92_{\pm 0.04}$
$\{\overset{\rightarrow\leftarrow}{s}_1, \overset{\rightarrow\leftarrow}{s}_0, \overset{\rightarrow\leftarrow}{b}_0\}$	$94.13_{\pm 0.06}$	$94.08_{\pm 0.05}$	$93.91_{\pm 0.07}$
$\{\overset{\rightarrow\leftarrow}{s}_0, \overset{\rightarrow\leftarrow}{b}_0\}$	$54.47_{\pm 0.36}$	$94.03_{\pm 0.12}$	$93.92_{\pm 0.07}$
$\{\overset{\rightarrow\leftarrow}{b}_0\}$	$47.11_{\pm 0.44}$	$52.39_{\pm 0.23}$	$79.15_{\pm 0.06}$

Min positions	Arc-standard	Arc-hybrid	Arc-eager
K&G 2016a	-	4	-
C&H 2016a	3	-	-
our work	3	**2**	**2**

Shi, Tianze, Liang Huang, and Lillian Lee. "Fast (er) Exact Decoding and Global Training for Transition-Based Dependency Parsing via a Minimal Feature Set." *In EMNLP* (2017).

Dependency Parsing(with feature set)

- Results



Shi, Tianze, Liang Huang, and Lillian Lee. "Fast (er) Exact Decoding and Global Training for Transition-Based Dependency Parsing via a Minimal Feature Set." *In EMNLP* (2017).

Dependency Parsing with exploration

- Parsing Model
  - (1) $$p(z_t \mid \mathbf{p}_t) = \frac{\exp\left(\mathbf{g}_{z_t}^{\top}\mathbf{p}_t + q_{z_t}\right)}{\sum_{z' \in \mathcal{A}(S,B)} \exp\left(\mathbf{g}_{z'}^{\top}\mathbf{p}_t + q_{z'}\right)}$$
  - (2) $$p(\boldsymbol{z} \mid \boldsymbol{w}) = \prod_{t=1}^{|\boldsymbol{z}|} p(z_t \mid \mathbf{p}_t)$$

Miguel Ballesteros, Yoav Goldberg, Chris Dyer, Noah A. Smith. "Training with exploration improves a greedy stack-LSTM parser." *In EMNLP*(2016).

Dependency Parsing with exploration

- Experiments
  - Baseline: Dyer et al.(2015)
  - Dynamic Oracle: for error states, estimate the best tree from the state, using it for oracle
  - Sample negative cases

Chris Dyer, Miguel Ballesteros, Wang Ling, Austin Matthews, Noah A. Smith, Transition-Based Dependency Parsing with Stack Long Short-Term Memory, In Proceedings of the 53rd ACL and the 7th IIJCNLP(2015)
Miguel Ballesteros, Yoav Goldberg, Chris Dyer, Noah A. Smith. "Training with exploration improves a greedy stack-LSTM parser." *In EMNLP*(2016).

Dependency Parsing with exploration

- Results

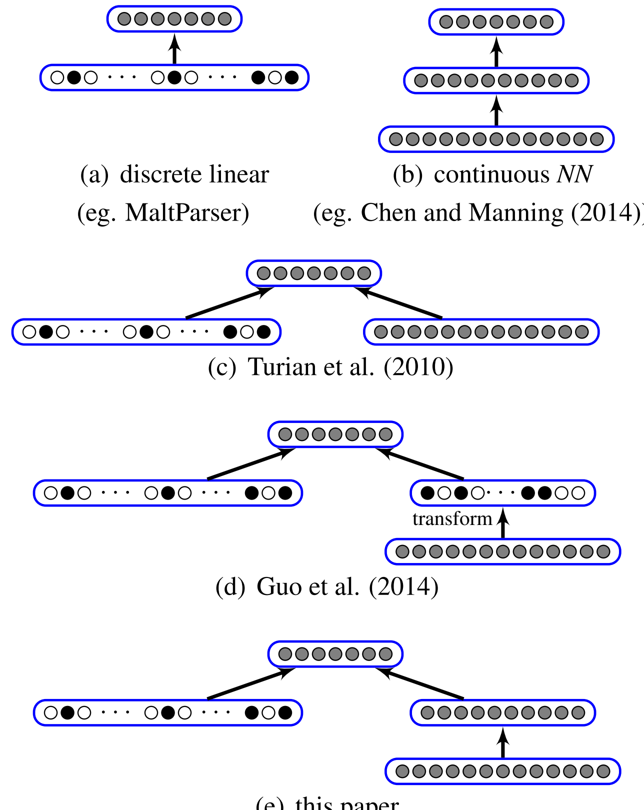
Method	English UAS	English LAS	Chinese UAS	Chinese LAS
Arc-standard (Dyer et al.)	92.40	90.04	85.48	83.94
Arc-hybrid (static)	92.08	89.80	85.66	84.03
Arc-hybrid (dynamic)	92.66	90.43	86.07	84.46
Arc-hybrid (dyn., $\alpha = 0.75$)	92.73	90.60	86.13	84.53
+ pre-training:				
Arc-standard (Dyer et al.)	93.04	90.87	86.85	85.36
Arc-hybrid (static)	92.78	90.67	86.94	85.46
Arc-hybrid (dynamic)	93.15	91.05	87.05	85.63
Arc-hybrid (dyn., $\alpha = 0.75$)	**93.56**	**91.42**	**87.65**	**86.21**

Miguel Ballesteros, Yoav Goldberg, Chris Dyer, Noah A. Smith. "Training with exploration improves a greedy stack-LSTM parser." *In EMNLP*(2016).

Part 5.4: Hybrid Models

Feature Integration

- Model



(a) discrete linear (eg. MaltParser)

(b) continuous *NN* (eg. Chen and Manning (2014))

(c) Turian et al. (2010)

(d) Guo et al. (2014)

(e) this paper

Meishan Zhang and Yue Zhang. *Combining Discrete and Continuous Features for Deterministic Transition-based Dependency Parsing*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Feature Integration

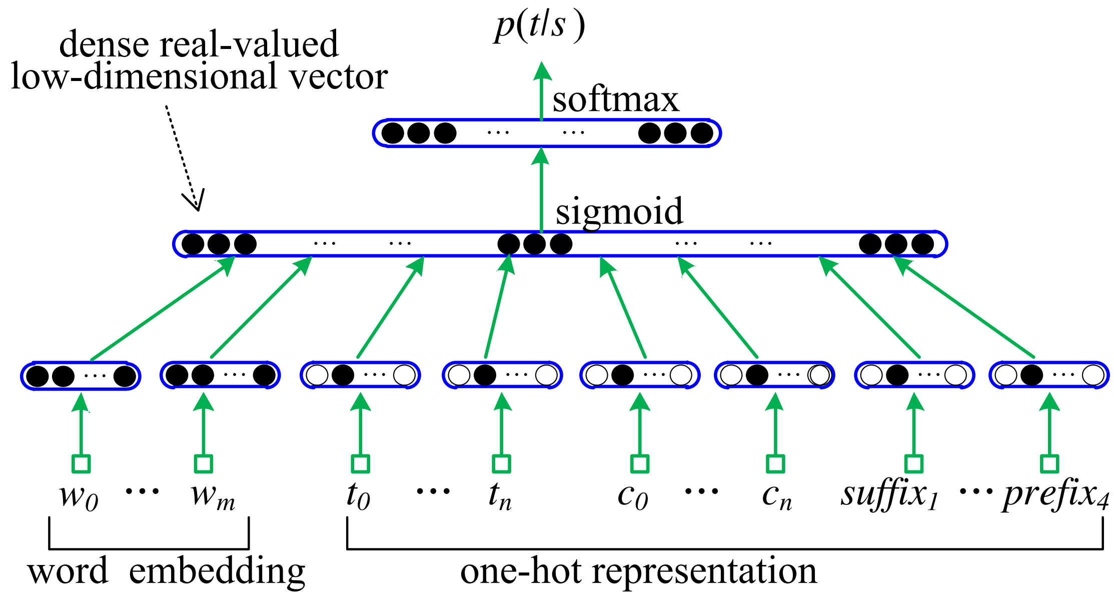
- Results

Domain	#Sent	#Word	TA
WSJ-train	30,060	731,678	97.03
WSJ-dev	1,336	32,092	96.88
WSJ-test	1,640	35,590	97.51
answers	1,744	28,823	91.93
newsgroups	1,195	20,651	93.75
reviews	1,906	28,086	92.66

Meishan Zhang and Yue Zhang. *Combining Discrete and Continuous Features for Deterministic Transition-based Dependency Parsing*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Feature Optimization

- Model
  - Combine discrete
  - Integrate neural



Wang, Zhiguo, Haitao Mi, and Nianwen Xue. "Feature Optimization for Constituent Parsing via Neural Networks." *ACL (1)*. 2015.

Feature Optimization

- Chinese Results

Type	System	F_1
Ours	Supervised*‡	83.2
	Pretrain-Finetune*‡	**86.6**
SI	Petrov and Klein (2007)	83.3
	Wang and Xue (2014)‡	83.6
SE	Zhu et al. (2013)‡	85.6
	Wang and Xue (2014)‡	86.3
RE	Charniak and Johnson (2005)	82.3
	Wang and Zong (2011)	85.7

Wang, Zhiguo, Haitao Mi, and Nianwen Xue. "Feature Optimization for Constituent Parsing via Neural Networks." *ACL (1)*. 2015.

Feature Optimization

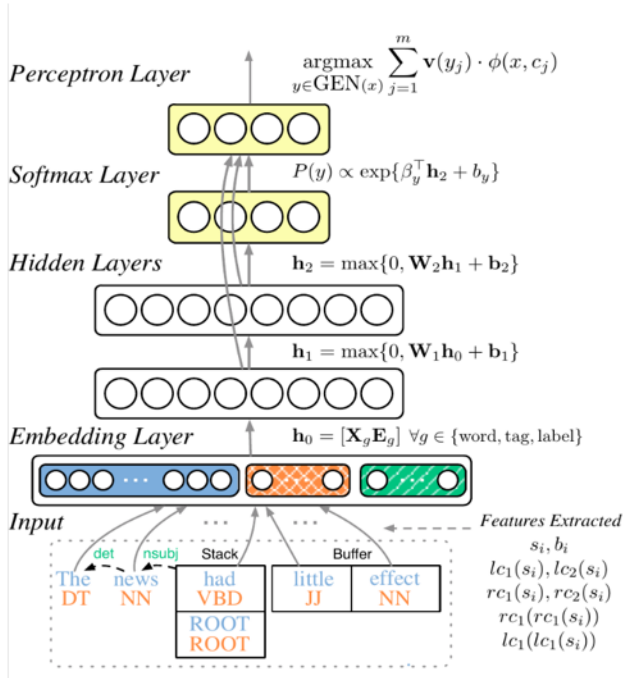
- English Results

Type	System	F_1
Ours	Supervised*‡	83.2
	Pretrain-Finetune*‡	**86.6**
SI	Petrov and Klein (2007)	83.3
	Wang and Xue (2014)‡	83.6
SE	Zhu et al. (2013)‡	85.6
	Wang and Xue (2014)‡	86.3
RE	Charniak and Johnson (2005)	82.3
	Wang and Zong (2011)	85.7

Wang, Zhiguo, Haitao Mi, and Nianwen Xue. "Feature Optimization for Constituent Parsing via Neural Networks." *ACL (1)*. 2015.

Google Hybrid Model

- Dependency parsing



Weiss, D., Alberti, C., Collins, M., & Petrov, S. (2015). Structured Training for Neural Network Transition-Based Parsing. ACL.

Google Hybrid Model

- Using Chen and Manning features for perceptron training
- Back-propagation pre-training

$$L(\Theta) = -\sum_j \log P(y_j \mid c_j, \Theta) + \lambda \sum_i \|\mathbf{W}_i\|_2^2$$

- Structured perceptron training

$$(h_1, h_2,\ P(y))$$

Weiss, D., Alberti, C., Collins, M., & Petrov, S. (2015). Structured Training for Neural Network Transition-Based Parsing. ACL.

Google Hybrid Model

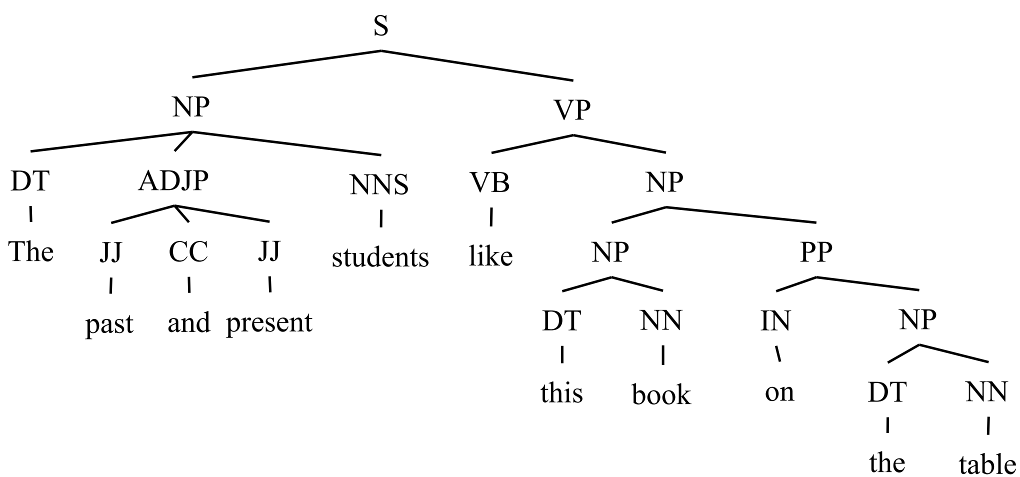
- Results

Method	UAS	LAS	Beam
Graph-based			
Bohnet (2010)	92.88	90.71	n/a
Martins et al. (2013)	92.89	90.55	n/a
Zhang and McDonald (2014)	93.22	91.02	n/a
Transition-based			
⋆Zhang and Nivre (2011)	93.00	90.95	32
Bohnet and Kuhn (2012)	93.27	91.19	40
Chen and Manning (2014)	91.80	89.60	1
S-LSTM (Dyer et al., 2015)	93.20	90.90	1
Our Greedy	93.19	91.18	1
Our Perceptron	**93.99**	**92.05**	8
Tri-training			
⋆Zhang and Nivre (2011)	92.92	90.88	32
Our Greedy	93.46	91.49	1
Our Perceptron	**94.26**	**92.41**	8

Weiss, D., Alberti, C., Collins, M., & Petrov, S. (2015). Structured Training for Neural Network Transition-Based Parsing. ACL.

Look-Ahead Features

- Model



(a)

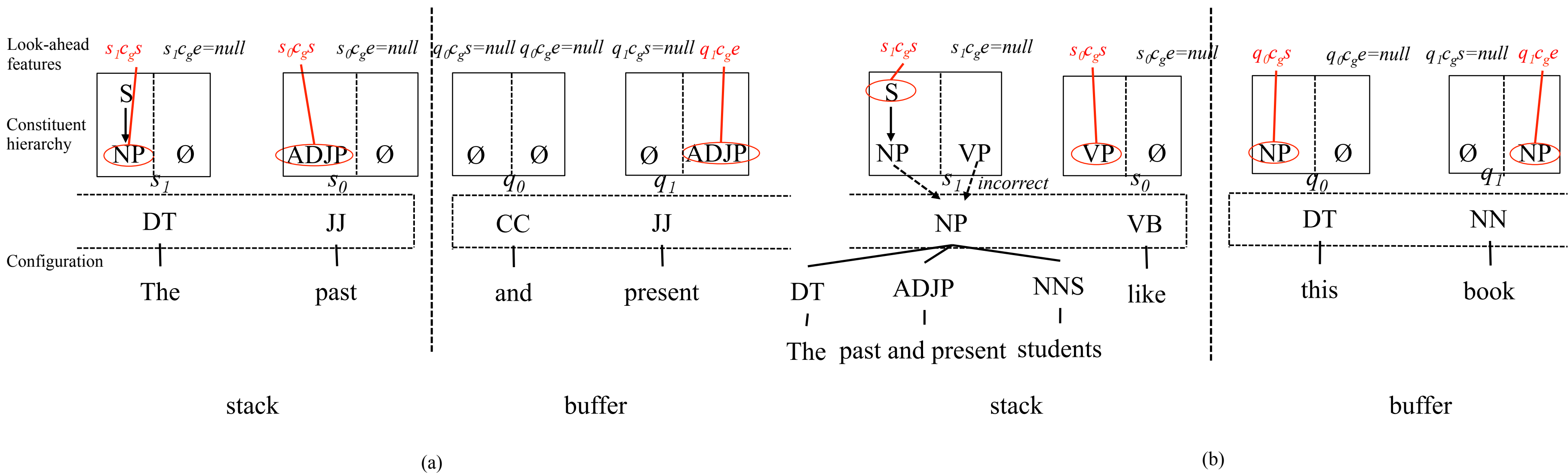
Word	***s*-type**	***e*-type**
The	[*s*: *S*→*NP*]	[*e*: Ø]
past	[*s*: *ADJP*]	[*e*: Ø]
and	[*s*: Ø]	[*e*: Ø]
present	[*s*: Ø]	[*e*: *ADJP*]
students	[*s*: Ø]	[*e*: *NP*]
like	[*s*: *VP*]	[*e*: Ø]
this	[*s*: *NP*→*NP*]	[*e*: Ø]
book	[*s*: Ø]	[*e*: *NP*]
on	[*s*: *PP*]	[*e*: Ø]
the	[*s*: *NP*]	[*e*: Ø]
table	[*s*: Ø	[*e*: *S*→*VP*→*NP*→*PP*→*NP*]

(b)

Jiangming Liu, Yue Zhang. 2017. Shift-Reduce Constituent Parsing with Neural Lookahead Features. In Transactions of the Association of Computational Linguistics (TACL)

Look-Ahead Features

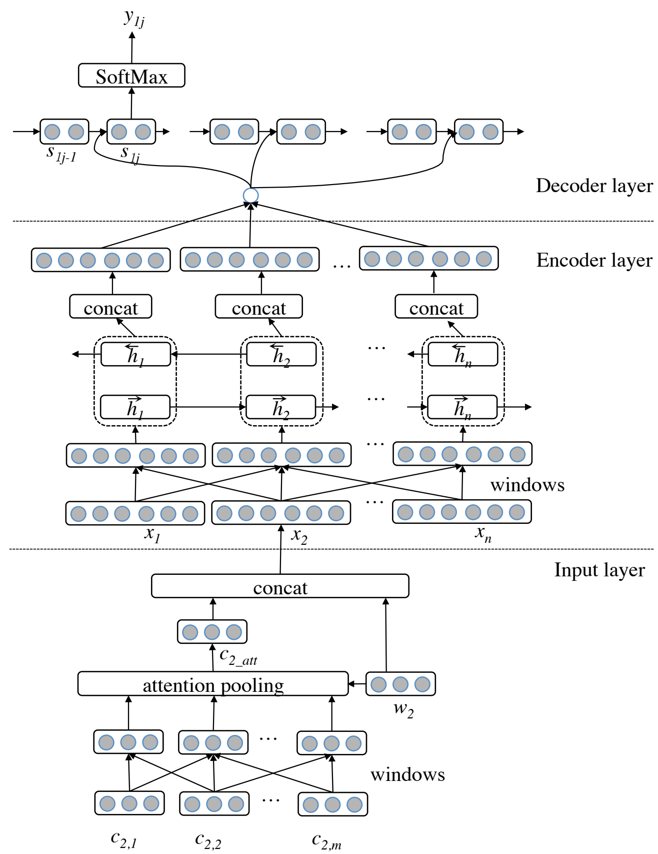
- Model



Jiangming Liu, Yue Zhang. 2017. Shift-Reduce Constituent Parsing with Neural Lookahead Features. In Transactions of the Association of Computational Linguistics (TACL)

Look-Ahead Features

- Model



Jiangming Liu, Yue Zhang. 2017. Shift-Reduce Constituent Parsing with Neural Lookahead Features. In Transactions of the Association of Computational Linguistics (TACL)

Look-Ahead Features

- Results

Parser	LR	LP	F_1
Fully-supervised			
Ratnaparkhi (1997)	86.3	87.5	86.9
Charniak (2000)	89.5	89.9	89.5
Collins (2003)	88.1	88.3	88.2
Sagae and Lavie (2005)†	86.1	86.0	86.0
Sagae and Lavie (2006)†	87.8	88.1	87.9
Petrov and Klein (2007)	90.1	90.2	90.1
Carreras et al. (2008)	90.7	91.4	91.1
Shindo et al. (2012)	N/A	N/A	91.1
Zhu et al. (2013)†	90.2	90.7	90.4
Socher et al. (2013)*	N/A	N/A	90.4
Vinyals et al. (2015)*	N/A	N/A	88.3
This work	**91.3**	**92.1**	**91.7**

Ensemble			
Shindo et al. (2012)	N/A	N/A	92.4
Vinyals et al. (2015)*	N/A	N/A	90.5
Rerank			
Charniak and Johnson (2005)	91.2	91.8	91.5
Huang (2008)	92.2	91.2	91.7
Semi-supervised			
McClosky et al. (2006)	92.1	92.5	92.3
Huang and Harper (2009)	91.1	91.6	91.3
Huang et al. (2010)	91.4	91.8	91.6
Zhu et al. (2013)†	91.1	91.5	91.3
Durrett and Klein (2015)*	N/A	N/A	91.1
Dyer et al. (2016)*†	N/A	N/A	92.4

Jiangming Liu, Yue Zhang. 2017. Shift-Reduce Constituent Parsing with Neural Lookahead Features. In Transactions of the Association of Computational Linguistics (TACL)

Look-Ahead Features

- Results

Parser	LR	LP	F_1
Fully-supervised			
Charniak (2000)	79.6	82.1	80.8
Bikel (2004)	79.3	82.0	80.6
Petrov and Klein (2007)	81.9	84.8	83.3
Zhu et al. (2013)†	82.1	84.3	83.2
Wang et al. (2015)‡	N/A	N/A	83.2
This work	**85.2**	**85.9**	**85.5**
Rerank			
Charniak and Johnson (2005)	80.8	83.8	82.3
Semi-supervised			
Zhu et al. (2013)†	84.4	86.8	85.6
Wand and Xue (2014)‡	N/A	N/A	86.3
Wang et al. (2015)‡	N/A	N/A	86.6
Dyer et al. (2016)*†	N/A	N/A	82.7

Jiangming Liu, Yue Zhang. 2017. Shift-Reduce Constituent Parsing with Neural Lookahead Features. In Transactions of the Association of Computational Linguistics (TACL)

Look-Ahead Features

- Results

Parser	#Sent/Second
Ratnaparkhi (1997)	Unk
Collins (2003)	3.5
Charniak (2000)	5.7
Sagae and Lavie (2005)	3.7
Sagae and Lavie (2006)	2.2
Petrov and Klein (2007)	6.2
Carreras et al. (2008)	Unk
Zhu et al. (2013)	89.5
This work	79.2

Jiangming Liu, Yue Zhang. 2017. Shift-Reduce Constituent Parsing with Neural Lookahead Features. In Transactions of the Association of Computational Linguistics (TACL)