Deep Learning and Lexical, Syntactic and Semantic Analysis

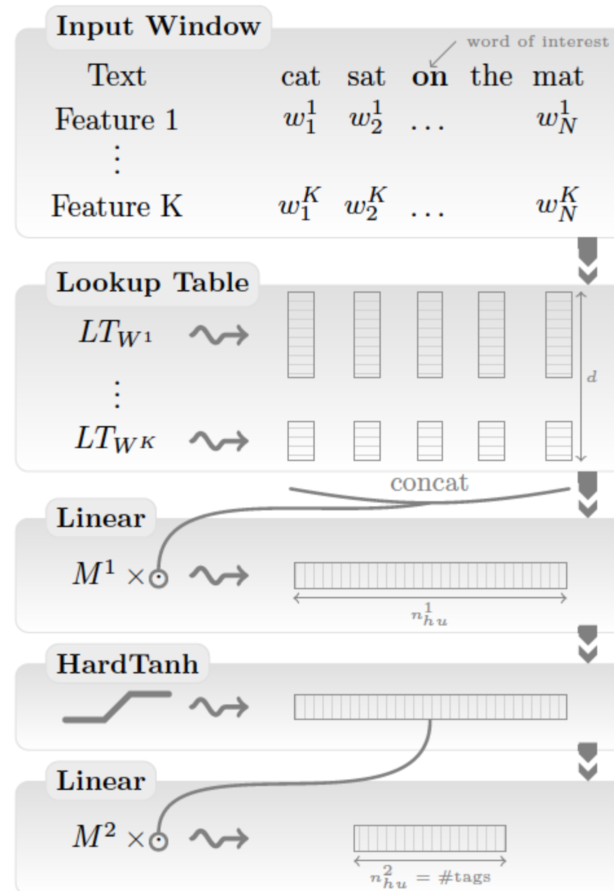
Wanxiang Che (HIT)
Yue Zhang (SUTD)

Part 3: Greedy Decoding

Part 3.1: Greedy Decoding for Tagging

Window Approach

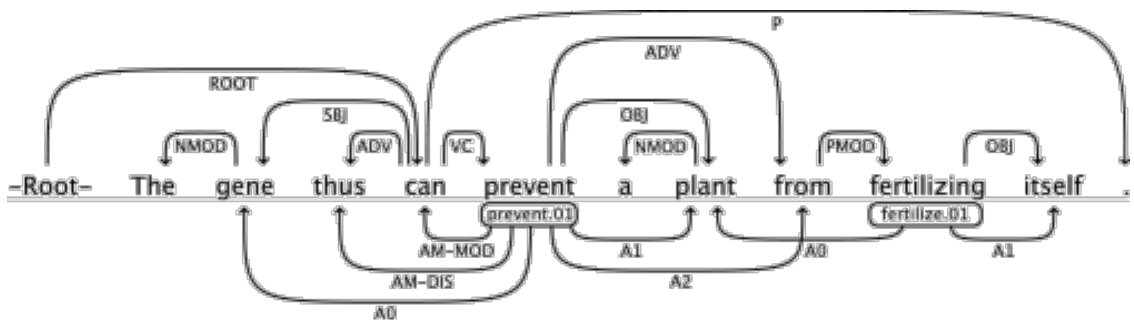
- Tasks
  - POS tagging, Chunking, NER, SRL
- Tag **one word** at a time
- Feed a **fixed-size** window of text around **each word** to tag
- Features
  - Words, POS tags, Suffix, Cascading, …



Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural Language Processing (Almost) from Scratch. J. Mach. Learn. Res. 12, 2493-2537.

Window Approach

- Works fine for most tasks
- How to deal with long-range dependencies?
  - E.g. in SRL, the verb of interest might be outside the window!



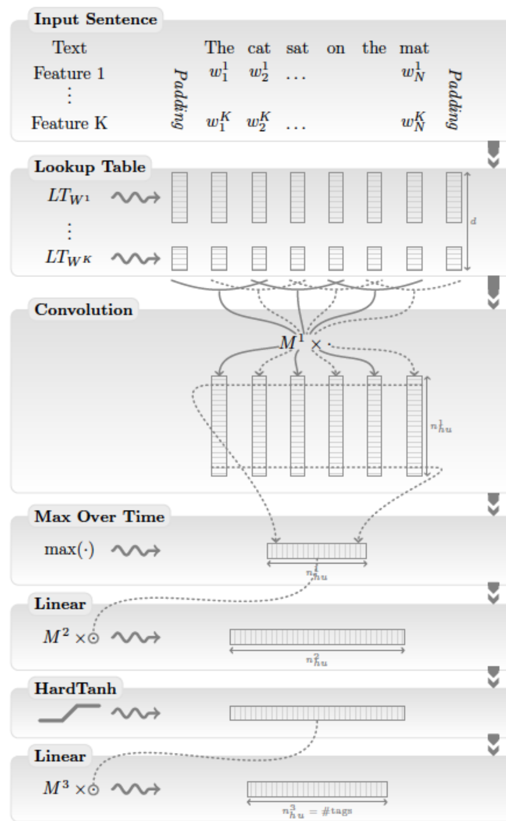
Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural Language Processing (Almost) from Scratch. J. Mach. Learn. Res. 12, 2493-2537.

Sentence Approach

- Tag one word at a time
  - add extra **relative position** features
- Feed the **whole sentence** to the network
- Convolutions to handle variable-length inputs
- **Max over** time to capture most relevant features
  - Outputs a fixed-sized feature vector

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural Language Processing (Almost) from Scratch. J. Mach. Learn. Res. 12, 2493-2537.

Sentence Approach



Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural Language Processing (Almost) from Scratch. J. Mach. Learn. Res. 12, 2493-2537.

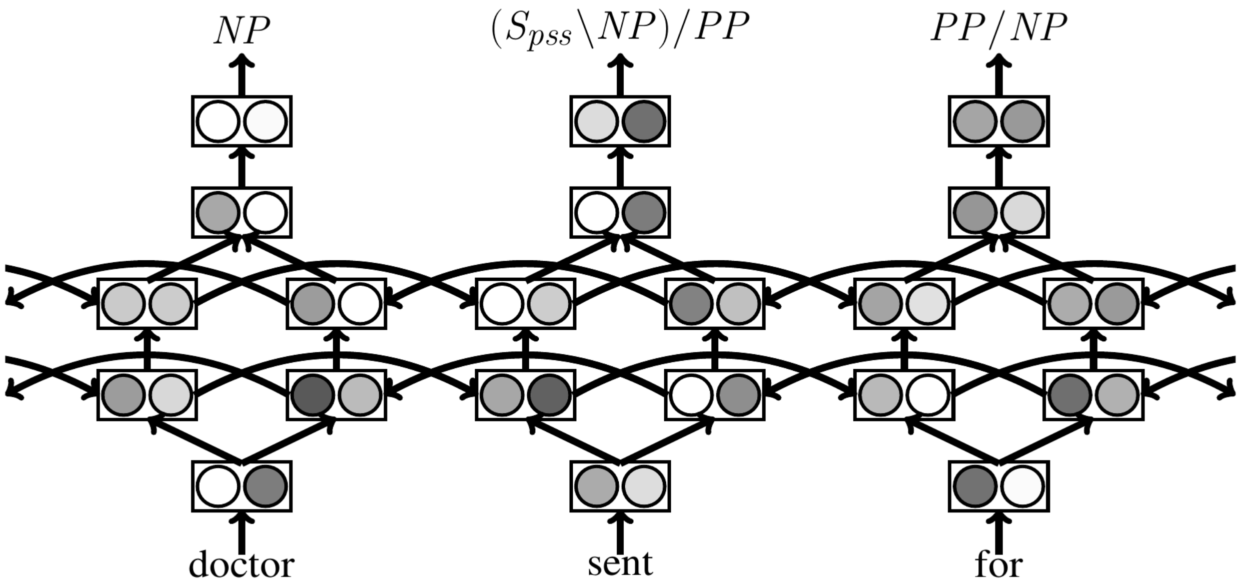
Results

Approach	POS (PWA)	Chunking (F1)	NER (F1)	SRL (F1)
Benchmark Systems	97.24	94.29	89.31	77.92
NN+WLL	96.31	89.13	79.53	55.40

- Window approach: POS, Chunking, NER
- Sentence approach: SRL
- WLL: Word-Level Log-Likelihood

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural Language Processing (Almost) from Scratch. J. Mach. Learn. Res. 12, 2493-2537.

CCG Supertagging



Lewis, M., & Steedman, M. (2014). Improved CCG Parsing with Semi-supervised Supertagging. TACL.
Xu, W., Auli, M., & Clark, S. (2015). CCG Supertagging with a Recurrent Neural Network. ACL.
Lewis, M., Lee, Kenton., & Zettlemoyer, L. (2016). LSTM CCG Parsing. NAACL.

CCG Supertagging

- No POS feature
- Reduce word sparsity
- Global context information

Model	Accuracy	Time
C&C (gold POS)	92.60	-
C&C (auto POS)	91.50	0.57
NN	91.10	21.00
RNN	92.63	-
RNN+dropout	93.07	2.02

Model	**Dev**	**Test**
C&C tagger	91.5	92.0
NN	91.3	91.6
RNN	93.1	93.0
LSTM	94.1	94.3
LSTM + Tri-training	**94.9**	**94.7**

Lewis, M., & Steedman, M. (2014). Improved CCG Parsing with Semi-supervised Supertagging. TACL.
Xu, W., Auli, M., & Clark, S. (2015). CCG Supertagging with a Recurrent Neural Network. ACL.
Lewis, M., Lee, Kenton., & Zettlemoyer, L. (2016). LSTM CCG Parsing. NAACL.

CCG Supertagging

Supertagger	Accuracy
Bidirectional RNNs	93.4
Forward LSTM only	83.5
Backward LSTM only	89.5
Bidirectional LSTMs	**94.1**

Word Class	NN	LSTM	LSTM+ Tri-training
All	91.32	94.14	**94.90**
Unseen Words	90.39	94.21	**95.26**
Unseen Usages	45.80	59.37	**62.46**
Prepositions	78.11	84.40	**85.98**
Verbs	82.55	87.85	**89.24**
Wh-words	90.47	92.09	**94.16**
Long range	74.80	83.99	**86.31**

Lewis, M., & Steedman, M. (2014). Improved CCG Parsing with Semi-supervised Supertagging. TACL.
Xu, W., Auli, M., & Clark, S. (2015). CCG Supertagging with a Recurrent Neural Network. ACL.
Lewis, M., Lee, Kenton., & Zettlemoyer, L. (2016). LSTM CCG Parsing. NAACL.

CCG Supertagging

- Parsing results

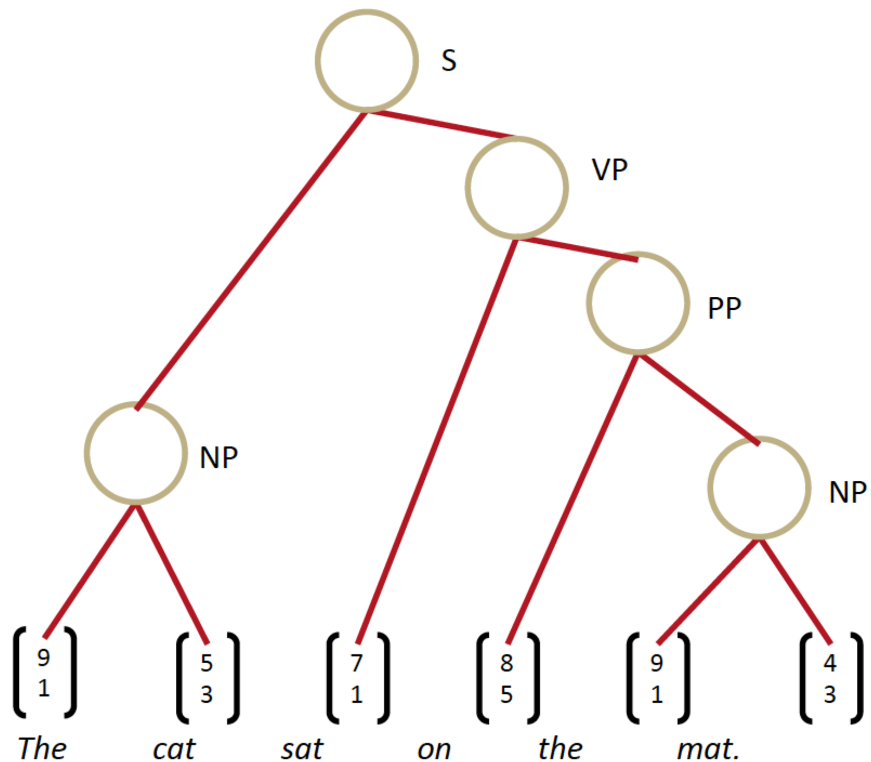
Model	P	R	F1
C&C	86.2	84.2	85.2
C&C + RNN	87.7	86.4	87.0
EASYCCG	83.7	83.0	83.3
Dependencies	86.5	85.8	86.1
LSTM	87.7	86.7	87.2
LSTM + Dependencies	88.2	87.3	87.8
LSTM + Tri-training	**88.6**	**87.5**	**88.1**
LSTM + Tri-training + Dependencies	88.2	87.3	87.8

Lewis, M., & Steedman, M. (2014). Improved CCG Parsing with Semi-supervised Supertagging. TACL.
Xu, W., Auli, M., & Clark, S. (2015). CCG Supertagging with a Recurrent Neural Network. ACL.
Lewis, M., Lee, Kenton., & Zettlemoyer, L. (2016). LSTM CCG Parsing. NAACL.

Part 3.2: Greedy Search for Constituent Parsing with RNN

Constituent Parsing with Recursive NN

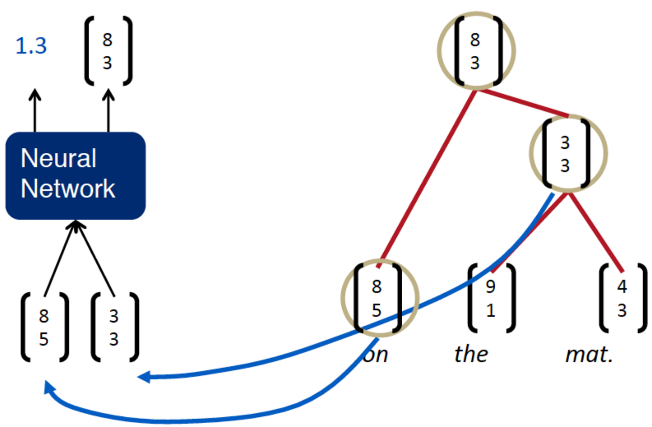
- Our goal



Richard Socher, Cliff Lin, Andrew Y. Ng, and Christopher D. Manning. Parsing Natural Scenes and Natural Language with Recursive Neural Networks. ICML 2011

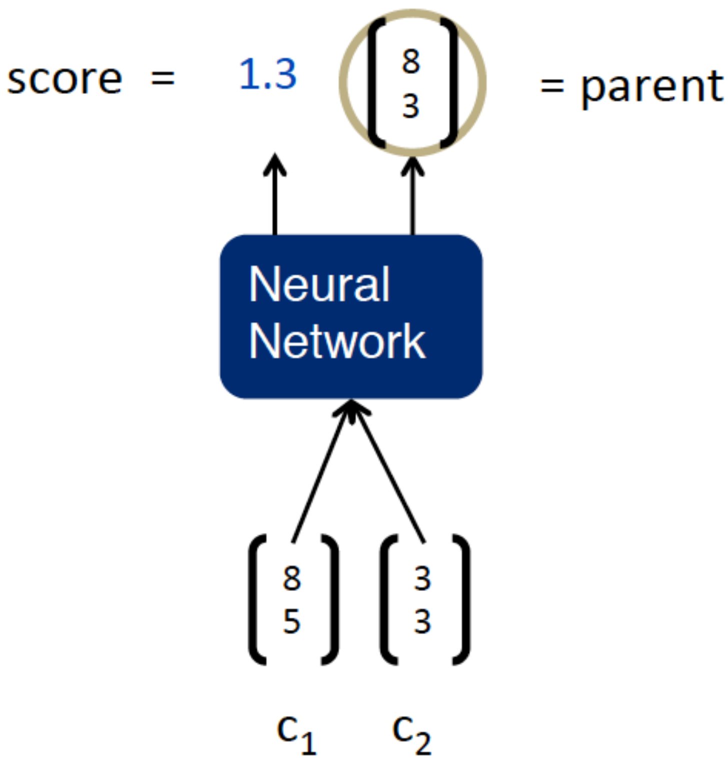
Recursive NN

- Inputs
  - Two candidate children's representations
- Outputs
  - The semantic representation if the two nodes are merged
  - Score of how plausible the new node would be



Richard Socher, Cliff Lin, Andrew Y. Ng, and Christopher D. Manning. Parsing Natural Scenes and Natural Language with Recursive Neural Networks. ICML 2011

RNN Definition

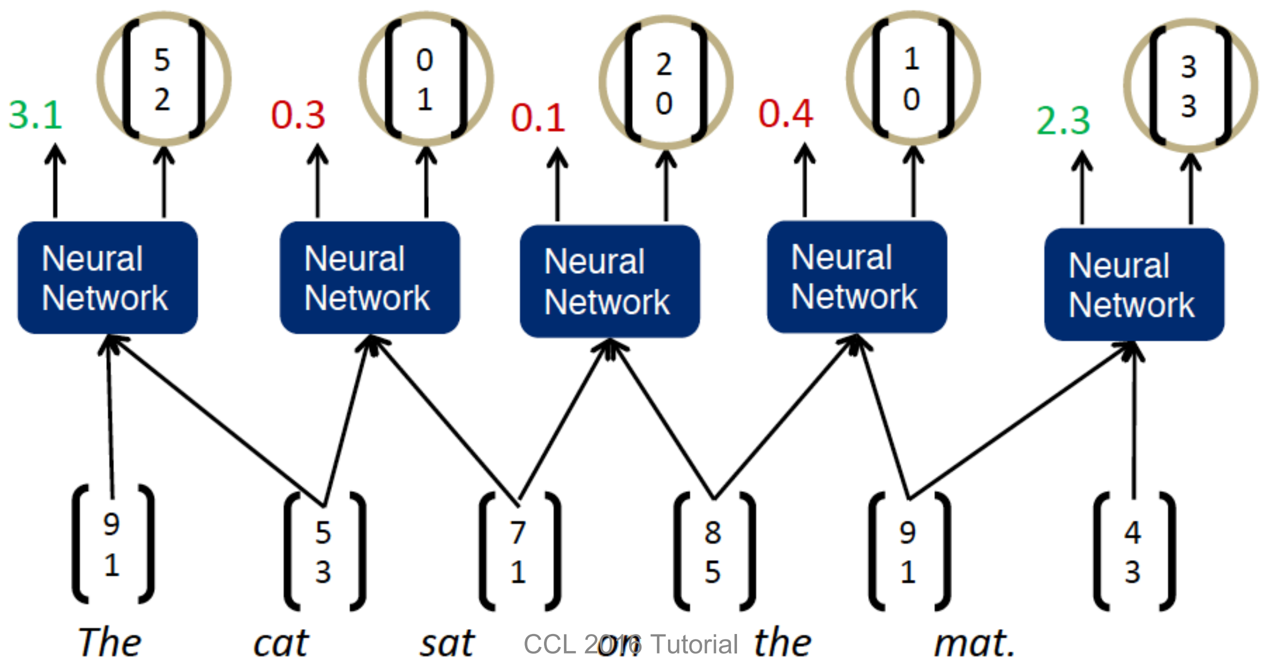


$$\begin{cases} \text{score} = U^T p \\ p = \tanh\left(W \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + b\right) \end{cases}$$

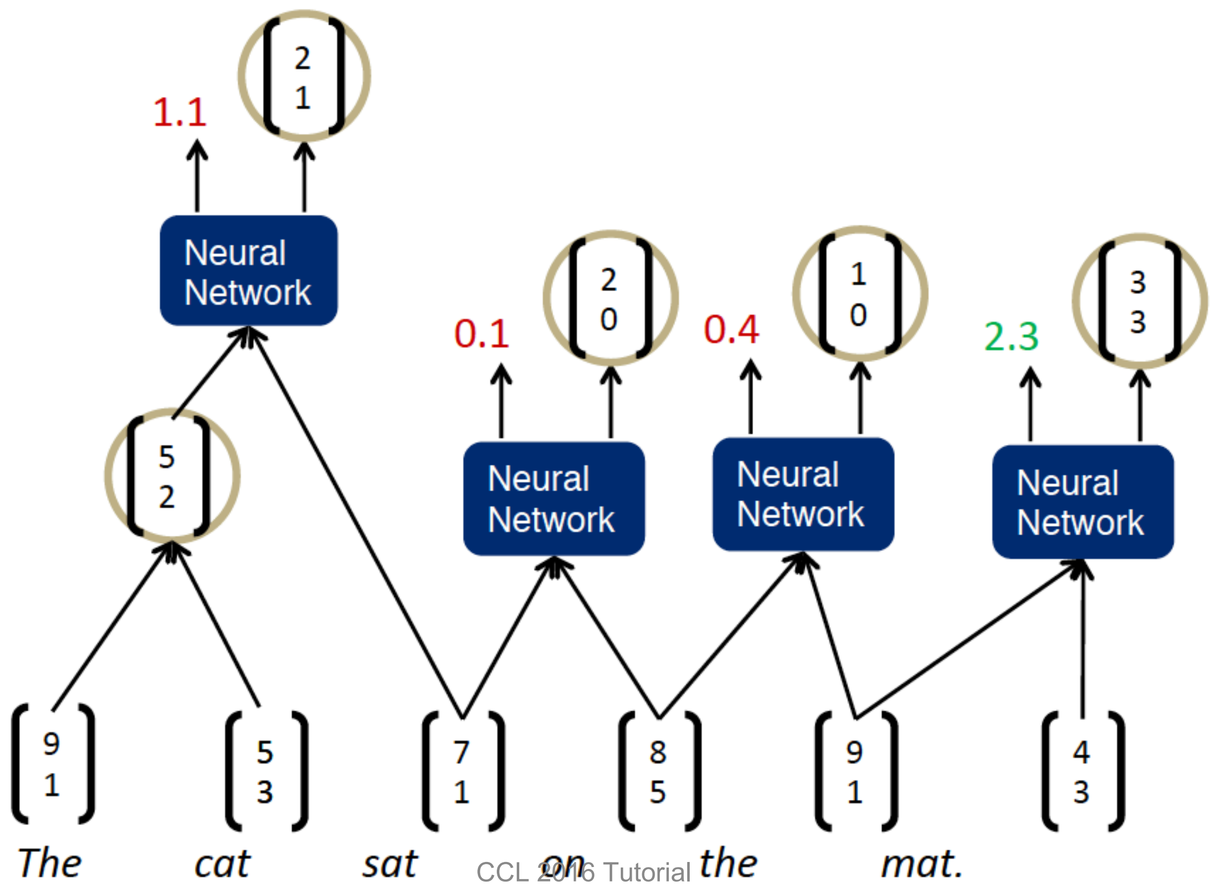
where *W* at all nodes of the tree are the same

Richard Socher, Cliff Lin, Andrew Y. Ng, and Christopher D. Manning. Parsing Natural Scenes and Natural Language with Recursive Neural Networks. ICML 2011

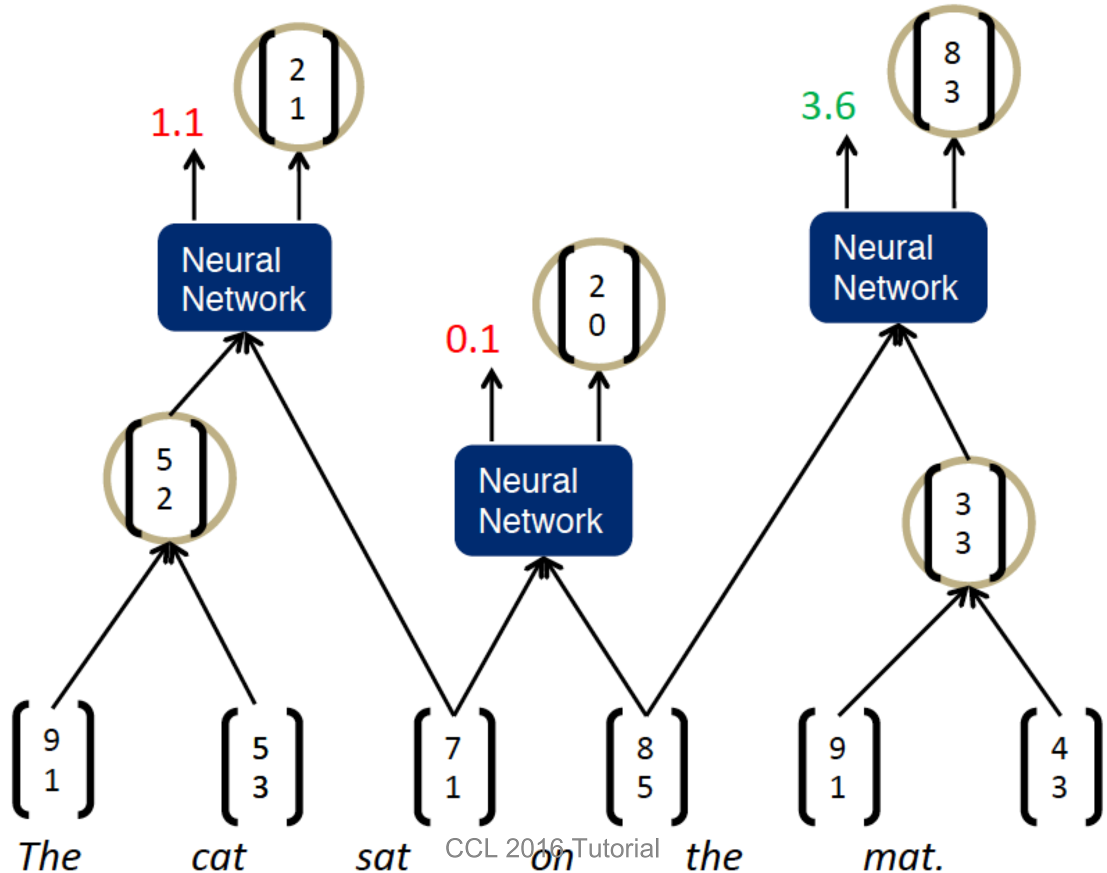
Parsing a sentence with an RNN



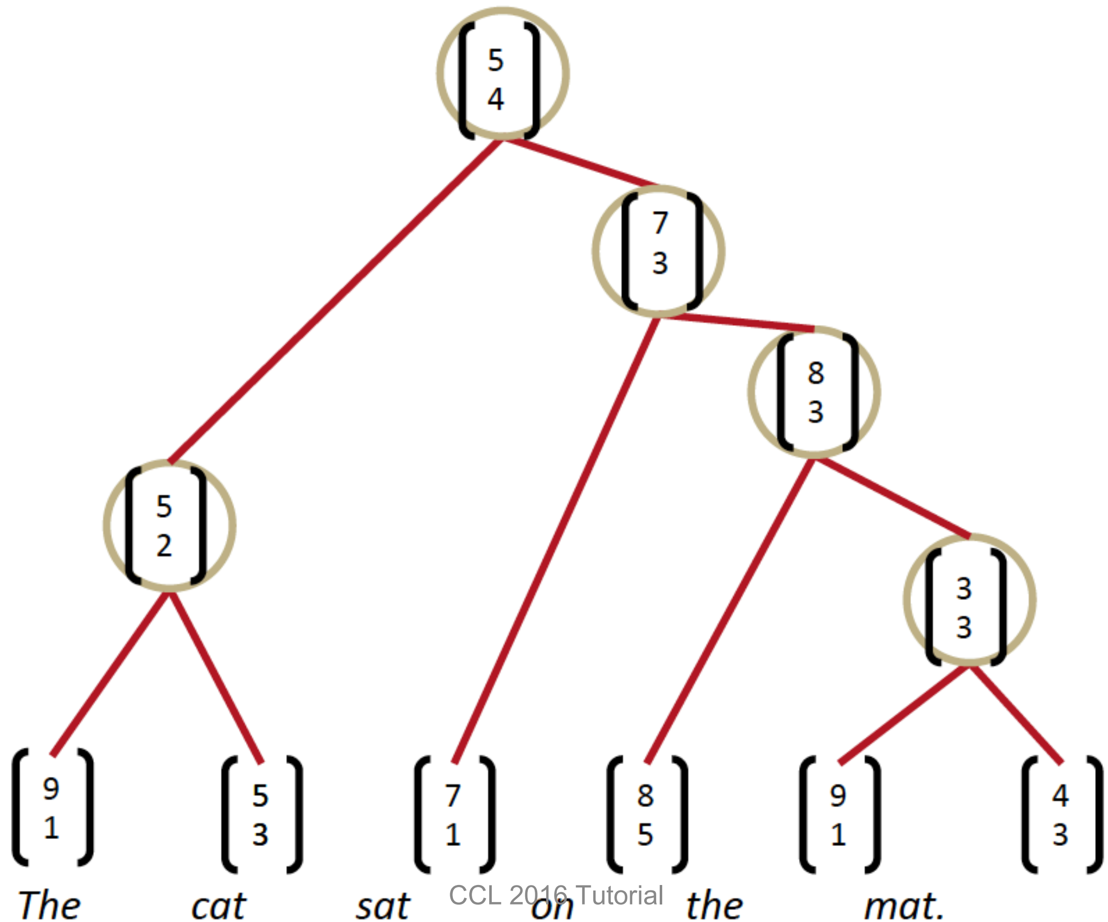
Parsing a sentence with an RNN



Parsing a sentence with an RNN

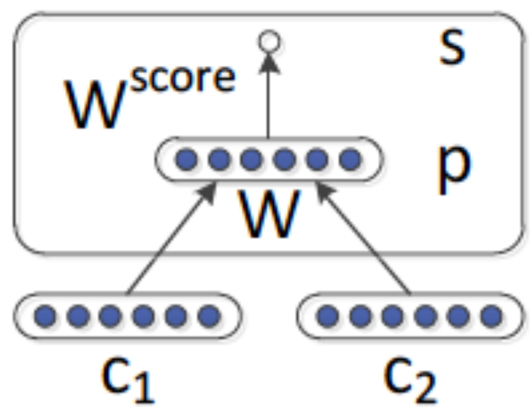


Parsing a sentence with an RNN



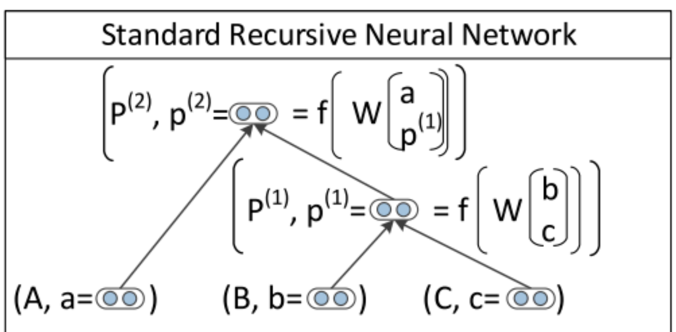
Discussion on Simple RNN

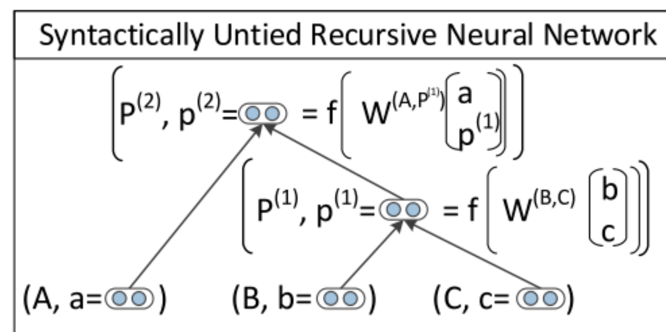
- The composition function with single weight matrix is the same for all categories, punctuation, etc.
- It could capture some phenomena but not adequate for more complex, higher order composition



Solution: Syntactically-Untied RNN (SU-RNN)

- Intuition
  - Condition the composition function on the syntactic categories
- Allows for different composition functions for pairs of syntactic categories, e.g. Adv + AdjP, VP + NP

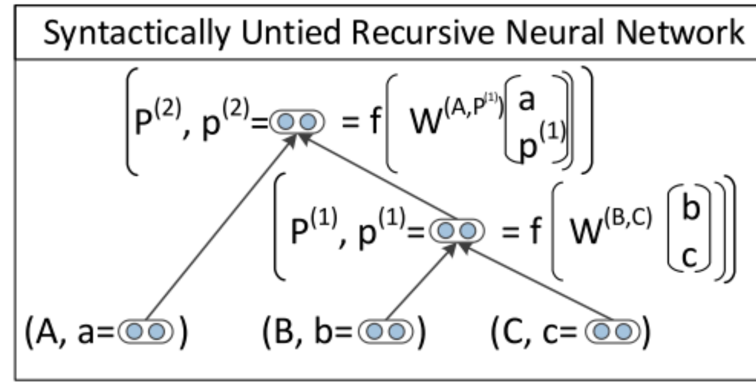




Richard Socher, John Bauer, Christopher D. Manning and Andrew Y. Ng. Parsing with Compositional Vector Grammars. ACL 2013.

Compositional Vector Grammars (CVG)

- PCFG + SU-RNN
- PCFG
  - Produce: k-best parsing trees
- SU-RNN
  - Re-ranking with SU-RNN



$$p^{(1)} = f\left(W^{(B,C)}\begin{bmatrix} b \\ c \end{bmatrix}\right)$$

$$s\left(p^{(1)}\right) = \left(v^{(B,C)}\right)^T p^{(1)} + \log P(P_1 \rightarrow B\ \ C)$$

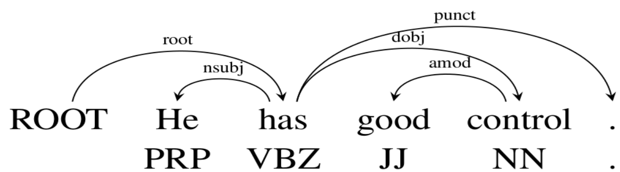
$$s(\texttt{CVG}(\theta, x, \hat{y})) = \sum_{d \in N(\hat{y})} s\left(p^d\right)$$

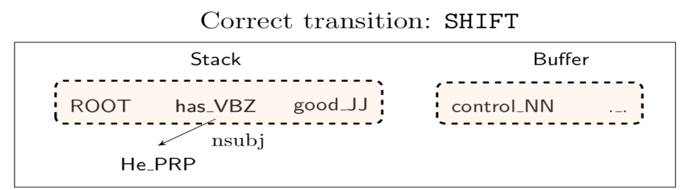
Richard Socher, John Bauer, Christopher D. Manning and Andrew Y. Ng. Parsing with Compositional Vector Grammars. ACL 2013.

Part 3.3: Transition-based Dependency Parsing with Greedy Search

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

Softmax layer:

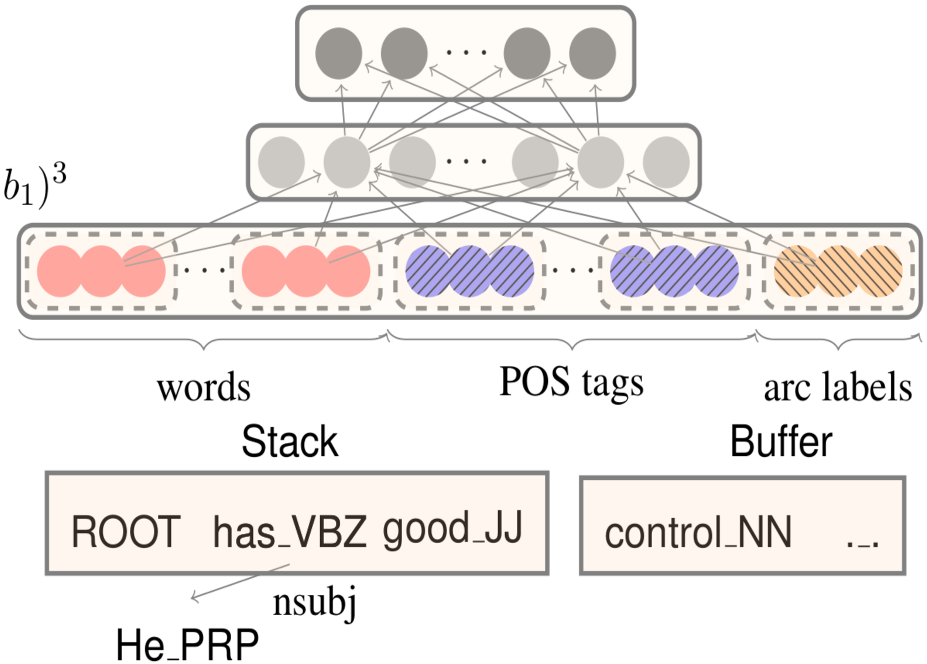
$$p = \texttt{softmax}(W_2 h)$$

Hidden layer:

$$h = (W_1^w x^w + W_1^t x^t + W_1^l x^l + b_1)^3$$

Input layer: $[x^w, x^t, x^l]$

Configuration



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

Dependency Parsing

- ZPar features (Zhang and Nivre, ACL 2011)

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

Dependency Parsing

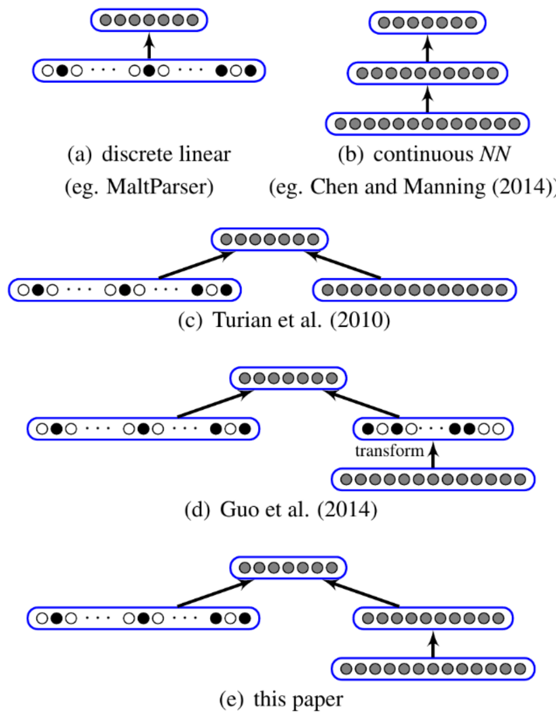
Parser	Dev		Test		Speed
	UAS	LAS	UAS	LAS	(sent/s)
standard	90.2	87.8	89.4	87.3	26
eager	89.8	87.4	89.6	87.4	34
Malt:sp	89.8	87.2	89.3	86.9	469
Malt:eager	89.6	86.9	89.4	86.8	448
MSTParser	91.4	88.1	90.7	87.6	10
Our parser	**92.0**	**89.7**	**91.8**	**89.6**	**654**

Parser	Dev		Test		Speed
	UAS	LAS	UAS	LAS	(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**

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

Dependency Parsing

- Chen and Manning with combined features



Zhang, M., & Zhang, Y. (2015). Combining Discrete and Continuous Features for Deterministic Transition-based Dependency Parsing. EMNLP.

Dependency Parsing

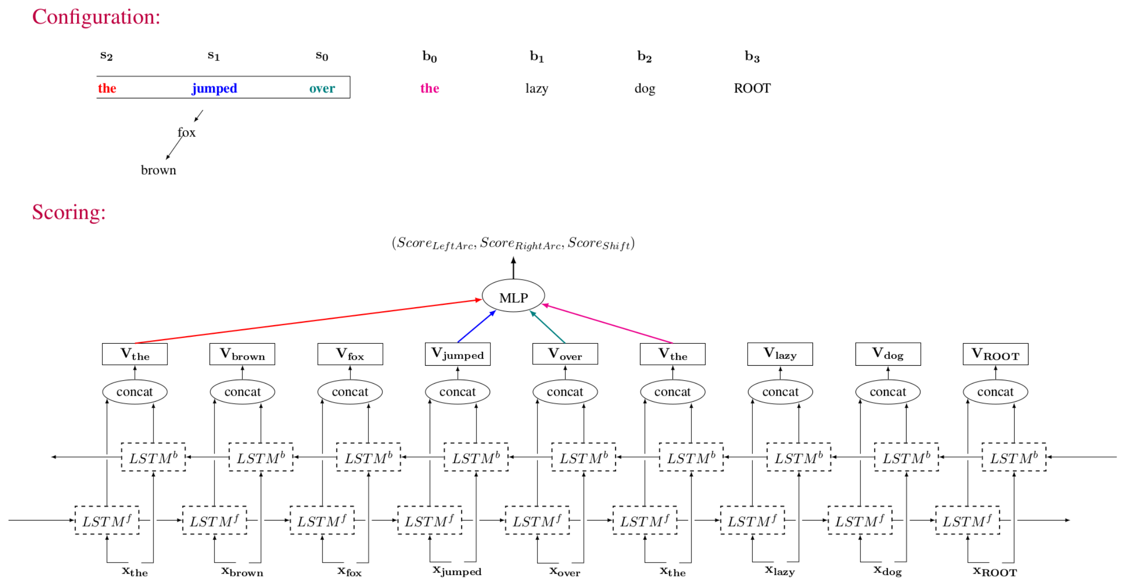
- Chen and Manning with combined features

System	UAS	LAS
L	89.36	88.33
NN	91.15	90.04
This	**91.80**	**90.68**
ZPar-local	89.94	88.92
Ma et al. (2014a)	90.38	–
Chen and Manning (2014)	91.17	89.99
Honnibal et al. (2013)	91.30	90.00
Ma et al. (2014a)*	91.32	–

Zhang, M., & Zhang, Y. (2015). Combining Discrete and Continuous Features for Deterministic Transition-based Dependency Parsing. EMNLP.

Dependency Parsing

- Chen and Manning with richer features



- 11 more LSTMs

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

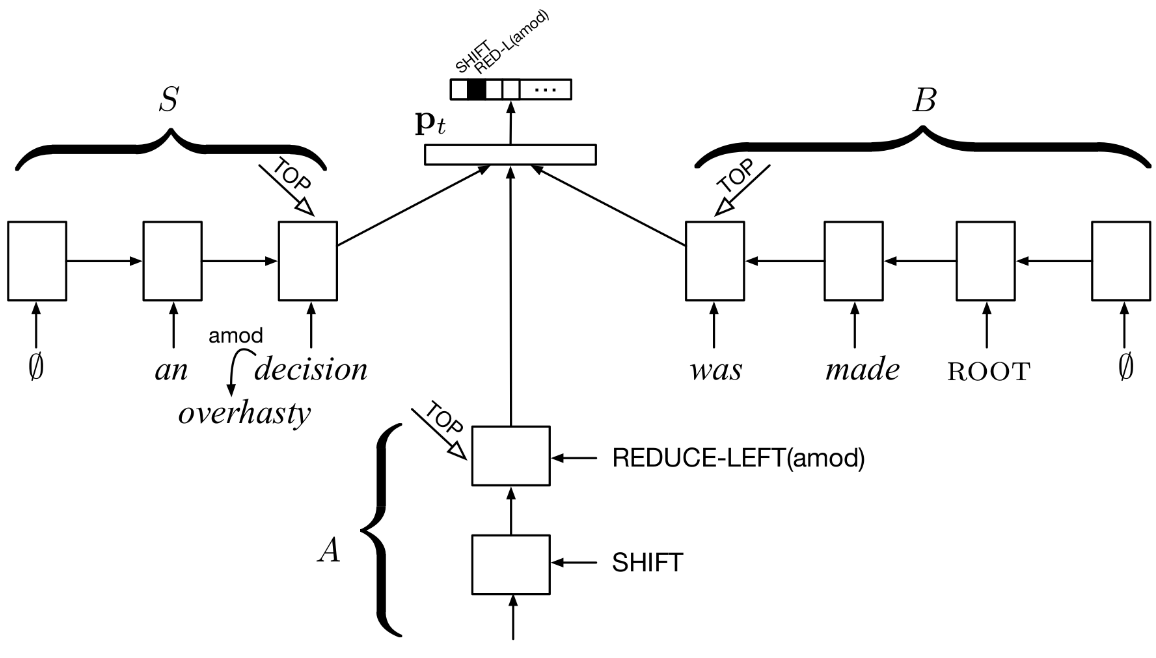
Dependency Parsing

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.

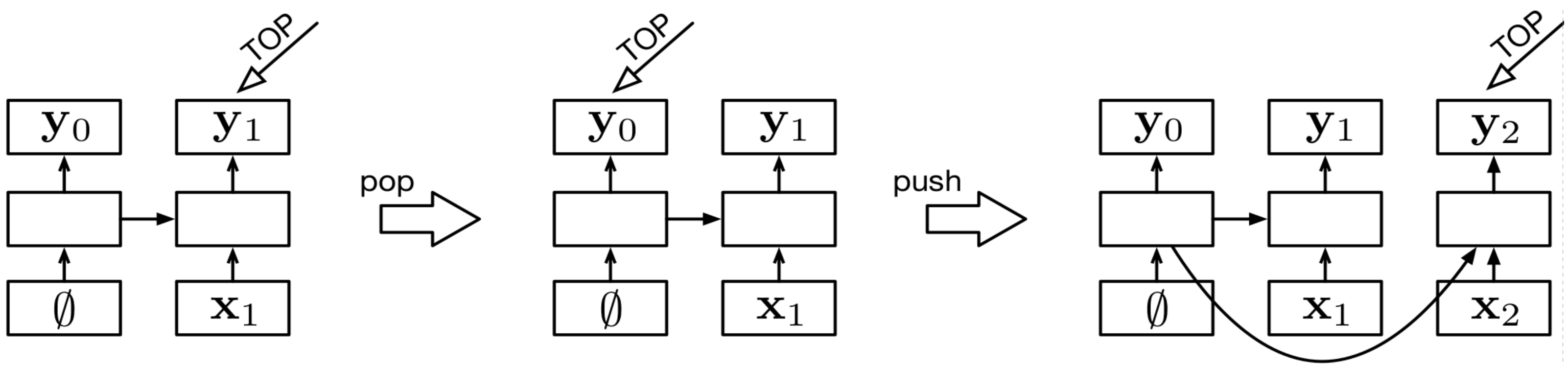
Dependency Parsing

- Chen and Manning with less features



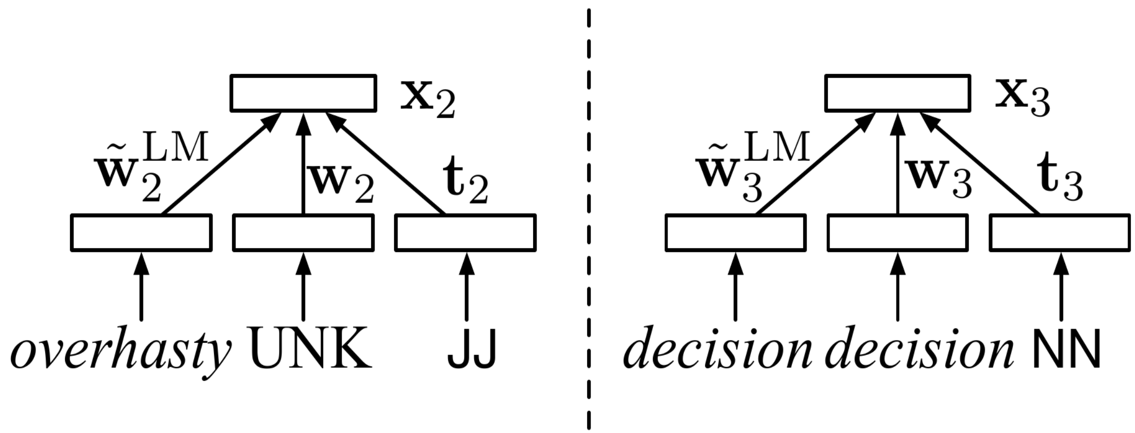
Dyer, C., Ballesteros, M., Ling, W., Matthews, A., & Smith, N. A. (2015). Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ACL.

Dependency Parsing



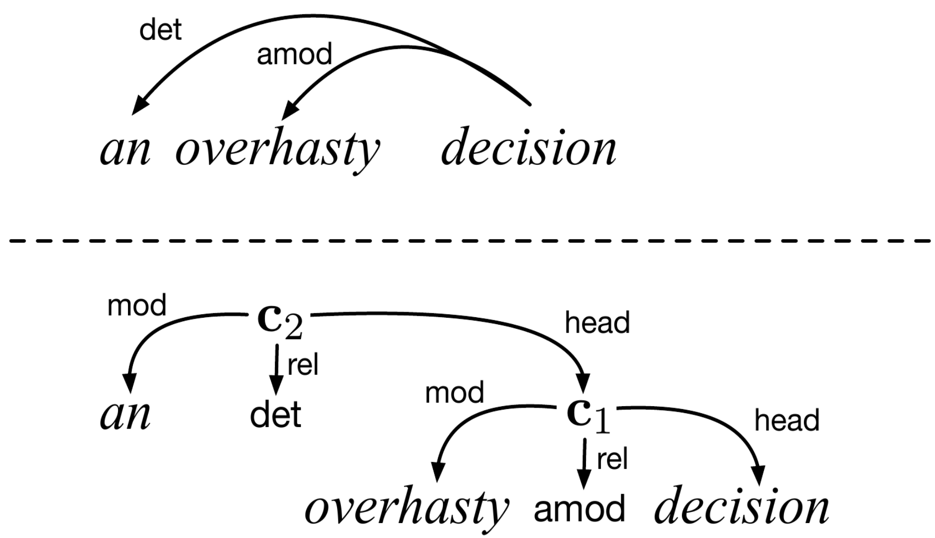
Dyer, C., Ballesteros, M., Ling, W., Matthews, A., & Smith, N. A. (2015). Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ACL.

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

	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

English parsing results (SD)

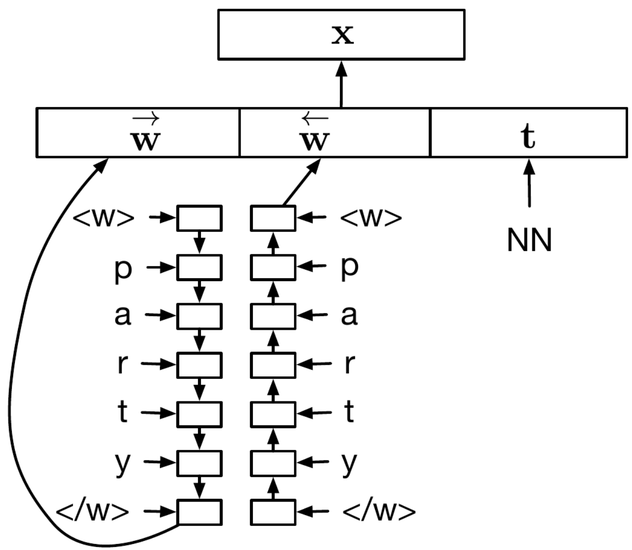
	Development		Test	
	UAS	LAS	UAS	LAS
S-LSTM	**87.2**	**85.9**	**87.2**	**85.7**
−POS	82.8	79.8	82.2	79.1
−pretraining	86.3	84.7	85.7	84.1
−composition	85.8	84.0	85.3	83.6
S-RNN	86.3	84.7	86.1	84.6
C&M (2014)	84.0	82.4	83.9	82.4

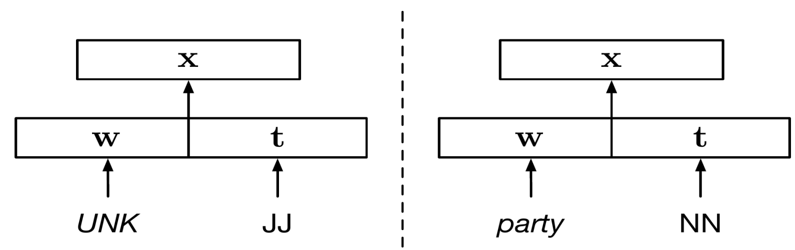
Chinese parsing results (CTB5)

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

UAS

Language	Words	Chars	Words + POS	Chars + POS
Arabic	86.14	**87.20**	**87.44**	87.07
Basque	78.42	**84.97**	83.49	**85.58**
French	84.84	**86.21**	**87.00**	86.33
German	88.14	**90.94**	91.16	**91.23**
Hebrew	79.73	**79.92**	**81.99**	80.76
Hungarian	72.38	**80.16**	78.47	**80.85**
Korean	78.98	**88.98**	87.36	**89.14**
Polish	73.29	**85.69**	**89.32**	88.54
Swedish	73.44	**75.03**	**80.02**	78.85
Turkish	71.10	**74.91**	77.13	**77.96**
Chinese	79.43	**80.36**	**85.98**	85.81
English	91.64	**91.98**	**92.94**	92.49
Average	79.79	**83.86**	85.19	**85.38**

LAS

Language	Words	Chars	Words + POS	Chars + POS
Arabic	82.73	**84.34**	**84.81**	84.36
Basque	67.08	**78.22**	74.31	**79.52**
French	80.32	**81.70**	**82.71**	81.51
German	85.36	**88.68**	**89.04**	88.83
Hebrew	69.42	**70.58**	**74.11**	72.18
Hungarian	62.14	**75.61**	69.50	**76.16**
Korean	67.48	**86.80**	83.80	**86.88**
Polish	65.13	**78.23**	**81.84**	80.97
Swedish	64.77	**66.74**	**72.09**	69.88
Turkish	53.98	**62.91**	62.30	**62.87**
Chinese	75.64	**77.06**	**84.36**	84.10
English	88.60	**89.58**	**90.63**	90.08
Average	71.89	**78.37**	79.13	**79.78**

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

Dependency Parsing

UAS

Language	Words	Chars	Words + POS	Chars + POS
Arabic	85.21	**86.08**	86.05	**86.07**
Basque	77.06	**85.19**	82.92	**85.22**
French	83.74	**85.34**	**86.15**	85.78
German	82.75	**86.80**	**87.33**	87.26
Hebrew	77.62	**79.93**	**80.68**	80.17
Hungarian	72.78	**80.35**	78.64	**80.92**
Korean	78.70	**88.39**	86.85	**88.30**
Polish	72.01	**83.44**	**87.06**	85.97
Swedish	76.39	**79.18**	**83.43**	83.24
Turkish	71.70	**76.32**	75.32	**76.34**
Chinese	79.01	**79.94**	**85.96**	85.30
English	91.16	**91.47**	**92.57**	91.63
Average	79.01	**85.36**	84.41	**84.68**

LAS

Language	Words	Chars	Words + POS	Chars + POS
Arabic	82.05	**83.41**	**83.46**	83.40
Basque	66.61	**79.09**	73.56	**78.61**
French	79.22	**80.92**	**82.03**	81.08
German	79.15	**84.04**	**84.62**	84.49
Hebrew	68.71	**71.26**	**72.70**	72.26
Hungarian	61.93	**75.19**	69.31	**76.34**
Korean	67.50	**86.27**	83.37	**86.21**
Polish	63.96	**76.84**	**79.83**	78.24
Swedish	67.69	**71.19**	**76.40**	74.47
Turkish	54.55	**64.34**	61.22	**62.28**
Chinese	74.79	**76.29**	**84.40**	83.72
English	88.42	**88.94**	**90.31**	89.44
Average	71.22	**78.15**	78.43	**79.21**

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

Dependency Parsing

	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

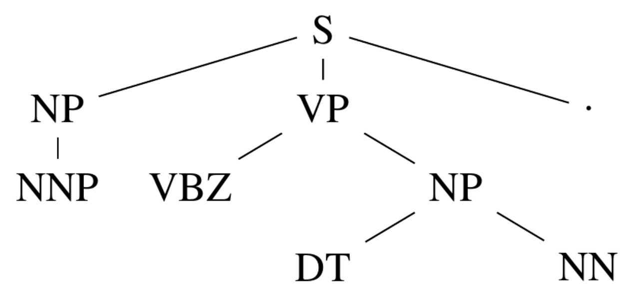
Ballesteros, M., Dyer, C., & Smith, N. A. (2015). Improved Transition-Based Parsing by Modeling Characters instead of Words with LSTMs. EMNLP.

Part 3.4: Sequence to Sequence with Greedy Search

Constituent Parsing

- Sequence to sequence

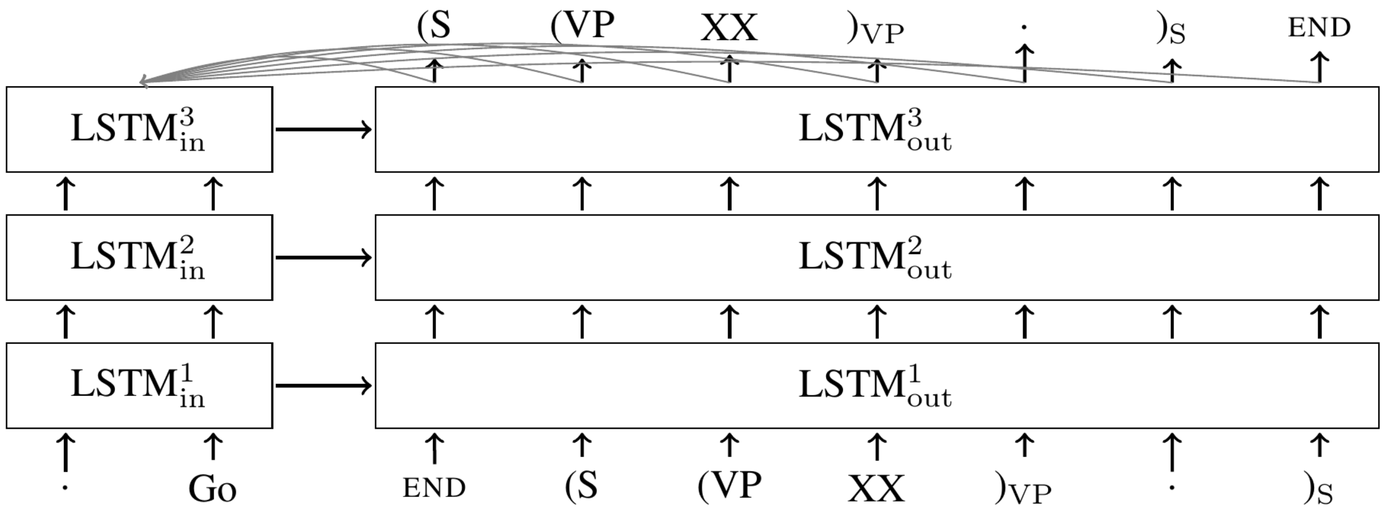
John has a dog . $\rightarrow$



John has a dog . $\rightarrow$ $(S\ (NP\ NNP\ )_{NP}\ (VP\ VBZ\ (NP\ DT\ NN\ )_{NP}\ )_{VP}\ .\ )_{S}$

Vinyals, O., Kaiser, L., Koo, T., Petrov, S., & Sutskever, I. (2015). Grammar as a Foreign Language. ICLR.

Constituent Parsing



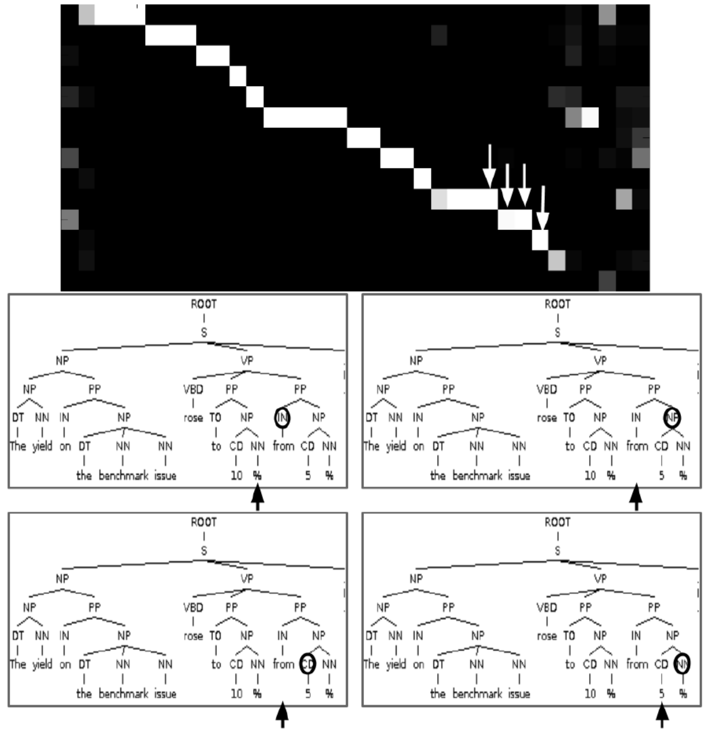
Vinyals, O., Kaiser, L., Koo, T., Petrov, S., & Sutskever, I. (2015). Grammar as a Foreign Language. ICLR.

Constituent Parsing

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**

Vinyals, O., Kaiser, L., Koo, T., Petrov, S., & Sutskever, I. (2015). Grammar as a Foreign Language. ICLR.

Constituent Parsing



Vinyals, O., Kaiser, L., Koo, T., Petrov, S., & Sutskever, I. (2015). Grammar as a Foreign Language. ICLR.