Joint Models for NLP

Yue Zhang

Westlake Institute for Advanced Study

Outline

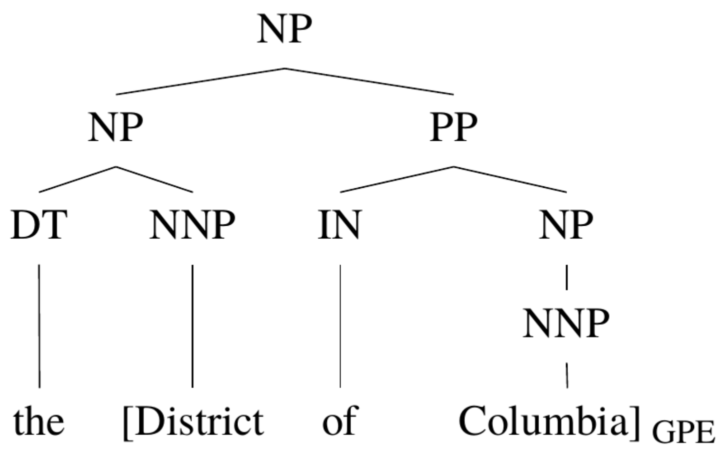
- Motivation
- Statistical Models
- Deep Learning Models

Outline

- Motivation
- Statistical Models
- Deep Learning Models

Motivation

- Related tasks in NLP
  - Constituents and named entities



Motivation

- Related tasks in NLP
  - NER, Chunking and POS Tagging

Sentence:	Joi	runs	the	MIT	Media	Lab	.
POS Tagging:	NNP	VBZ	DT	NNP	NNP	NNP	.
NER:	PER	O	O	B-ORG	I-ORG	I-ORG	O
Chunking:	S	S	S	B	I	E	O

Motivation

- Pipelines in NLP
  - Segmentation ⟹ POS tagging

布朗访问上海

⇩

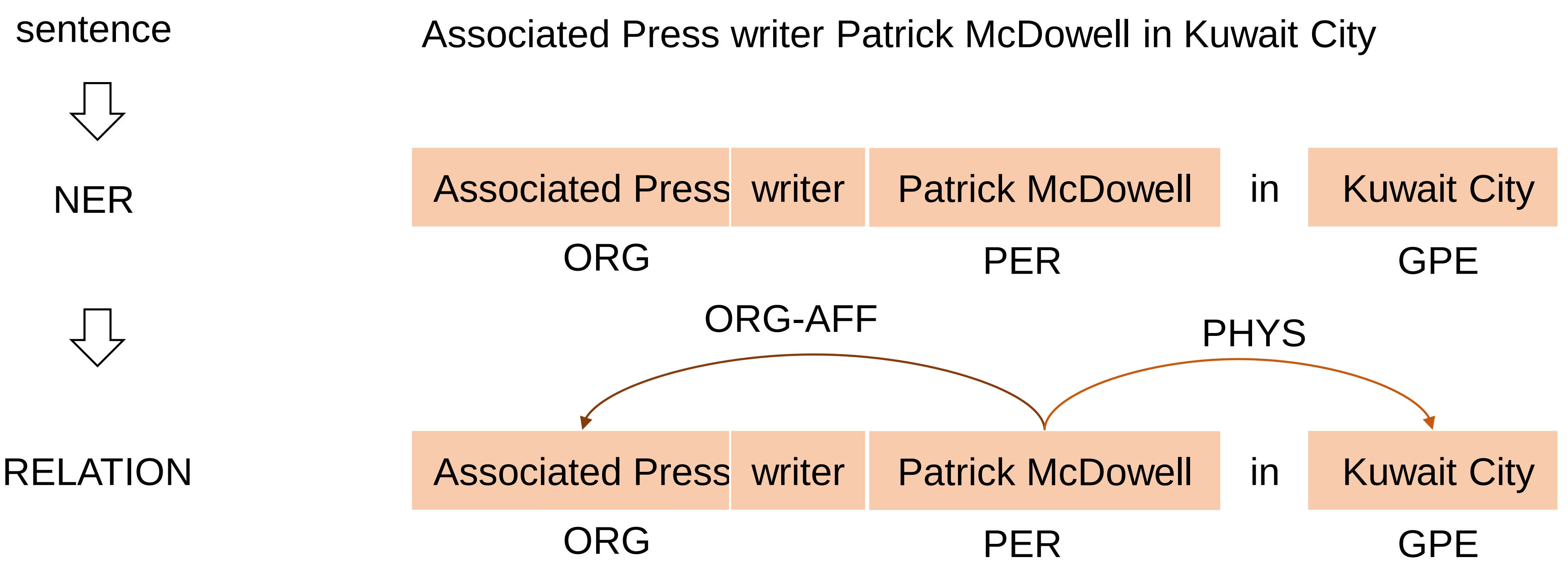
布朗/ 访问/ 上海/

⇩

布朗/NR 访问/VV 上海/NR

Motivation

- Pipelines in NLP
  - Entity and Relation



Motivation

- Pipelines in NLP
  - Entity and Sentiment

sentence	So excited to meet my baby Farah !!!
⇩	
NER	So excited to meet my [baby Farah] !!! PER
⇩	
Sentiment	So excited to meet my [baby Farah]+ !!! PER + POSITIVE

Motivation

- Joint model
  - Reduce error propagation
  - Allow information exchange between tasks
- Challenge
  - Joint learning
  - Search

Solutions

Learning

Search

	Joint	**Separate**
Joint	Statistical Neural	Statistical
Separate	Neural	

Outline

- Motivation
- Statistical Models
- Deep Learning Models

Statistical Models

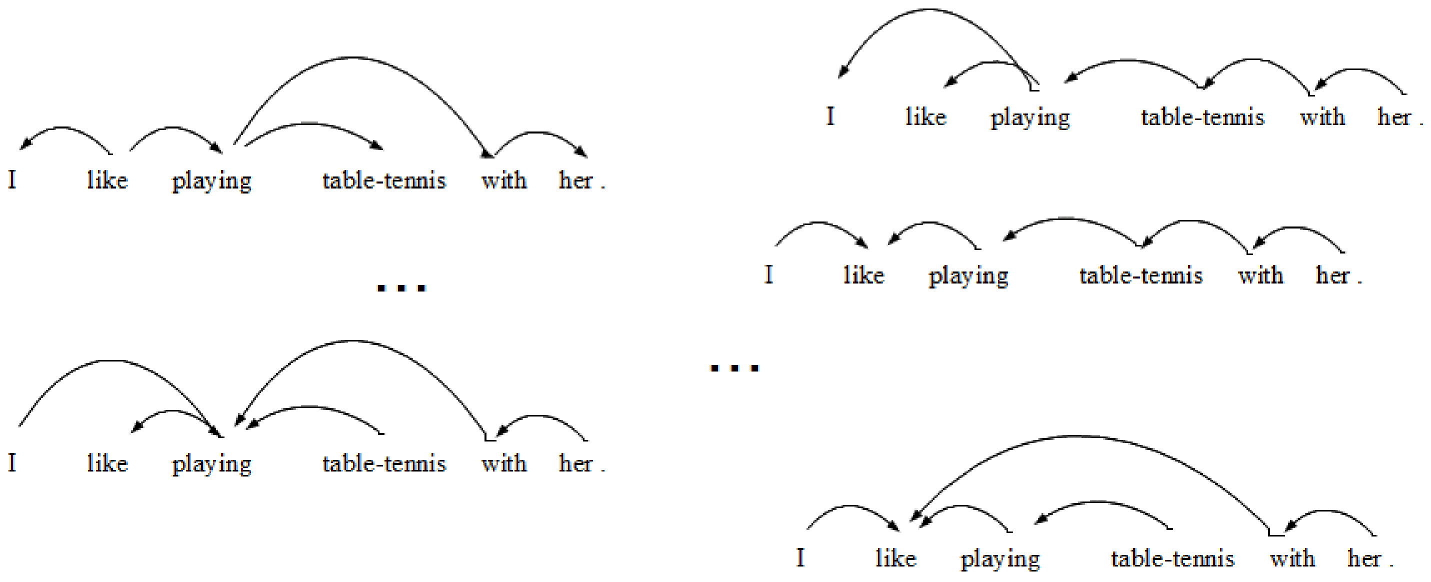
- Graph-Based Methods
- Transition-Based Methods

Statistical Models

- Graph-Based Methods
- Transition-Based Methods

Graph-Based Methods

- Traditional solution
  - Score each candidate, select the highest-scored output
  - Search-space typically exponential



Graph-Based Methods

- Joint Label Structure
- Reranking
- Joint Modeling (Multi task)
- Joint Modeling (Single task)

Graph-Based Methods

- Joint Label Structure
- Reranking
- Joint Modeling (Multi task)
- Joint Modeling (Single task)

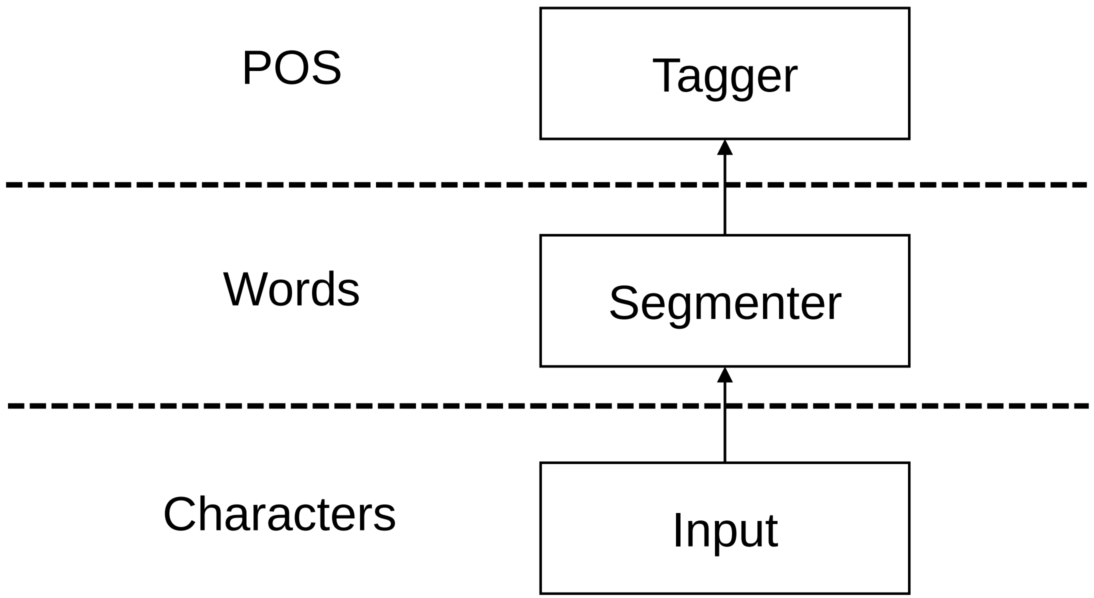
Graph-Based Methods

- Joint Label Structure
- Reranking
- Joint Modeling (Multi task)
- Joint Modeling (S…

Joint Learning , Joint Search

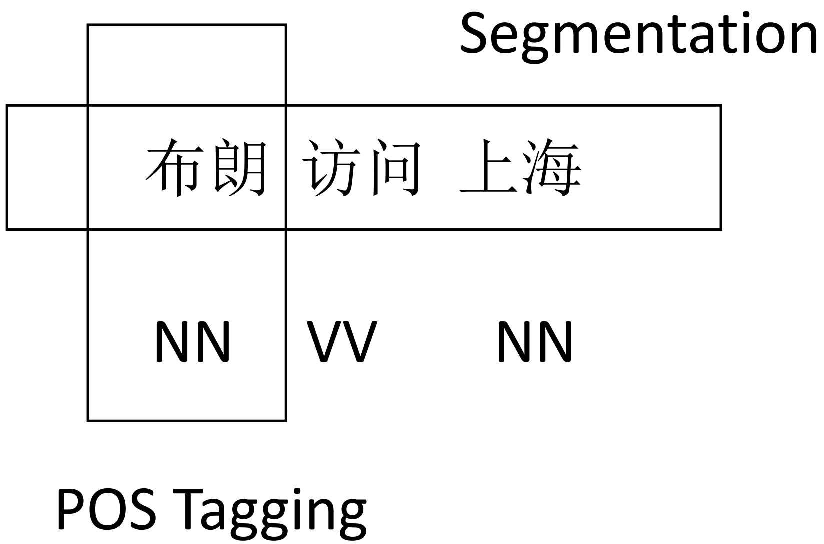
Joint Segmentation and POS tagging

- Tasks



Joint Segmentation and POS tagging

- Task



Ng, Hwee Tou, and Jin Kiat Low. "Chinese part-of-speech tagging: One-at-a-time or all-at-once? word-based or character-based?." *EMNLP*. 2004.

Joint Segmentation and POS tagging

- Collapsing labels

BE BE BE

布朗 访问 上海

NN VV NN

B-NN E-NN B-VV E-VV B-NN E-NN

⇑ ⇑ ⇑ ⇑ ⇑ ⇑

布 朗 访 问 上 海

Ng, Hwee Tou, and Jin Kiat Low. "Chinese part-of-speech tagging: One-at-a-time or all-at-once? word-based or character-based?." *EMNLP*. 2004.

Joint Segmentation and POS tagging

- All-at-Once, Character-Based POS Tagger and Segmenter : Feature

(a) $C_n \ (n = -2,-1,0,1,2)$

(b) $C_n C_{n+1} (n = -2,-1,0,1)$

(c) $C_{-1} C_1$

(d) $W_0 C_0$

(e) $Pu(C_0)$

(f) $T(C_{-2})T(C_{-1})T(C_0)T(C_1)T(C_2)$

(g) $B(C_{-1W_0})POS(C_{-1W_0})$

(h) $B(C_{-2W_0})POS(C_{-2W_0})B(C_{-1W_0})POS(C_{-1W_0})$

Ng, Hwee Tou, and Jin Kiat Low. "Chinese part-of-speech tagging: One-at-a-time or all-at-once? word-based or character-based?." *EMNLP*. 2004.

Joint Segmentation and POS tagging

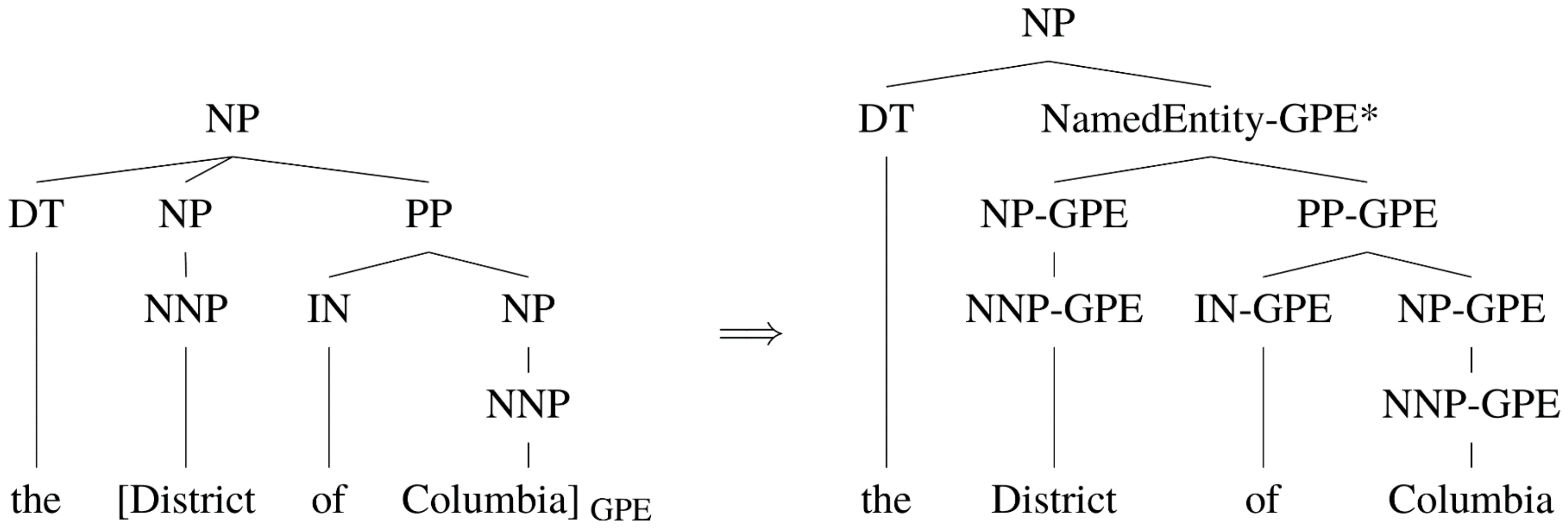
- Results on CTB

Method	Word Seg F-measure (%)	POS Accuracy (%)	Total Testing Time
One-at-a-Time Word-Based	95.1	84.1	1 min 20 secs
One-at-a-Time Char-Based	95.1	91.7	1 min 50 secs
All-At-Once Char-Based	95.2	91.9	20 mins

Ng, Hwee Tou, and Jin Kiat Low. "Chinese part-of-speech tagging: One-at-a-time or all-at-once? word-based or character-based?." *EMNLP*. 2004.

Joint Parsing and NER

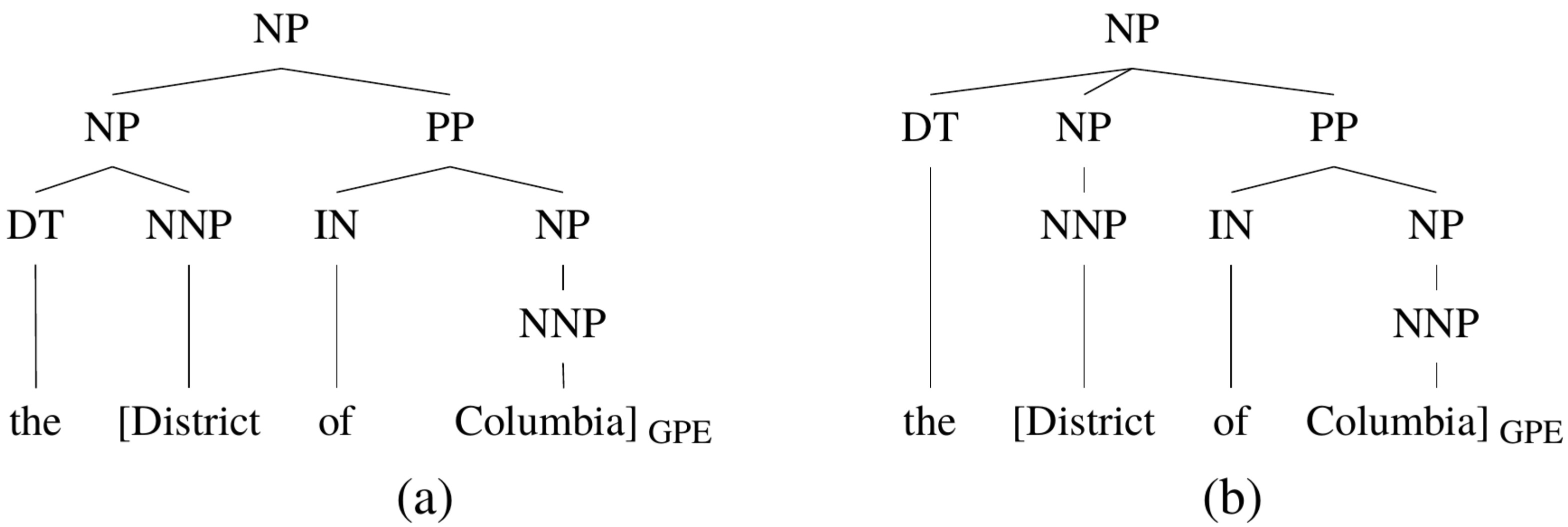
- A joint model of both parsing and named entity recognition.



Finkel, Jenny Rose, and Christopher D. Manning. "Joint parsing and named entity recognition." *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009.

Joint Parsing and NER

- A feature-based CRF-CFG parser operating over tree structures augmented with NER information.



Finkel, Jenny Rose, and Christopher D. Manning. "Joint parsing and named entity recognition." *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009.

Joint Parsing and NER

- Results:
  - On OntoNotes

		Parse Labeled Bracketi						Training
		Precision	Recall	F				Time
ABC	Just Parse	70.18%	70.12%	**70.15%**		–		25m
	Just NER		–		76.84%	72.32%	74.51%	
	Joint Model	69.76%	70.23%	69.99%	77.70%	72.32%	**74.91%**	45m
CNN	Just Parse	76.92%	77.14%	77.03%		–		16.5h
	Just NER		–		75.56%	76.00%	75.78%	
	Joint Model	77.43%	77.99%	**77.71%**	78.73%	78.67%	**78.70%**	31.7h
MNB	Just Parse	63.97%	67.07%	65.49%		–		12m
	Just NER		–		72.30%	54.59%	62.21%	
	Joint Model	63.82$	67.46%	**65.59%**	71.35%	62.24%	**66.49%**	19m
NBC	Just Parse	59.72%	63.67%	61.63%		–		10m
	Just NER		–		67.53%	60.65%	63.90%	
	Joint Model	60.69%	65.34%	**62.93%**	71.43%	64.81%	**67.96%**	17m
PRI	Just Parse	76.22%	76.49%	76.35%		–		2.4h
	Just NER		–		82.07%	84.86%	83.44%	
	Joint Model	76.88%	77.95%	**77.41%**	86.13%	86.56%	**86.34%**	4.2h
VOA	Just Parse	76.56%	75.74%	76.15%		–		2.3h
	Just NER		–		82.79%	75.96%	79.23%	
	Joint Model	77.58%	77.45%	**77.51%**	88.37%	87.98%	**88.18%**	4.4h

Finkel, Jenny Rose, and Christopher D. Manning. "Joint parsing and named entity recognition." *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009.

Graph-Based Methods

- Joint Label Structure
- Reranking
- Joint Modeling (Multi task)
- Joint Modeling (Single task)

Graph-Based Methods

- Joint Label Structure
- Reranking
- Joint Modeling (Multi task)
- Joint Modeling (S

Separate Learning , **Joint** Search

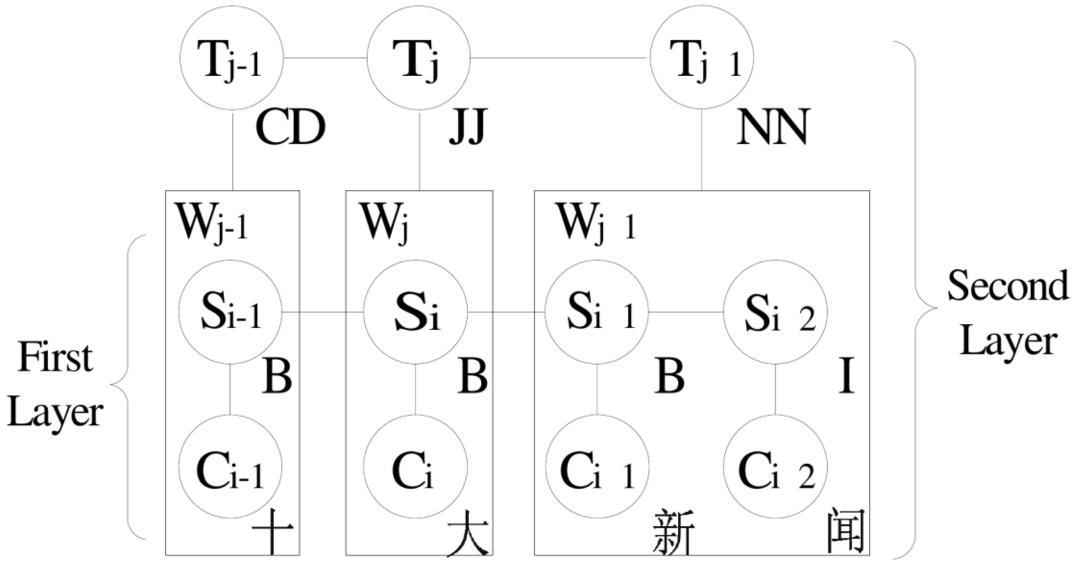
Joint Segmentation and POS Tagging

- Two separate CRF taggers.
- Separately trained, reranking.
- Use tag sequence score to rank segmentation.

Shi, Yanxin, and Mengqiu Wang. "A Dual-layer CRFs Based Joint Decoding Method for Cascaded Segmentation and Labeling Tasks." *IJcAI*. 2007.

Joint Segmentation and POS Tagging

- Dual-layer CRFs



Shi, Yanxin, and Mengqiu Wang. "A Dual-layer CRFs Based Joint Decoding Method for Cascaded Segmentation and Labeling Tasks." *IJcAI*. 2007.

Joint Segmentation and POS Tagging

- Results on Segmentation

	1	2	3	4	5	6
Baseline	97.3%	97.2%	95.4%	96.7%	96.2%	93.1%
Joint decoding	97.4%	97.3%	95.7%	96.9%	96.4%	93.4%

	7	8	9	10	average
Baseline	95.9%	94.8%	95.7%	96.2 %	95.85%
Joint decoding	96.0%	95.2%	95.9%	96.3%	96.05%

	AS			CTB		
	P	*R*	*F1*	*P*	*R*	*F1*
Baseline	96.7%	96.8%	96.7%	88.5%	88.3%	88.4%
Joint Decoding	96.9%	96.7%	96.8%	89.4%	88.7%	89.1%

	PK			HK		
	P	*R*	*F1*	*P*	*R*	*F1*
Baseline	94.9%	94.9%	94.9%	94.9%	95.5%	95.2%
Joint Decoding	95.3%	95.0%	95.2%	95.0%	95.4%	95.2%

	ASo	CTBo	HKo	PKo	S-Avg	O-Avg
S01		88.1%		**95.3%**	91.7%	**92.2%**
S02		**91.2%**			**91.2%**	89.1%
S03	87.2%	82.9%	88.6%	92.5%	87.8%	**94.1%**
S04				93.7%	93.7%	**95.2%**
S07				94.0%	94.0%	**95.2%**
S08			**95.6%**	93.8%	94.7%	**95.2%**
S10		90.1%		**95.9%**	**93.0%**	92.2%
S11	90.4%	88.4%	87.9%	88.6%	88.8%	**94.1%**
Peng *et al.* '04	95.7%	89.4%	94.6%	94.6%	93.6%	**94.1%**
Our System	**96.8%**	89.1%	95.2%	95.2%		94.1%

Shi, Yanxin, and Mengqiu Wang. "A Dual-layer CRFs Based Joint Decoding Method for Cascaded Segmentation and Labeling Tasks." *IJcAI*. 2007.

Joint Segmentation and POS Tagging

- Results on POS Tagging

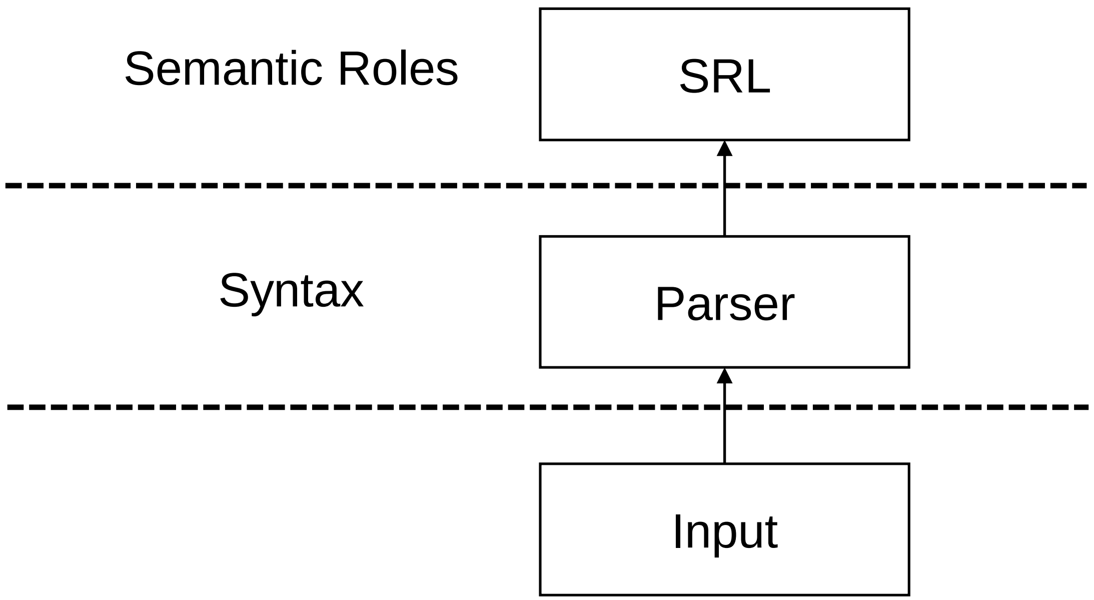
	1	2	3	4	5	6
Baseline	93.8%	93.7%	90.2%	92.0%	93.3%	87.2%
Joint Decoding	94.0%	93.9%	90.4%	92.2%	93.4%	87.5%

	7	8	9	10	average
Baseline	92.2%	90.8%	91.5%	92.0 %	91.67%
Joint Decoding	92.4%	91.0%	91.7%	92.1%	91.86%

Shi, Yanxin, and Mengqiu Wang. "A Dual-layer CRFs Based Joint Decoding Method for Cascaded Segmentation and Labeling Tasks." *IJcAI*. 2007.

Joint Parsing and SRL

- Task



Sutton, Charles, and Andrew McCallum. "Joint parsing and semantic role labeling." *Proceedings of the Ninth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2005.

Joint Parsing and SRL

- Rerank k-best parse trees from a probabilistic parser using an SRL system.

Sutton, Charles, and Andrew McCallum. "Joint parsing and semantic role labeling." *Proceedings of the Ninth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2005.

Joint Parsing and SRL

- Overall results

	Precision	Recall	$F_{\beta=1}$
Development	64.43%	63.11%	63.76
Test WSJ	68.57%	64.99%	66.73
Test Brown	62.91%	54.85%	58.60
Test WSJ+Brown	67.86%	63.63%	65.68

- Did *not* beat a pipeline baseline

  Many subsequent CoNLL shared tasks show difficulties for this joint task

Sutton, Charles, and Andrew McCallum. "Joint parsing and semantic role labeling." *Proceedings of the Ninth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2005.

Graph-Based Methods

- Joint Label Structure
- Reranking
- Joint Modeling (Multi task)
- Joint Modeling (Single task)

Graph-Based Methods

- Joint Label Structure
- Reranking
- Joint Modeling (Multi task)
- Joint Modeling (S

Separate Learning , **Joint** Search

Joint Modeling

- Joint Search, separate training
- Search complex problem
  - ILP
  - BP
  - Dual Decomposition

Auli, Michael, and Adam Lopez. "A comparison of loopy belief propagation and dual decomposition for integrated CCG supertagging and parsing." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

Joint Entity and Sentiment

- Task:
  - Opinion linking relations
    - The numberic subscripts denote linking relations, one of IS-ABOUT OR IS-FROM
  - Opinion entities:
    - Opinion expressions: O
    - Opinion targets: T
    - Opinion holders: H

jointly identifies opinion-related entities, as well as opinion linking relations

[The workers]$_{[H_{1,2}]}$ were irked $_{[O_1]}$ by [the government report]$_{[T_1]}$ and were worried$_{[O_2]}$ as they went about their daily chores.

Yang, Bishan, and Claire Cardie. "Joint Inference for Fine-grained Opinion Extraction." *ACL (1)*. 2013.

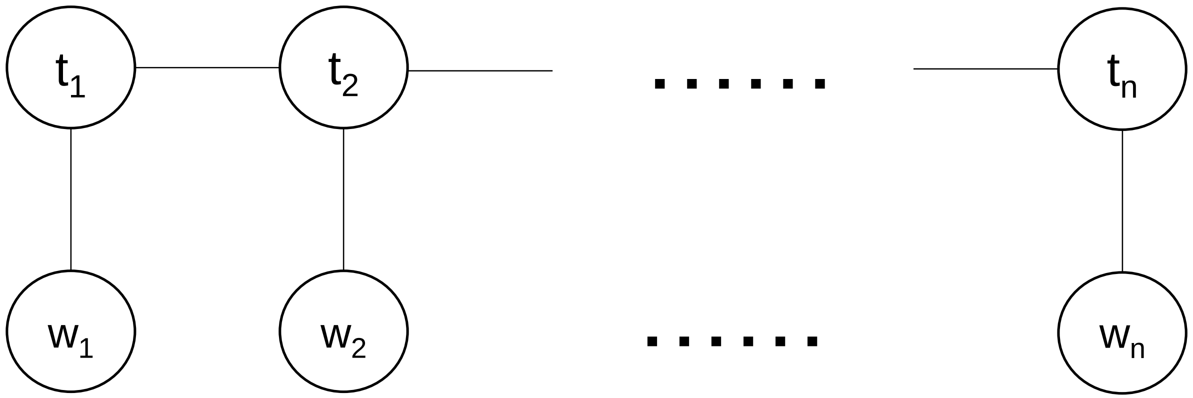
Joint Entity and Sentiment

- Model
  - Formulate the task of **opinion entity identification** as a sequence labeling problem and employ conditional random fields (CRFs) to learn the probability of a sequence assignment y for a given sentence x;
  - Treat the **relation extraction** problem as a combination of two binary classification problems and use L1-regularized logistic regression to train the classifiers;
  - Optimize the joint objective function which is defined as a linear combination of the potentials from different predictors with a parameter λ to balance the contribution of these two components: opinion entity identification and opinion relation extraction.

Yang, Bishan, and Claire Cardie. "Joint Inference for Fine-grained Opinion Extraction." *ACL (1)*. 2013.

Joint Entity and Sentiment

- CRF



D – Opinion expression

T – Opinion target

H – Opinion Holder

N – Opinion None

Yang, Bishan, and Claire Cardie. "Joint Inference for Fine-grained Opinion Extraction." *ACL (1)*. 2013.

Joint Entity and Sentiment

- A classification model for opinion target relation
- A classification model for opinion holder relation
- Syntactic and semantic features are used

Yang, Bishan, and Claire Cardie. "Joint Inference for Fine-grained Opinion Extraction." *ACL (1)*. 2013.

Joint Entity and Sentiment

- Joint scoring function by linear interpolation

$$
\begin{aligned}
Score = \; & \lambda \cdot Score_{(entity)} \\
& +(1-\lambda) \cdot Score_{(relation)}
\end{aligned}
$$

Yang, Bishan, and Claire Cardie. "Joint Inference for Fine-grained Opinion Extraction." *ACL (1)*. 2013.

Joint Entity and Sentiment

- ILP for search
  - Constraint 1: Uniqueness
  - Constraint 2: Non-overlapping
  - Constraint 3: Consistency between the opinion-arg and opinion-implicit-arg classifiers
  - Constraint 4: Consistency between opinion-arg classifier and opinion entity extractor
  - Constraint 5: Consistency between the opinion-implicit-arg classifier and opinion entity extractor

Yang, Bishan, and Claire Cardie. "Joint Inference for Fine-grained Opinion Extraction." *ACL (1)*. 2013.

Joint Entity and Sentiment

- Results on MPQA

	Opinion Expression			Opinion Target			Opinion Holder		
Method	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
CRF	82.21	66.15	73.31	73.22	48.58	58.41	72.32	49.09	58.48
CRF+Adj	82.21	66.15	73.31	80.87	42.31	55.56	75.24	48.48	58.97
CRF+Syn	82.21	66.15	73.31	81.87	30.36	44.29	78.97	40.20	53.28
CRF+RE	83.02	48.99	61.62	85.07	22.01	34.97	78.13	40.40	53.26
Joint-Model	71.16	77.85	**74.35***	75.18	57.12	**64.92****	67.01	66.46	**66.73****
CRF	66.60	52.57	58.76	44.44	29.60	35.54	65.18	44.24	52.71
CRF+Adj	66.60	52.57	58.76	49.10	25.81	33.83	68.03	43.84	53.32
CRF+Syn	66.60	52.57	58.76	50.26	18.41	26.94	74.60	37.98	50.33
CRF+RE	69.27	40.09	50.79	60.45	15.37	24.51	75	38.79	51.13
Joint-Model	57.39	62.40	**59.79***	49.15	38.33	**43.07****	62.73	62.22	**62.47****

Yang, Bishan, and Claire Cardie. "Joint Inference for Fine-grained Opinion Extraction." *ACL (1)*. 2013.

Joint Entity and Sentiment

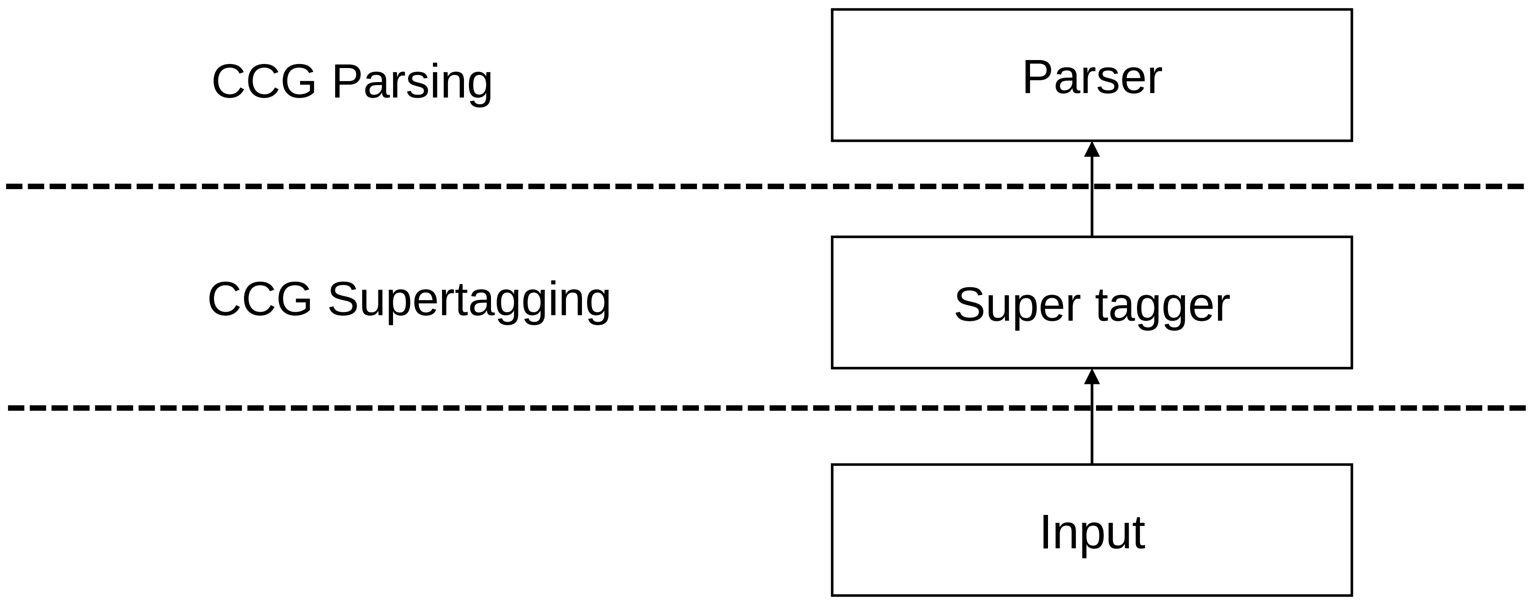
- Results on MPQA

	IS-ABOUT			IS-FROM		
Method	**P**	**R**	**F1**	**P**	**R**	**F1**
CRF+Adj	73.65	37.34	49.55	70.22	41.58	52.23
CRF+Syn	76.21	28.28	41.25	77.48	36.63	49.74
CRF+RE	78.26	20.33	32.28	74.81	37.55	50.00
CRF+Adj-merged-10-best	25.05	61.18	35.55	30.28	62.82	40.87
CRF+Syn-merged-10-best	41.60	45.66	43.53	48.08	54.03	50.88
CRF+RE-merged-10-best	51.60	33.09	40.32	47.73	54.40	50.84
Joint-Model	64.38	51.20	**57.04****	64.97	58.61	**61.63****

Yang, Bishan, and Claire Cardie. "Joint Inference for Fine-grained Opinion Extraction." *ACL (1)*. 2013.

Joint Supertagging and Parsing

- Tasks



Joint Supertagging and Parsing

- **CCG parsing** (for English, Chinese and other languages) is to find the syntactic structures of written text based on combinatory categorial grammars.

Marcel proved completeness

NP (S\ NP)/NP NP

S\ NP

S

Supper tagging and parsing

Auli, Michael, and Adam Lopez. "A comparison of loopy belief propagation and dual decomposition for integrated CCG supertagging and parsing." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

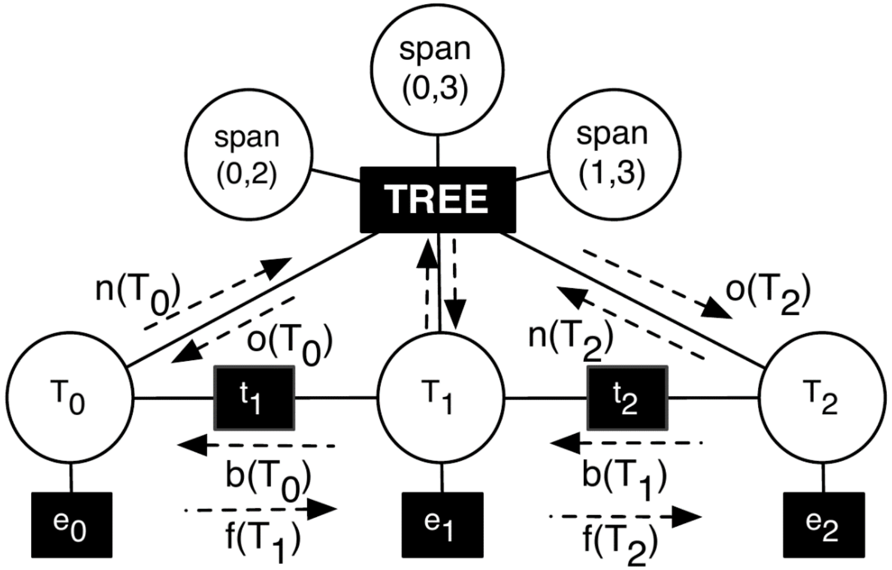
Joint Supertagging and Parsing

- CCG traditionally done by supertagging -> parsing

Auli, Michael, and Adam Lopez. "A comparison of loopy belief propagation and dual decomposition for integrated CCG supertagging and parsing." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

Joint Supertagging and Parsing

- Loopy belief propagation
- Factor graph for the combined parsing and supertagging model



Auli, Michael, and Adam Lopez. "A comparison of loopy belief propagation and dual decomposition for integrated CCG supertagging and parsing." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

Joint Supertagging and Parsing

- Dual decomposition: Lagrangian method for constraint optimization

$$\arg\max_{y \in Y, z \in Z} f(y) + g(z)$$

$$\text{such that } y(i,t) = z(i,t) \text{ for all } (i,t) \in I$$

$$L(u) = \max_{y \in Y}(f(y) - \sum_{i,t} u(i,t) y(i,t)) + \max_{z \in Z}(f(z) + \sum_{i,t} u(i,t) z(i,t))$$

Auli, Michael, and Adam Lopez. "A comparison of loopy belief propagation and dual decomposition for integrated CCG supertagging and parsing." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

Joint Supertagging and Parsing

- Results

	section 00 (dev)						section 23 (test)					
	AST			Reverse			AST			Reverse		
	LF	UF	ST	LF	UF	ST	LF	UF	ST	LF	UF	ST
Baseline	87.38	93.08	94.21	87.36	93.13	93.99	87.73	93.09	94.33	87.65	93.06	94.01
C&C '07	87.24	93.00	94.16	-	-	-	87.64	93.00	94.32	-	-	-
$BP_{k=1}$	87.70	93.28	94.44	**88.35**	93.69	**94.73**	**88.20**	**93.28**	**94.60**	88.78	93.66	94.81
$BP_{k=25}$	87.70	93.31	94.44	88.33	**93.72**	94.71	88.19	93.27	94.59	88.80	93.68	94.81
$DD_{k=1}$	87.40	93.09	94.23	87.38	93.15	94.03	87.74	93.10	94.33	87.67	93.07	94.02
$DD_{k=25}$	**87.71**	**93.32**	**94.44**	88.29	93.71	94.67	88.14	93.24	94.59	**88.80**	**93.68**	**94.82**

BP: Belief Propagation

DD: Dual Decomposition

Auli, Michael, and Adam Lopez. "A comparison of loopy belief propagation and dual decomposition for integrated CCG supertagging and parsing." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

Graph-Based Methods

- Joint Label Structure
- Reranking
- Joint Modeling (Multi task)
- Joint Modeling (Single task)

Graph-Based Methods

- Joint Label Structure
- Reranking
- Joint Modeling (Multi task)
- Joint Modeling (S

Joint Learning ， **Joint** Search

Joint Modeling (Single task)

- A Single Model

$Score = \Phi(\boldsymbol{y}) \bullet \vec{\omega}$

where $\boldsymbol{y}$ is the model features

Joint Segmentation and POS Tagging

- Task

Input	我喜欢读书	Ilikereadingbooks
Output	我/PN 喜欢/V 读/V 书/N	I/PN like/V reading/V books/N

Zhang, Yue, and Stephen Clark. "Joint Word Segmentation and POS Tagging Using a Single Perceptron." *ACL*. 2008.

Joint Segmentation and POS Tagging

- Feature templates for the baseline segmentor

1	word w	9	word w immediately before character c
2	word bigram w_1w_2	10	character c immediately before word w
3	single-character word w	11	the starting characters c_1 and c_2 of two consecutive words
4	a word of length l with starting character c	12	the ending characters c_1 and c_2 of two consecutive words
5	a word of length l with ending character c	13	a word of length l with previous word w
6	space-separated characters c_1 and c_2	14	a word of length l with next word w
7	character bigram c_1c_2 in any word		
8	the first / last characters c_1 / c_2 of any word		

Zhang, Yue, and Stephen Clark. "Joint Word Segmentation and POS Tagging Using a Single Perceptron." *ACL*. 2008.

Joint Segmentation and POS Tagging

- Feature templates for the baseline POS tagger

1	tag t with word w	11	tag t on a word containing char c (not the starting or ending character)
2	tag bigram t_1t_2		
3	tag trigram $t_1t_2t_3$	12	tag t on a word starting with char c_0 and containing char c
4	tag t followed by wo		
5	word w followed by	13	tag t on a word ending with char c_0 and containing char c
6	word w with tag t ar		
7	word w with tag t ar	14	tag t on a word containing repeated char cc
8	tag t on single-chara ter trigram c_1wc_2	15	tag t on a word starting with character category g
9	tag t on a word starting with char c	16	tag t on a word ending with character category g
10	tag t on a word ending with char c		

Zhang, Yue, and Stephen Clark. "Joint Word Segmentation and POS Tagging Using a Single Perceptron." *ACL*. 2008.

Joint Segmentation and POS Tagging

- Perceptron with both segmentation and POS features

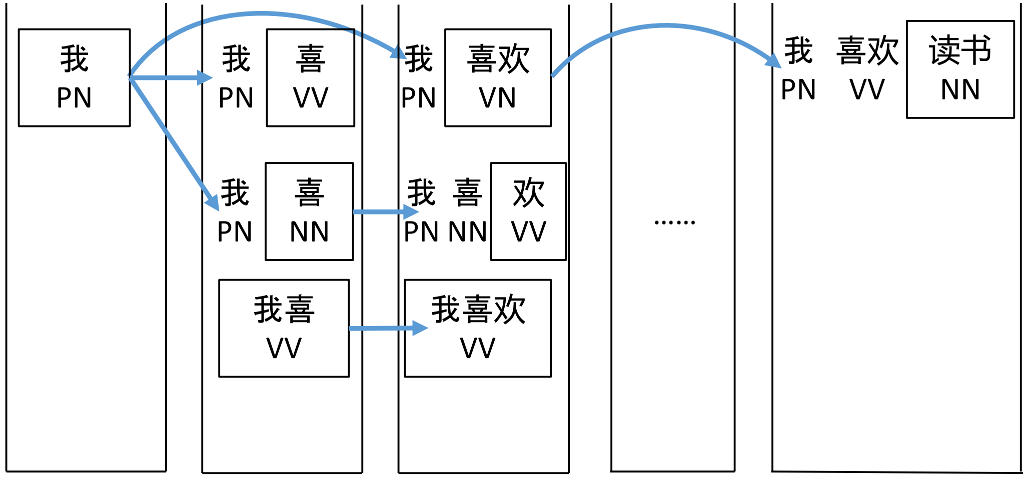
Inputs: training examples (x_i, y_i)
Initialization: set $\vec{w} = 0$
Algorithm:
  for $t = 1..T$, $i = 1..N$
    calculate $z_i = \arg\max_{y \in \mathrm{GEN}(x_i)} \Phi(y) \cdot \vec{w}$
    if $z_i \neq y_i$
      $\vec{w} = \vec{w} + \Phi(y_i) - \Phi(z_i)$
Outputs: $\vec{w}$

The perceptron learning algorithm

Zhang, Yue, and Stephen Clark. "Joint Word Segmentation and POS Tagging Using a Single Perceptron." *ACL*. 2008.

Joint Segmentation and POS Tagging

- The decoding algorithm for the joint word segmentor and POS tagger, $\text{agendas}[i]$ stores the best sequences that end at i



Algorithm:

for $end\_index = 1$ to $sent$.length:

  foreach tag:

    for $start\_index =$

     $\max(1,\ end\_index - maxlen[tag] + 1)$

    to end_index:

      $word = sent[start\_index..end\_i$

      if $(word, tag)$ consistent with tag

        for $item \in agendas[start\_ind$

          $item_1 = item$

          $item_1$.append(($word$,tag))

          $agendas[end\_index]$.insert($item_1$)

Outputs: $agendas[sent.\text{length}]$.best_item

Zhang, Yue, and Stephen Clark. "Joint Word Segmentation and POS Tagging Using a Single Perceptron." *ACL*. 2008.

Joint Segmentation and POS Tagging

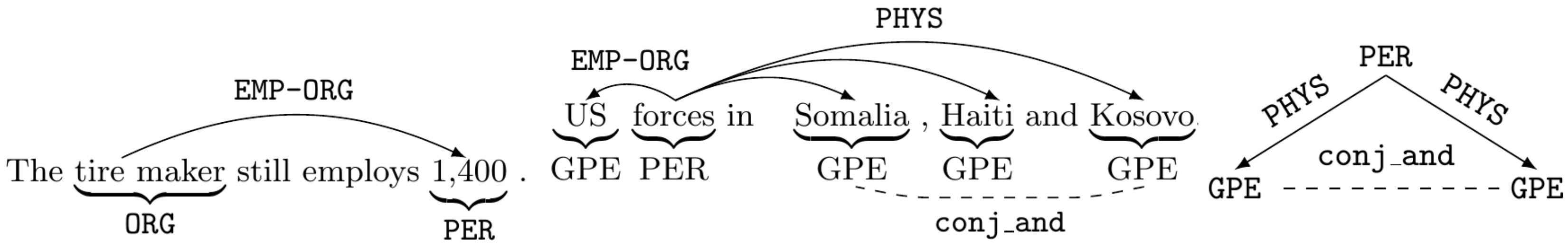
- Results by 10-fold cross validation using CTB

Model	SF	TF	TA
Baseline+ (Ng)	95.1	–	91.7
Joint+ (Ng)	95.2	–	91.9
Baseline+* (Shi)	95.85	91.67	–
Joint+* (Shi)	96.05	91.86	–
Baseline (ours)	95.20	90.33	92.17
Joint (ours)	95.90	91.34	93.02

Zhang, Yue, and Stephen Clark. "Joint Word Segmentation and POS Tagging Using a Single Perceptron." *ACL*. 2008.

Joint Entity Relation Extraction

- Task



(a) Interactions between Two Tasks

(b) Example of Global Feature

Li, Qi, and Heng Ji. "Incremental Joint Extraction of Entity Mentions and Relations." *ACL (1)*. 2014.

Joint Entity Relation Extraction

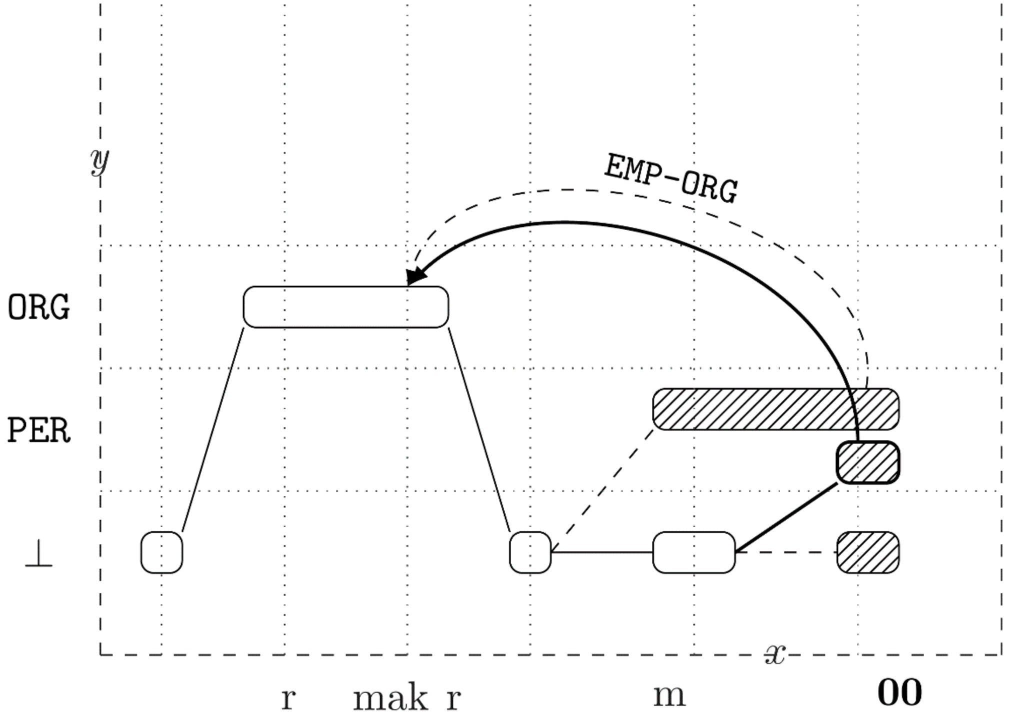
- A Single Model

$$\hat{y} = \operatorname*{argmax}_{y' \in \mathcal{Y}(x)} \mathbf{f}(x, y') \cdot \mathbf{w}$$

Li, Qi, and Heng Ji. "Incremental Joint Extraction of Entity Mentions and Relations." *ACL (1)*. 2014.

Joint Entity Relation Extraction

- Beam Search



Li, Qi, and Heng Ji. "Incremental Joint Extraction of Entity Mentions and Relations." *ACL (1)*. 2014.

Joint Entity Relation Extraction

- Feature
  - Local features
    - Gazetteer features
    - Case features
    - Contextual features
    - Parsing-based features
  - Global entity mention features
    - Coreference consistency
    - Neighbor coherence
    - Part-of-whole consistency
  - Global relation features
    - Role coherence
    - Triangle constraint
    - Inter-dependent compatibility
    - Neighbor coherence

Li, Qi, and Heng Ji. "Incremental Joint Extraction of Entity Mentions and Relations." *ACL (1)*. 2014.

Joint Entity Relation Extraction

- Experiments
  - Data:
    - Training data: ACE’05
    - Validation data: ACE’04

Li, Qi, and Heng Ji. "Incremental Joint Extraction of Entity Mentions and Relations." *ACL (1)*. 2014.

Joint Entity Relation Extraction

- Results on ACE

Model	Entity Mention (%)			Relation (%)			Entity Mention + Relation (%)		
Score	P	R	F_1	P	R	F_1	P	R	F_1
Pipeline	83.2	73.6	78.1	67.5	39.4	49.8	65.1	38.1	48.0
Joint w/ Local	84.5	76.0	80.0	68.4	40.1	50.6	65.3	38.3	48.3
Joint w/ Global	85.2	76.9	**80.8**	68.9	41.9	**52.1**	65.4	39.8	**49.5**
Annotator 1	91.8	89.9	90.9	71.9	69.0	70.4	69.5	66.7	68.1
Annotator 2	88.7	88.3	88.5	65.2	63.6	64.4	61.8	60.2	61.0
Inter-Agreement	85.8	87.3	86.5	55.4	54.7	55.0	52.3	51.6	51.9

Li, Qi, and Heng Ji. "Incremental Joint Extraction of Entity Mentions and Relations." *ACL (1)*. 2014.

Statistical Models

- Graph-Based Methods
- Transition-Based Methods

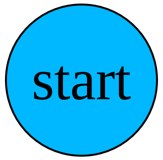
A Transition System

- Automata
  - State
    - Start state —— an empty structure
    - End state —— the output structure
    - Intermediate states —— partially constructed structures
  - Actions
    - Change one state to another

Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Transition System

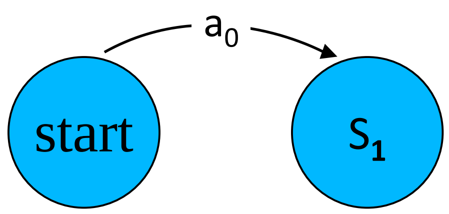
- Automata



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Transition System

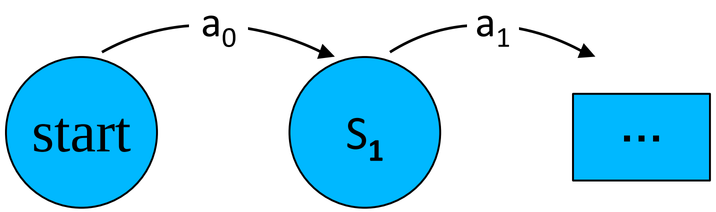
- Automata



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Transition System

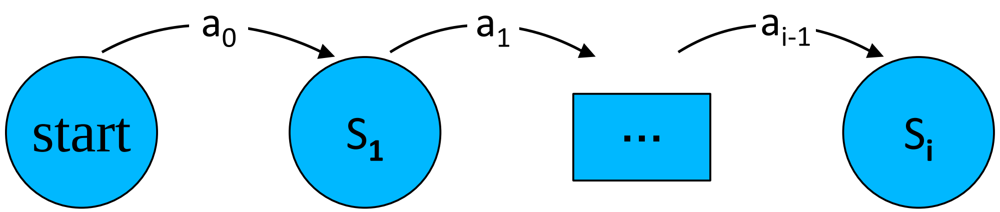
- Automata



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Transition System

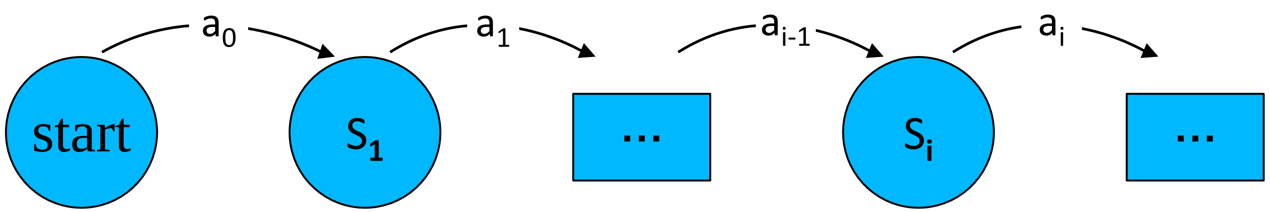
- Automata



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Transition System

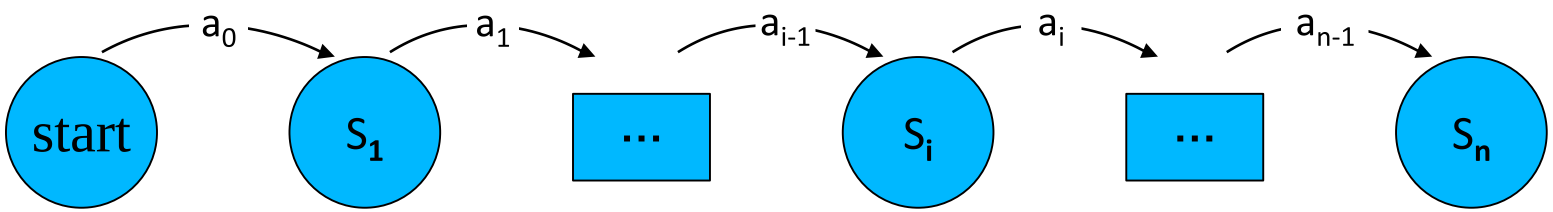
- Automata



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Transition System

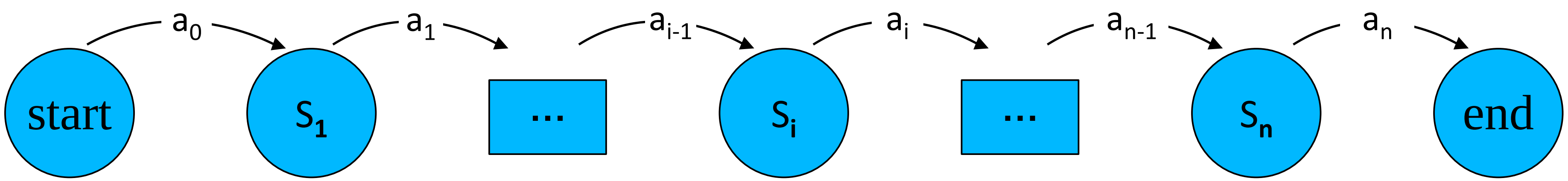
- Automata



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Transition System

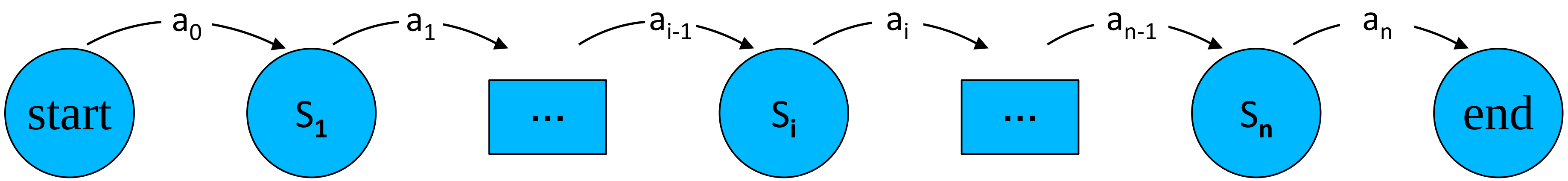
- Automata



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Transition System

- State
  - Corresponds to partial results during decoding
    - start state, end state, S_i



- Actions
  - The operations that can be applied for state transition
  - Construct output incrementally
    - a_i

Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Dependency Parsing

- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT

He does it here

Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Dependency Parsing

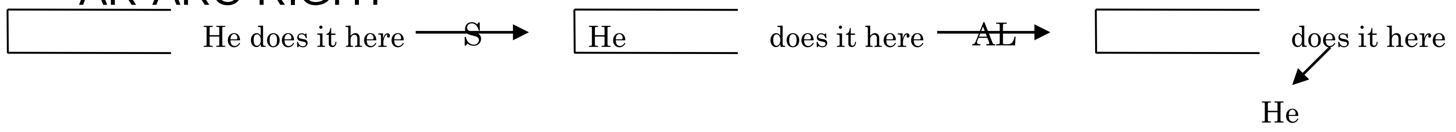
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT

[ ] He does it here $\xrightarrow{S}$ [He] does it here

Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Dependency Parsing

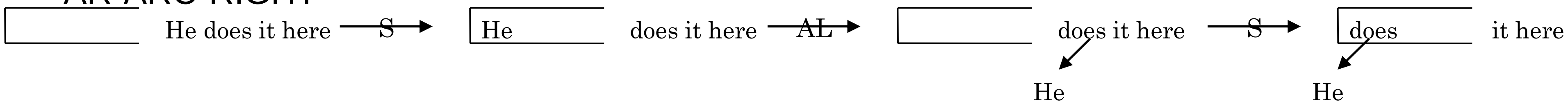
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Dependency Parsing

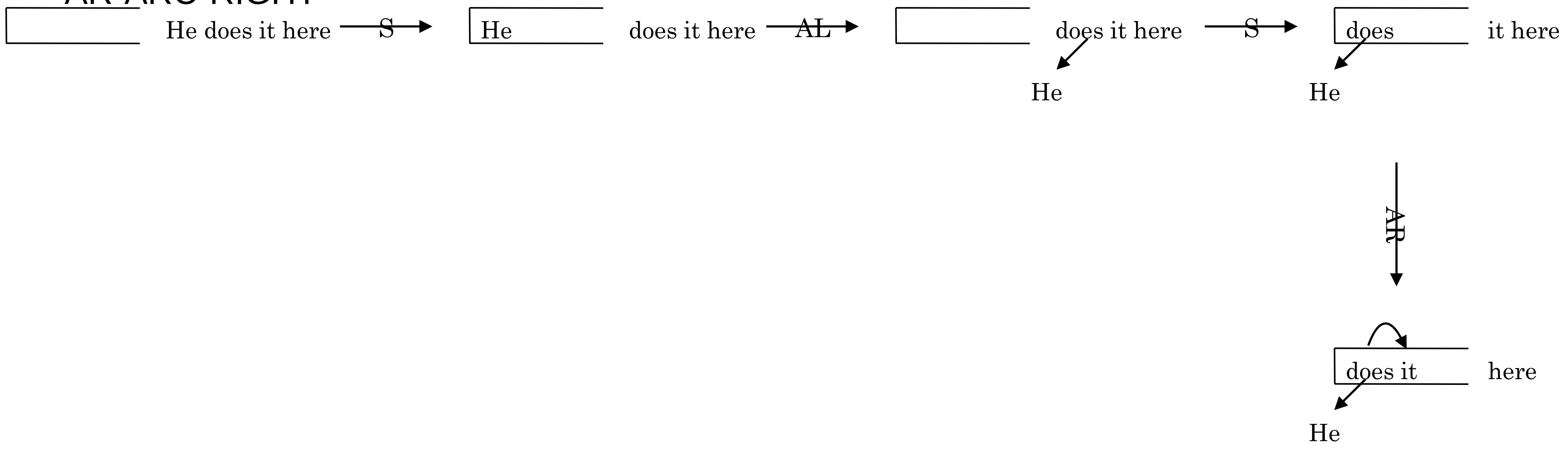
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Dependency Parsing

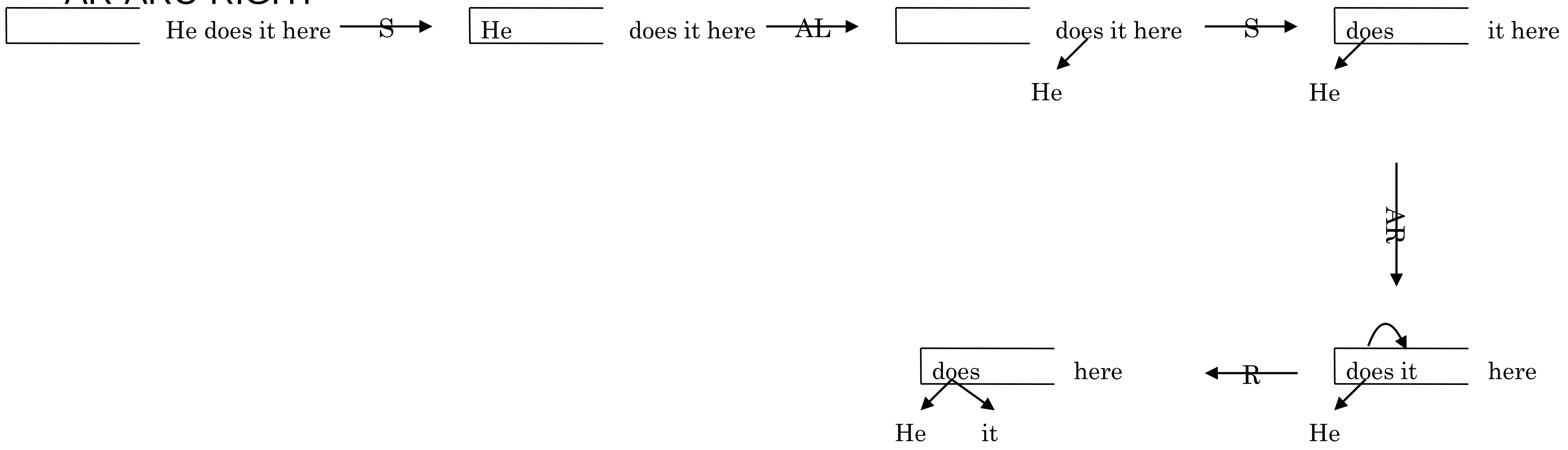
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Dependency Parsing

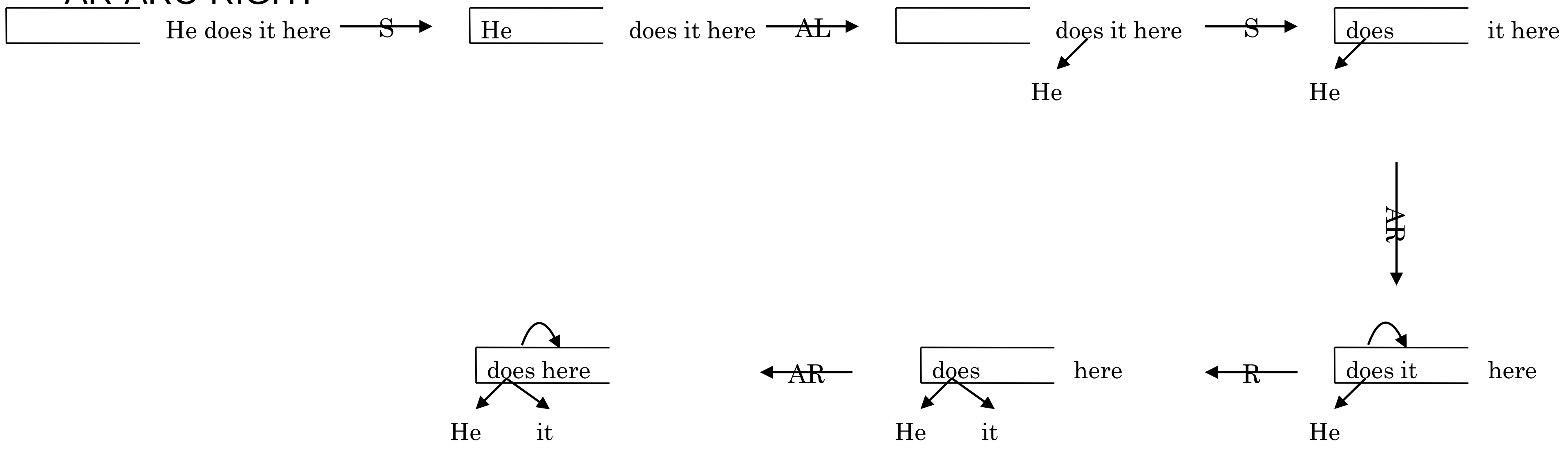
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Dependency Parsing

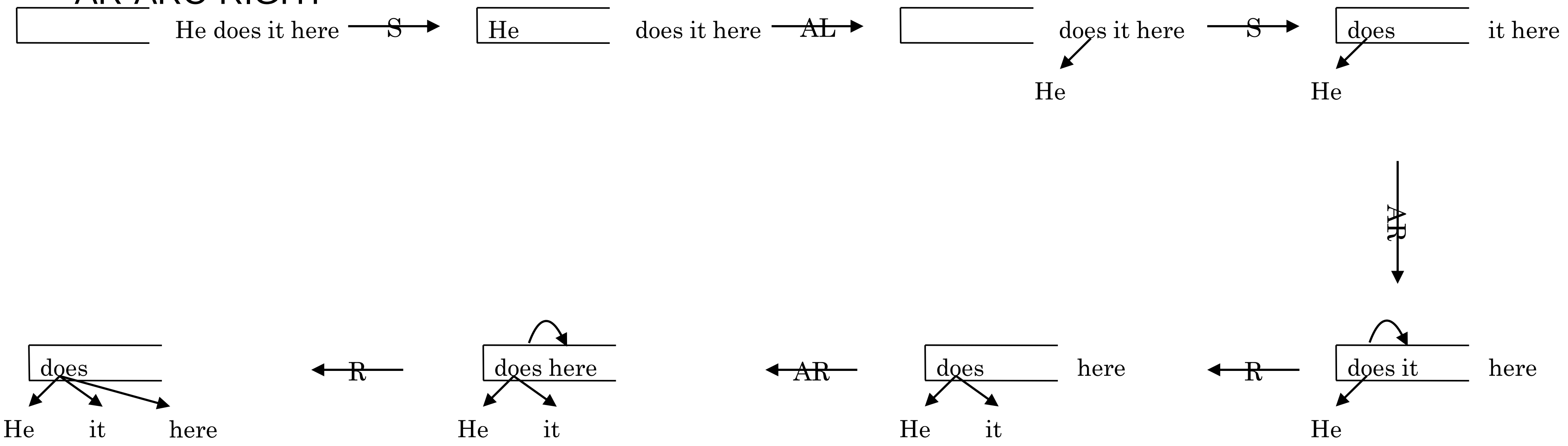
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Dependency Parsing

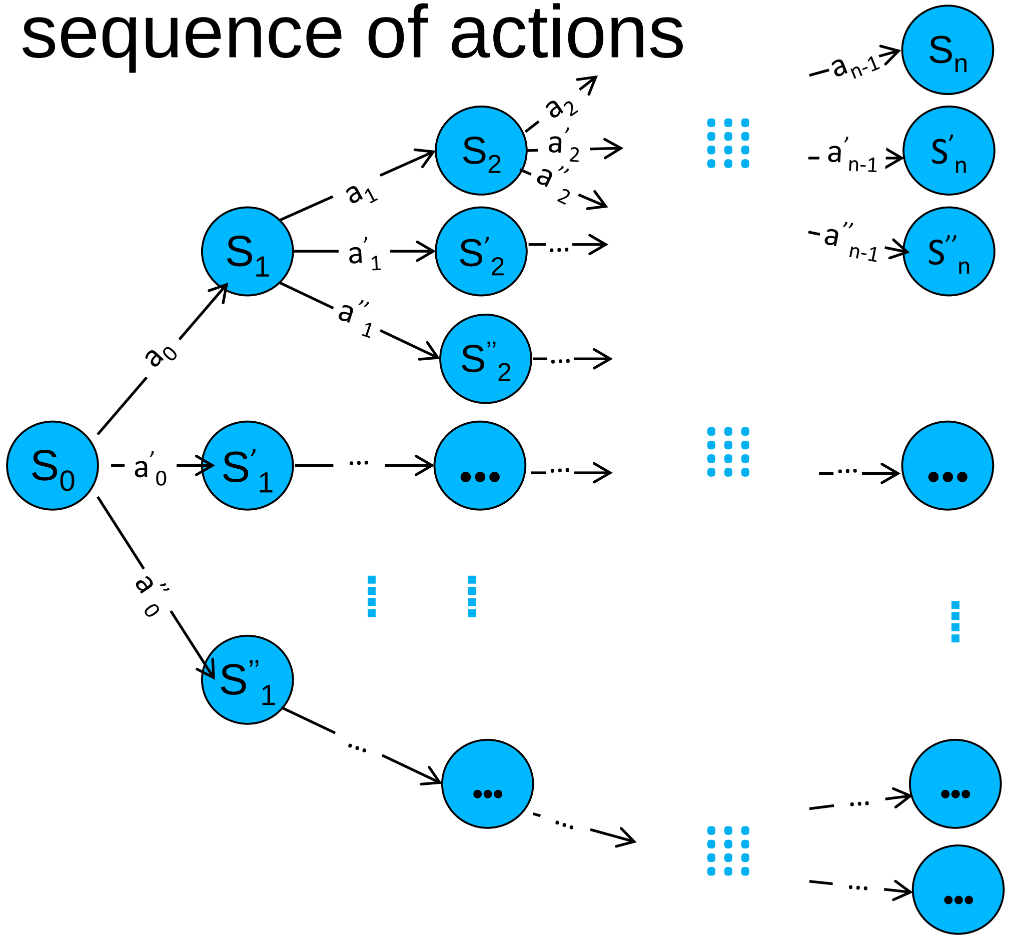
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

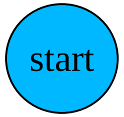
Search Space

- Find the best sequence of actions
- Exponential



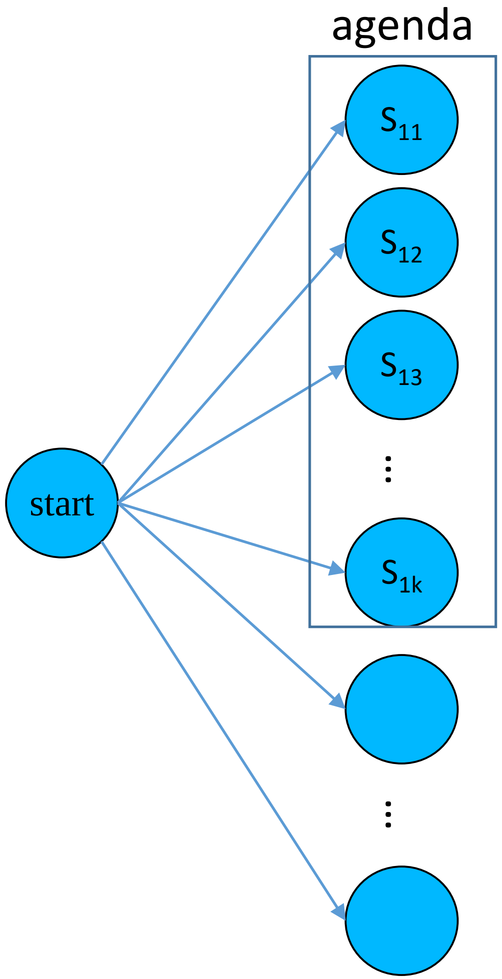
Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework



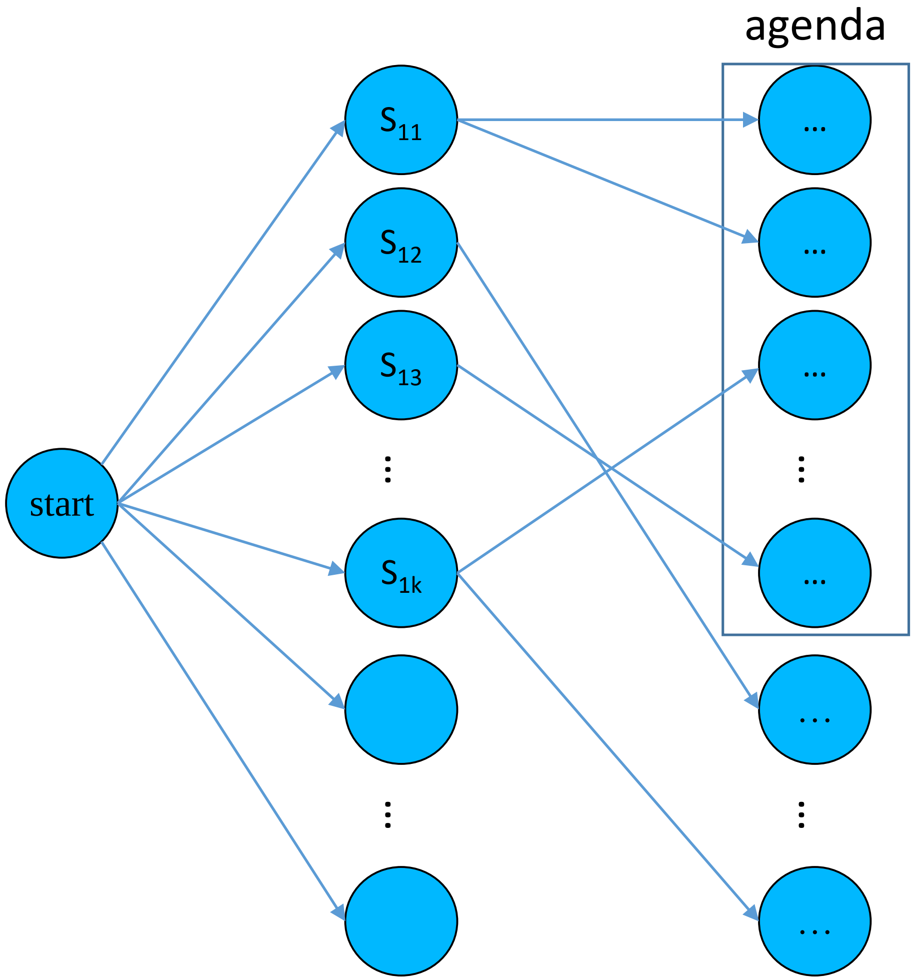
Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework



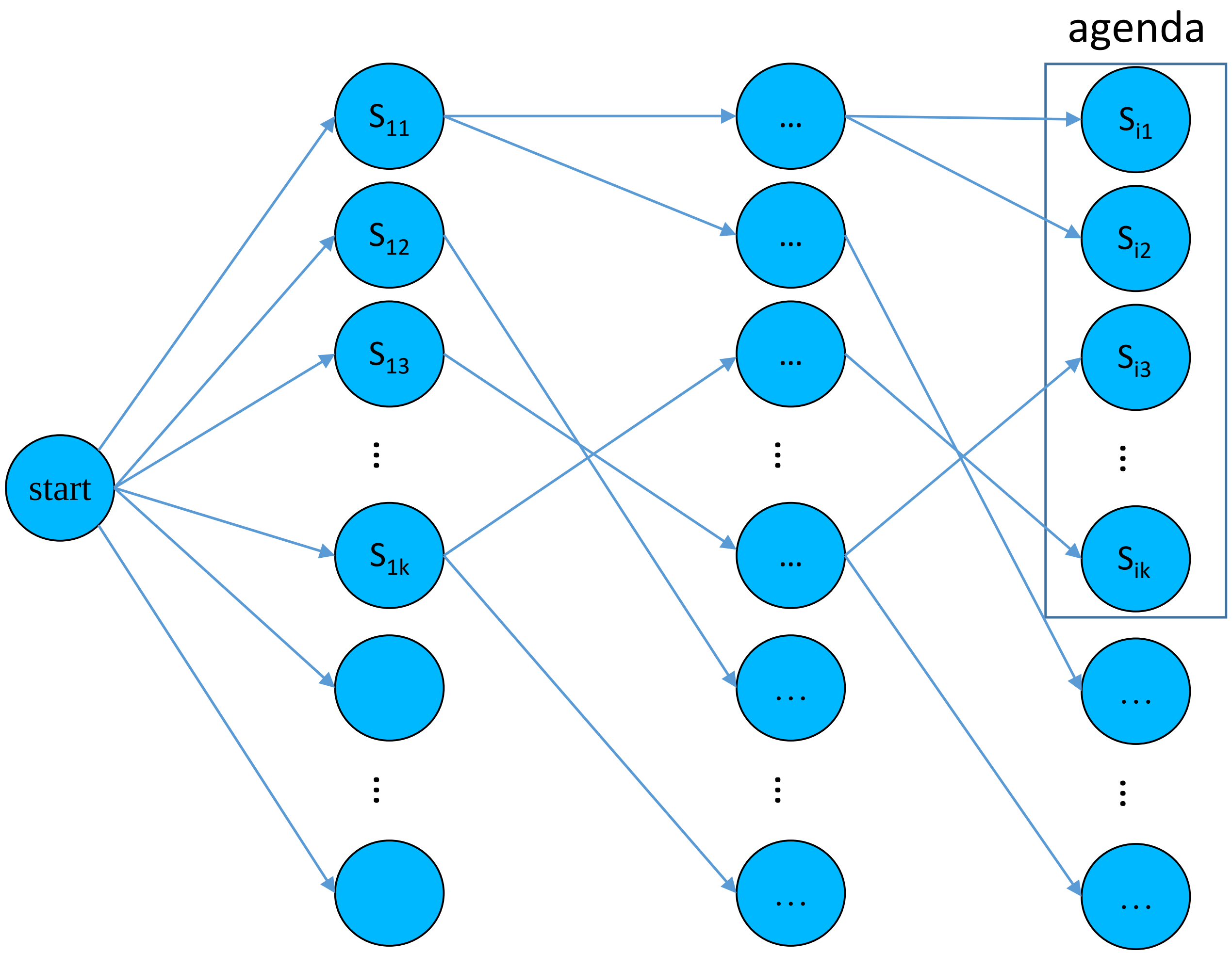
Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework



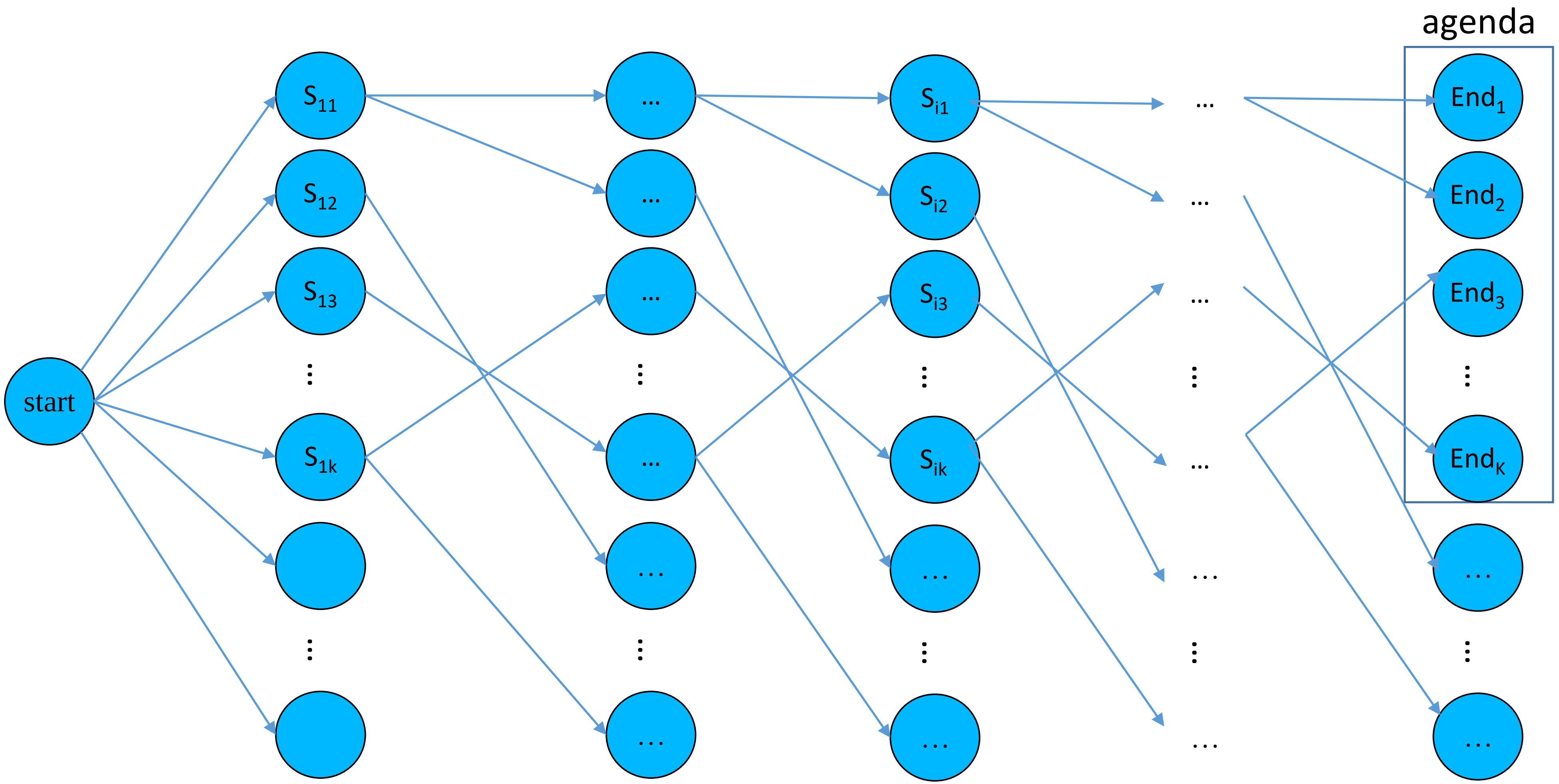
Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

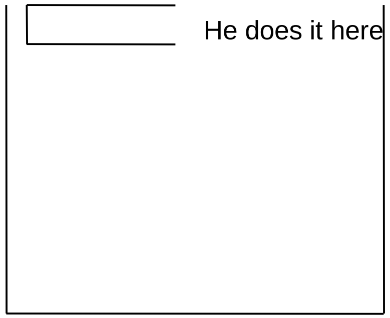
A Learning+Search Framework



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

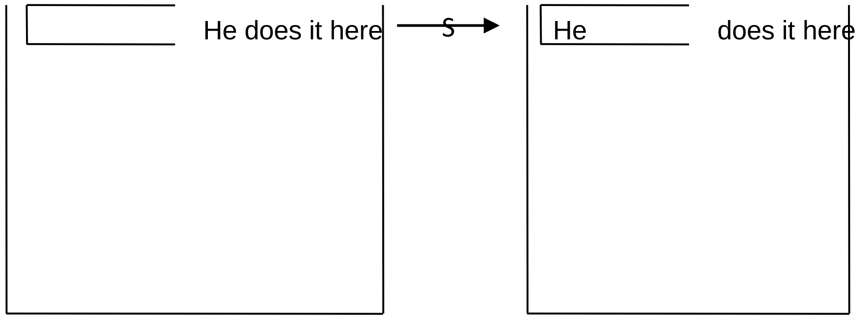
- Dependency Parsing Example
  - Decoding



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

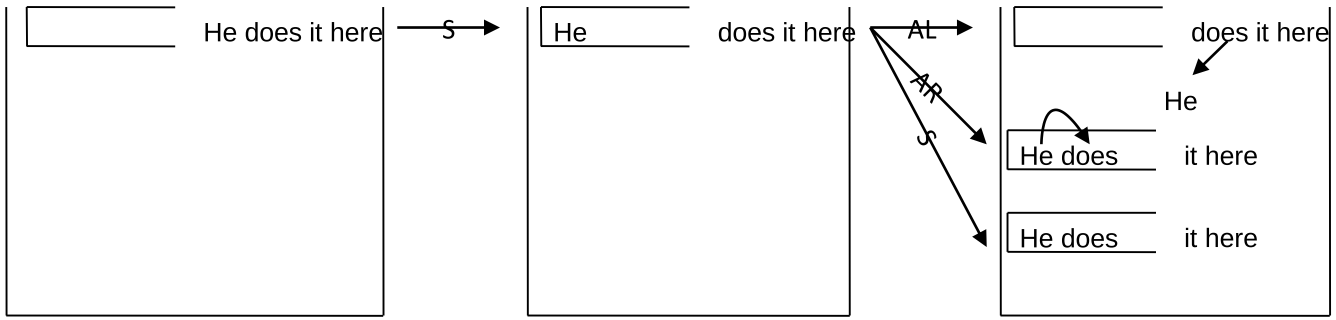
- Dependency Parsing Example
  - Decoding



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

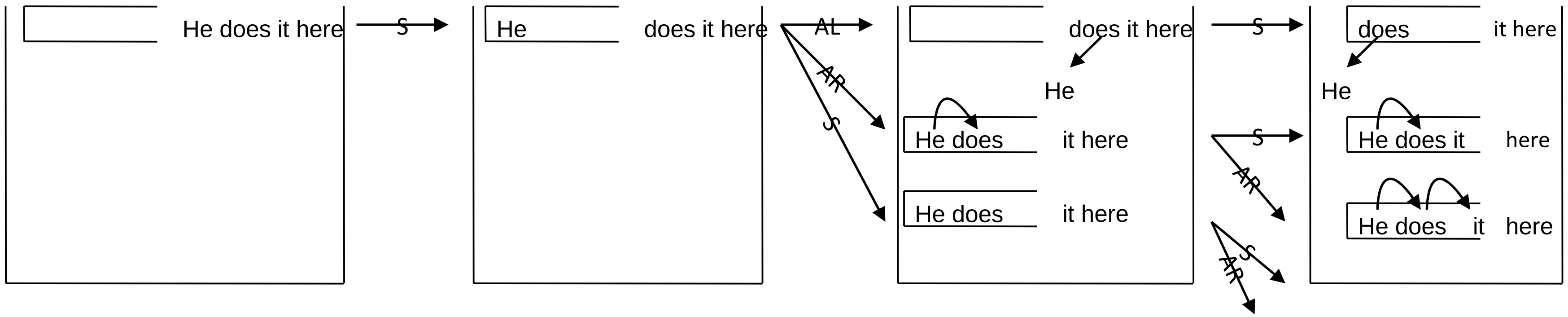
- Dependency Parsing Example
  - Decoding



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

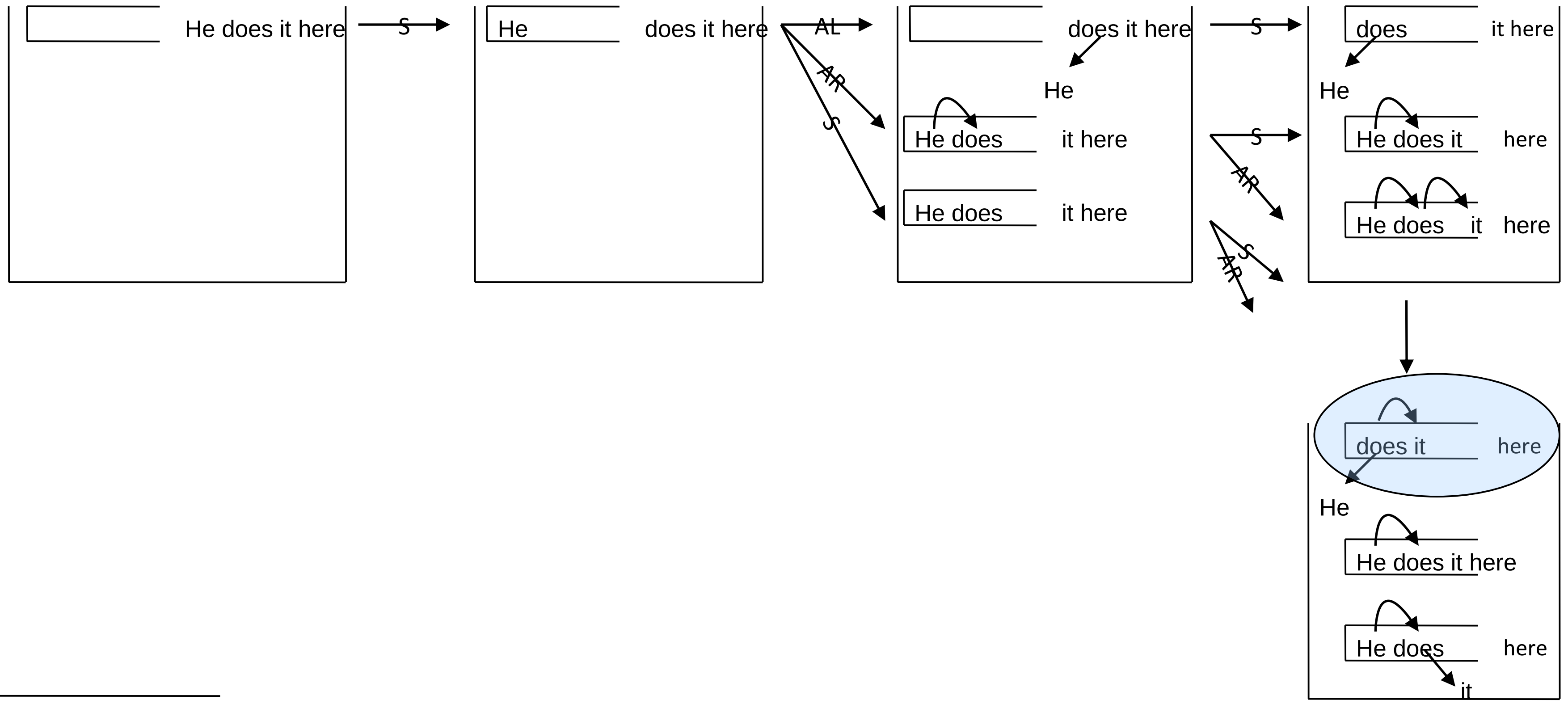
- Dependency Parsing Example
  - Decoding



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

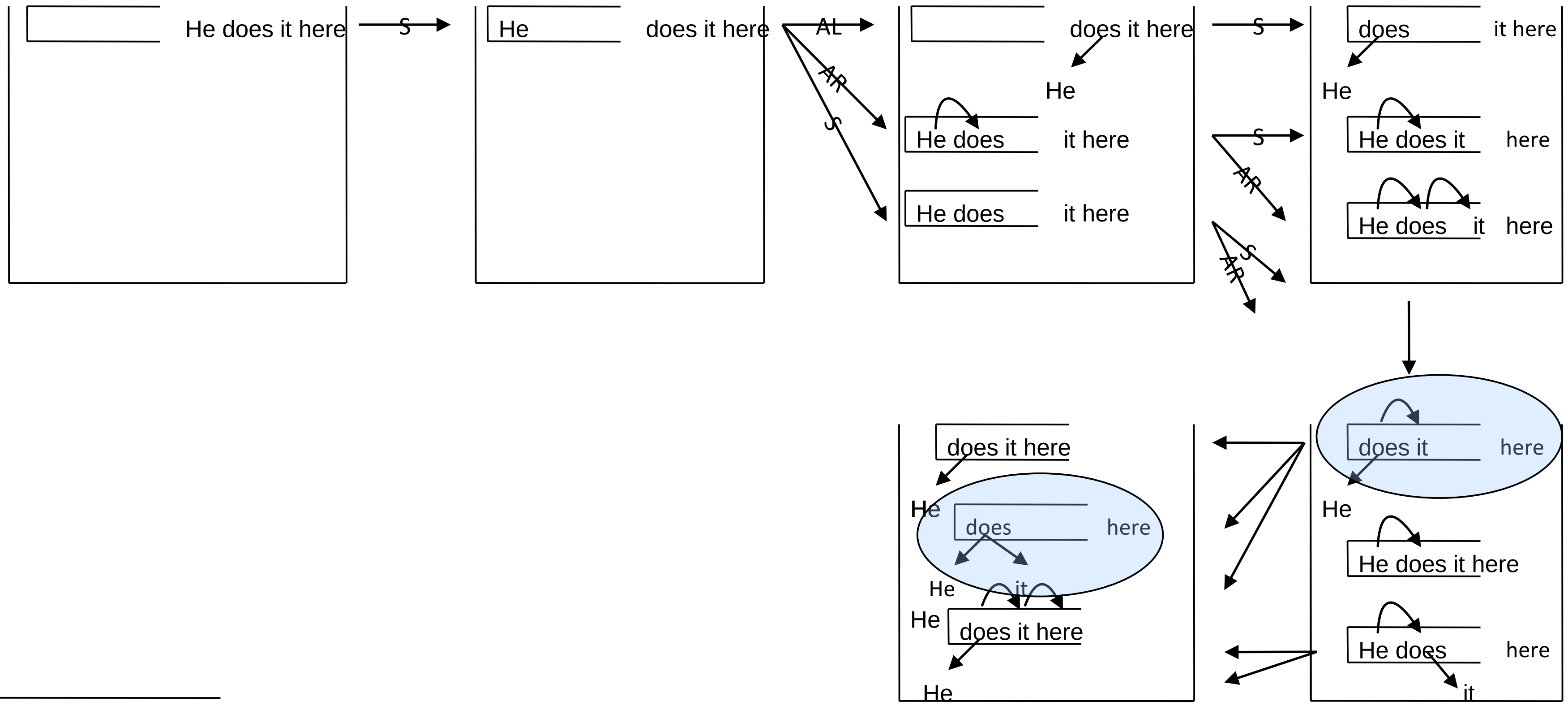
- Dependency Parsing Example
  - Decoding



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

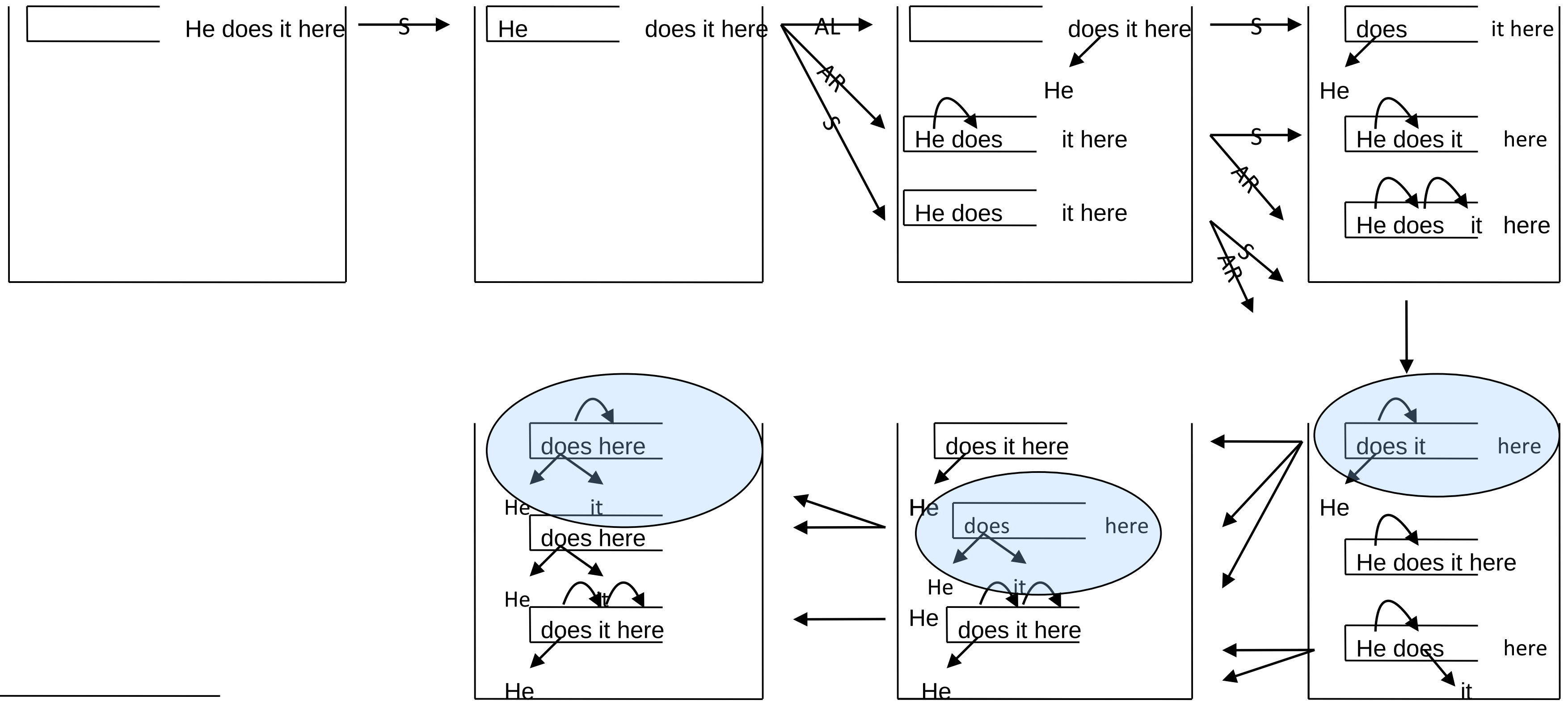
- Dependency Parsing Example
  - Decoding



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

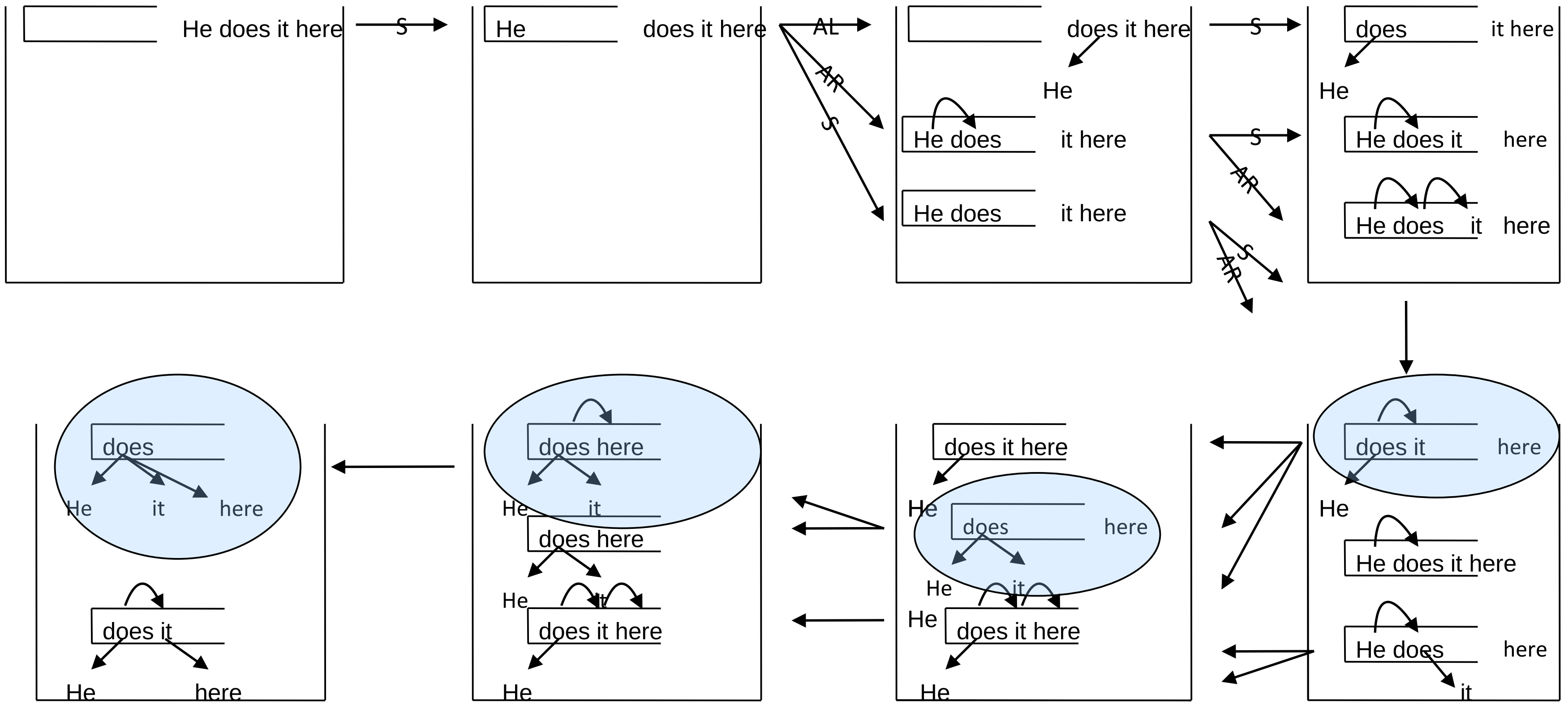
- Dependency Parsing Example
  - Decoding



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

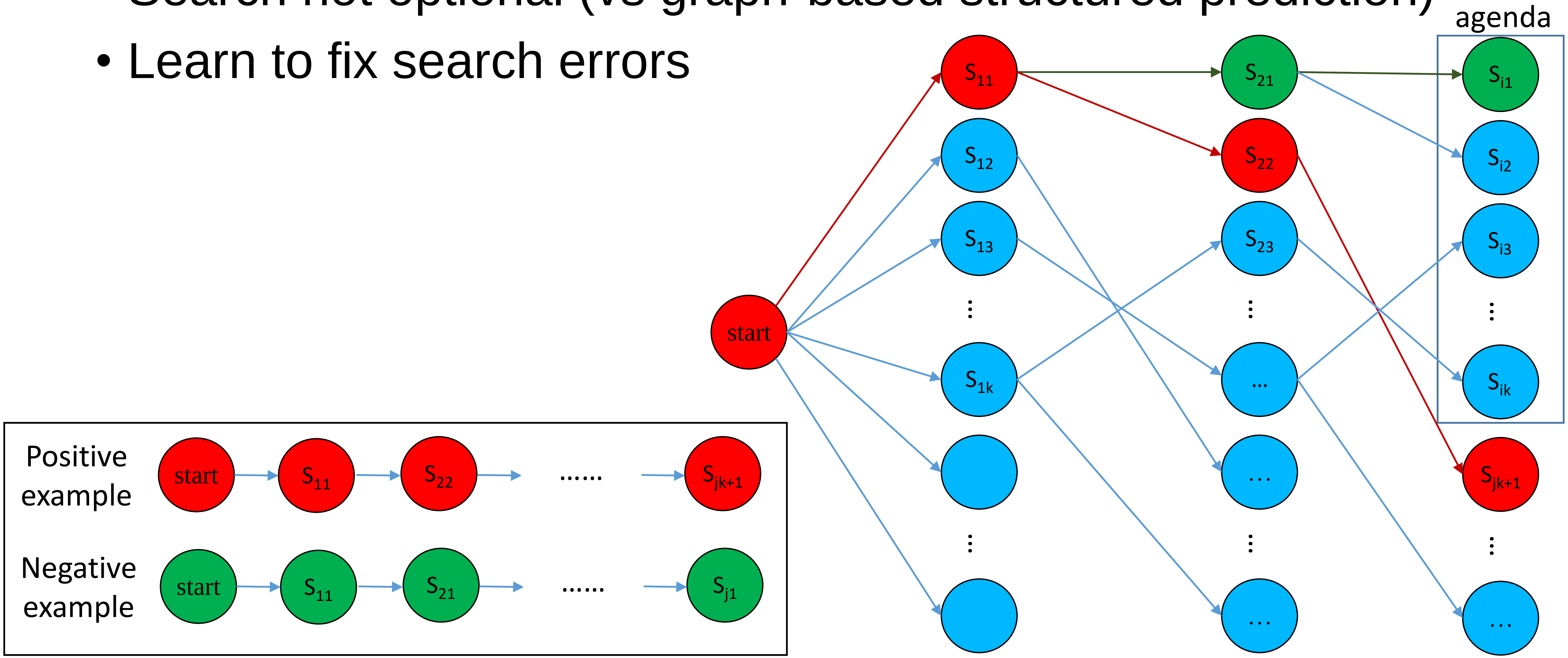
- Dependency Parsing Example
  - Decoding



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

- Search not optional (vs graph-based structured prediction)
- Learn to fix search errors



A Learning+Search Framework

- Advantages
  - Low computation complexity
  - Arbitrary non-local features
  - Learning-guided-search

Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

- State-of-the-art **accuracies** and **speeds**
  - Constituent parsing
  - Dependency parsing
  - Word Segmentation
  - CCG parsing
- Enable joint models
  - Address complex search space and use joint features, which have been difficult for traditional models

---

Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

- State-of-the-art **accuracies** and **speeds**
  - Constituent parsing
  - Dependency parsing
  - Word Segmentation
  - CCG parsing
- Enabl
  - and use joint features, which have
  - models

Joint Learning

Joint Search

Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

A Learning+Search Framework

- Global Normalization for Neural Structured Prediction
  - Zhou et al., (2015)
  - Watanabe et al., (2015)
  - Andor et al., (2016)
  - Rush et al., (2016)

Hao Zhou, Yue Zhang, Shujian Huang and Jiajun Chen. A Neural Probabilistic Structured-Prediction Model for Transition-based Dependency Parsing. In Proceedings of ACL 2015, Beijing, China, July.

Watanabe, Taro, and Eiichiro Sumita. "Transition-based neural constituent parsing." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Vol. 1. 2015.

Andor Daniel, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, Michael Collins "Globally normalized transition-based neural networks." arXiv preprint arXiv:1603.06042 (2016).

Wiseman, Sam, and Alexander M. Rush. "Sequence-to-sequence learning as beam-search optimization." arXiv preprint arXiv:1606.02960 (2016).

Joint Segmentation and POS Tagging

- The transition system
  - State
    - Partial segmented results
    - Unprocessed characters
  - Two actions
    - Separate (t) : t is a POS tag
    - Append

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- The transition system
  - Initial state

我喜欢读书

我/PN 喜欢/V 读/V 书/N
[I] [like] [reading] [books]

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- The transition system
  - Separate(PN)

我/PN	喜欢读书

我/PN 喜欢/V 读/V 书/N
[I] [like] [reading] [books]

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- The transition system
  - Separate (V)

我/PN 喜/V	欢读书

我/PN 喜欢/V 读/V 书/N
[I] [like] [reading] [books]

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- The transition system
  - Append

我/PN 喜欢/V	读书

我/PN 喜欢/V 读/V 书/N
[I] [like] [reading] [books]

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- The transition system
  - Separate (V)

我/PN 喜欢/V 读/V

书

我/PN 喜欢/V 读/V 书/N
[I] [like] [reading] [books]

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- The transition system
  - Separate (N)

我/PN 喜欢/V 读/V 书/N

我/PN	喜欢/V	读/V	书/N
[I]	[like]	[reading]	[books]

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- The transition system
  - End state

我/PN　喜欢/V　读/V　书/N

我/PN 喜欢/V 读/V 书/N
[I] [like] [reading] [books]

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- Segmentation Feature templates

Feature templates for the word segmentor.

	Feature template	When c_0 is
1	w_{-1}	separated
2	$w_{-1}w_{-2}$	separated
3	w_{-1}, where $len(w_{-1}) = 1$	separated
4	$start(w_{-1})len(w_{-1})$	separated
5	$end(w_{-1})len(w_{-1})$	separated
6	$end(w_{-1})c_0$	separated
7	$c_{-1}c_0$	appended
8	$begin(w_{-1})end(w_{-1})$	separated
9	$w_{-1}c_0$	separated
10	$end(w_{-2})w_{-1}$	separated
11	$start(w_{-1})c_0$	separated
12	$end(w_{-2})end(w_{-1})$	separated
13	$w_{-2}len(w_{-1})$	separated
14	$len(w_{-2})w_{-1}$	separated

w = word; c = character. The index of the current character is 0.

Non-local

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- POS Feature templates

POS feature templates for the joint segmentor and POS-tagger.

	Feature template	when c_0 is
1	$w_{-1}t_{-1}$	separated
2	$t_{-1}t_0$	separated
3	$t_{-2}t_{-1}t_0$	separated
4	$w_{-1}t_0$	separated
5	$t_{-2}w_{-1}$	separated
6	$w_{-1}t_{-1}end(w_{-2})$	separated
7	$w_{-1}t_{-1}c_0$	separated
8	$c_{-2}c_{-1}c_0t_{-1}$, where $len(w_{-1}) = 1$	separated
9	c_0t_0	separated
10	$t_{-1}start(w_{-1})$	separated
11	t_0c_0	separated or appended
12	$c_0t_0start(w_0)$	appended
13	$ct_{-1}end(w_{-1})$, where $c \in w_{-1}$ and $c \neq end(w_{-1})$	separated
14	$c_0t_0cat(start(w_0))$	separated
15	$ct_{-1}cat(end(w_{-1}))$, where $c \in w_{-1}$ and $c \neq end(w_{-1})$	appended
16	$c_0t_0c_{-1}t_{-1}$	separated
17	$c_0t_0c_{-1}$	appended

w = word; c = character; t = POS-tag. The index of the current character is 0.

Word-level

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation and POS Tagging

- Experiments on CTB 5

	SF	JF
K09 (error-driven)	97.87	93.67
This work	**97.78**	**93.67**
Zhang 2008	97.82	93.62
K09 (baseline)	97.79	93.60
J08a	97.85	93.41
J08b	97.74	93.37
N07	97.83	93.32

SF = segmentation F-score; JF = joint segmentation and POS-tagging F-score

Yue Zhang and Stephen Clark. *A Fast Decoder for Joint Word Segmentation and POS-tagging Using a Single Discriminative Model*. In proceedings of EMNLP 2010. Massachusetts, USA. October.

Joint Segmentation/Tagging/Chunking

- Input 他到达北京机场。
  Output [NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
  [He] [arrived] [Beijing airport] [.]
- Chunking knowledge can potentially improve segmentation/tagging.
- Chunk cluster features to avoid sparsity.

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: initial state

chunk buffer	word buffer	character buffer
		他 到 达 北 京 机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: SEP(NR)

chunk buffer	word buffer	character buffer
	[他/NR	到 达 北 京 机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: FIN W

chunk buffer | word buffer | character buffer

chunk buffer	word buffer	character buffer
	[他/NR]	到 达 北 京 机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: SEP(NP)

chunk buffer	word buffer	character buffer
[NP 他/NR]		到 达 北 京 机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: SEP(VV)

chunk buffer	word buffer	character buffer
[NP 他/NR]	[到/VV	达 北 京 机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: APP W

chunk buffer	word buffer	character buffer
[NP 他/NR]	[到达/VV	北 京 机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: FIN W

chunk buffer	word buffer	character buffer
[NP 他/NR]	[到达/VV]	北 京 机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: SEP(VP)

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV]		北 京 机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: SEP(NR)

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV]	[北/NR	京 机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: APP W

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV]	[北京/NR	机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: FIN W

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV]	[北京/NR]	机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: SEP(NP)

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV] [NP 北京/NR]		机 场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: SEP(NN)

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV] [NP 北京/NR]	[机/NN	场 。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: APP W

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV] [NP 北京/NR]	[机场/NN	。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: FIN W

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV] [NP 北京/NR]	[机场/NN]	。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: APP C

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN]		。

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: SEP(PU)

chunk buffer

[NP 他/NR]
[VP 到达/VV]
[NP 北京/NR 机场/NN]

word buffer

[。/PU

character buffer

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: FIN W

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN]	[。/PU]	

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking
  - Action: SEP(O)

chunk buffer	word buffer	character buffer
[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]		

[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]
[He] [arrived] [Beijing airport] [.]

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Character-based chunking feature template

ID	**Feature Templates**
1	C_0
2	$C_0 \cdot T_0$
3	$C_0 \cdot POSset(C_0)$
4	C_0, where $len(C_0) = 1$
5	$C_0 \cdot N_0 w$
6	$C_0 \cdot N_0 w \cdot T_0$
7	$C_{-1} \cdot C_0$
8	$T_{-1} \cdot C_0$
9	$C_{-1} \cdot T_0$
10	$C_0 \cdot end\_word(C_{-1})$
11	$C_{-1} \cdot len(C_0)$
12	$C_0 \cdot len(C_{-1})$
13	$C_0 \cdot end\_word(C_{-1}) \cdot T0$
14	$C_{-1} \cdot T_{-1} \cdot C_0 \cdot T_0$
15	$w_{-2} \cdot w_{-1}$

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation/Tagging/Chunking

- Results on CTB

	SEG	POS	CHUNK
Pipeline	88.81	80.64	69.02
Pipeline-C	88.81	80.64	68.82
Pipeline-Semi-C	88.81	80.64	69.45
Joint	89.85	81.94	70.96
Joint-C	89.83	81.78	70.63
Joint-Semi-C	**90.67**	**82.45**	**72.09**

Chen Lyu, Yue Zhang and Donghong Ji. *Joint Word Segmentation, POS-Tagging and Syntactic Chunking.* In Proceedings of the AAAI 2016, Phoenix, Arizona, USA, February.

Joint Segmentation, Tagging and Normalization

- Text normalization is introduced as a pre-processing step for microblog processing, which transforms informal words into their standard forms. For example, “tmrw” has been frequently used in tweets for is for “tomorrow”.
- Task

pear

工作鸭梨大啊！ ⇒ 工作/NN 压力/NN 大/VA 啊/SP !/PU

work stress big

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Normalization dictionary

鸭梨- 压力
pear - pressure
孩纸- 孩子
child paper - child
围脖- 微博
neckerchief - microblog
盆友- 朋友
basin friend - friend
…………

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Transition actions for joint segmentation, tagging and normalization
  - Actions: initial state

word buffer

character buffer

工 作 鸭 梨 大 啊 ！

工作/NN 压力/NN 大/VA 啊/SP !/PU
Work stress big ah !

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Transition actions for joint segmentation, tagging and normalization
  - Actions: SEP(工, NN)

word buffer	character buffer
工/NN	作 鸭 梨 大 啊 ！

工作/NN 压力/NN 大/VA 啊/SP !/PU
Work stress big ah !

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Transition actions for joint segmentation, tagging and normalization
  - Actions: APP(作)

word buffer	character buffer
工作/NN	鸭 梨 大 啊 ！

工作/NN 压力/NN 大/VA 啊/SP !/PU
Work stress big ah !

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Transition actions for joint segmentation, tagging and normalization
  - Actions: SEP(鸭, NN)

word buffer

工作/NN 鸭/NN

character buffer

梨 大 啊 ！

工作/NN 压力/NN 大/VA 啊/SP !/PU
Work stress big ah !

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Transition actions for joint segmentation, tagging and normalization
  - Actions: APP(梨)

word buffer

工作/NN 鸭梨/NN

character buffer

大 啊 ！

工作/NN 压力/NN 大/VA 啊/SP !/PU
Work stress big ah !

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Transition actions for joint segmentation, tagging and normalization
  - Actions: SEPS(大, VA, 压力)

word buffer

工作/NN 压力/NN 大/VA

鸭梨/NN

character buffer

啊 ！

工作/NN 压力/NN 大/VA 啊/SP !/PU

Work stress big ah !

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Transition actions for joint segmentation, tagging and normalization
  - Actions: SEP(啊, SP)

word buffer

工作/NN 压力/NN 大/VA 啊/SP

character buffer

!

工作/NN 压力/NN 大/VA 啊/SP !/PU
Work stress big ah !

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Transition actions for joint segmentation, tagging and normalization
  - Actions: SEP(！, PU)

word buffer

character buffer

工作/NN 压力/NN 大/VA 啊/SP ！/PU

工作/NN	压力/NN	大/VA	啊/SP	!/PU
Work	stress	big	ah	!

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

- Features
  - The segmentation feature templates of Zhang and Clark (2011)
  - Extracting language model features by using word-based language model learned from a large quantity of standard texts

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, Tagging and Normalization

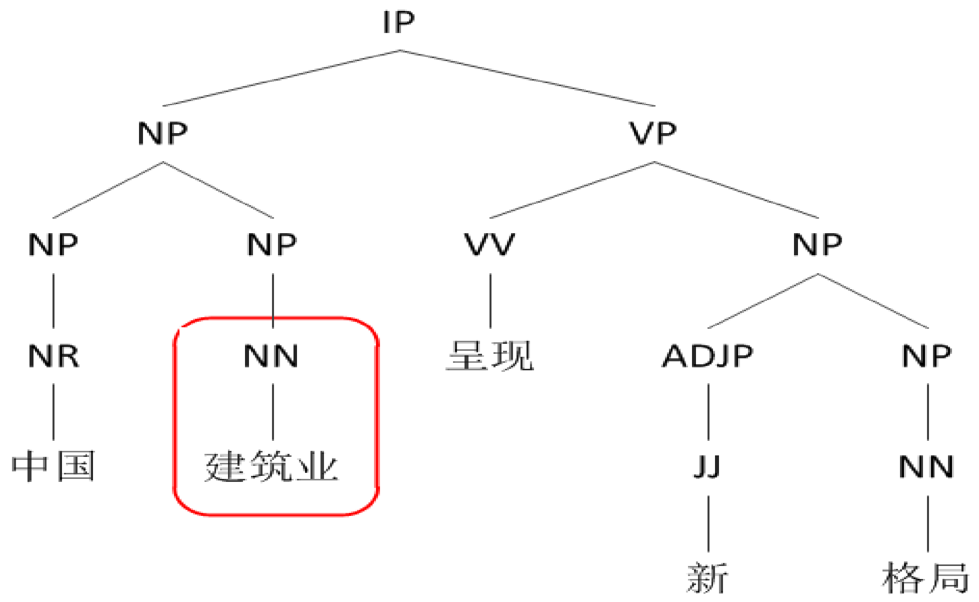
- Results on CTB

	Seg-F	POS-F	Nor-F
Stanford	**0.9058**	0.8163	
ST	0.8934	0.8263	
S;N;T	0.8885	0.8197	0.4058
SN;T	0.8945	0.8287	0.4207
SNT	0.8995	**0.8296**	**0.4391**
ST+lm	0.9162	0.8401	
S;N;T+lm	0.9132	0.8341	0.6276
SN;T+lm	0.9240	0.8439	0.6392
SNT+lm	**0.9261**	**0.8459**	**0.6413**

Tao Qian, Yue Zhang, Meishan Zhang and Donghong Ji. *A Transition-based Model for Joint Segmentation, POS-tagging and Normalization*. In proceedings of EMNLP 2015, Lisboa, Portugal, September.

Joint Segmentation, POS-tagging and Constituent Parsing

- Traditional: word-based Chinese parsing

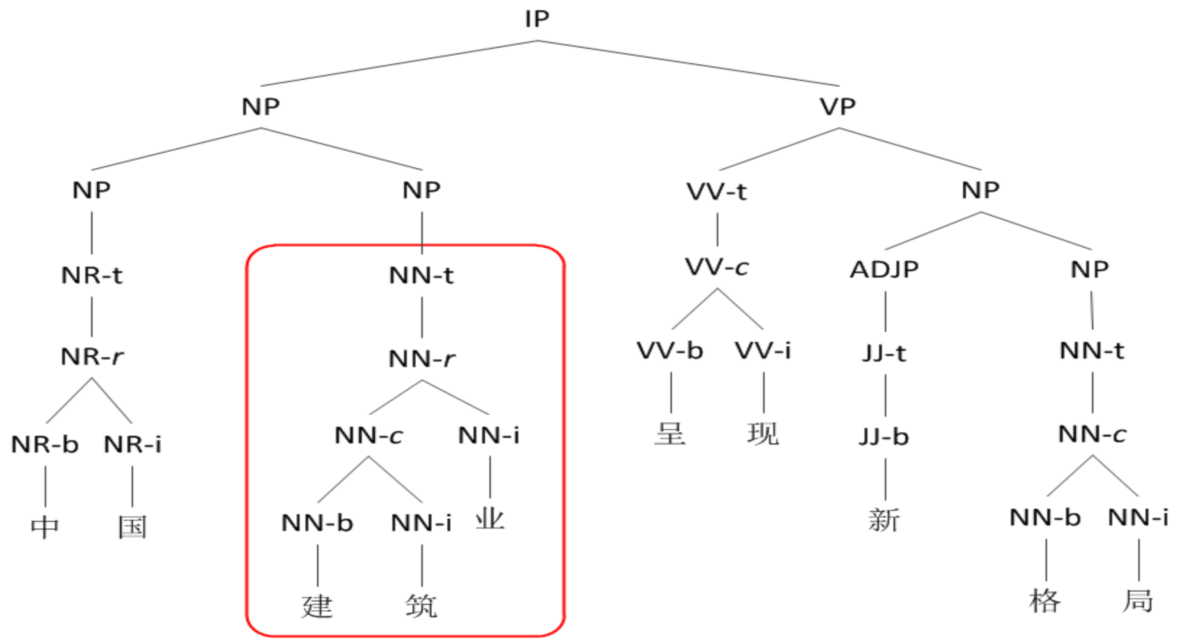


CTB-style word-based syntax tree for “中国 (China) 建筑业 (architecture industry) 呈现 (show) 新 (new) 格局 (pattern)”.

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

- This: character-based Chinese parsing



Character-level syntax tree with hierarchal word structures for “中 (middle) 国 (nation) 建 (construction) 筑 (building) 业 (industry) 呈 (present) 现 (show) 新 (new) 格 (style) 局 (situation)”.

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

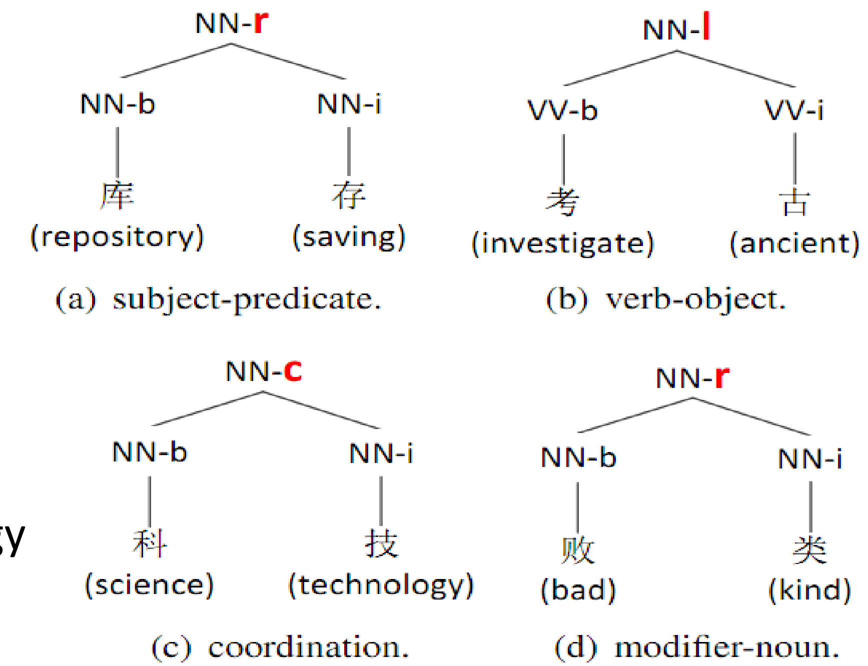
- Why character-based?
  - Chinese words have syntactic structures.

库存：stock

考古：archaeology

科技：Science and technology

败类：scum



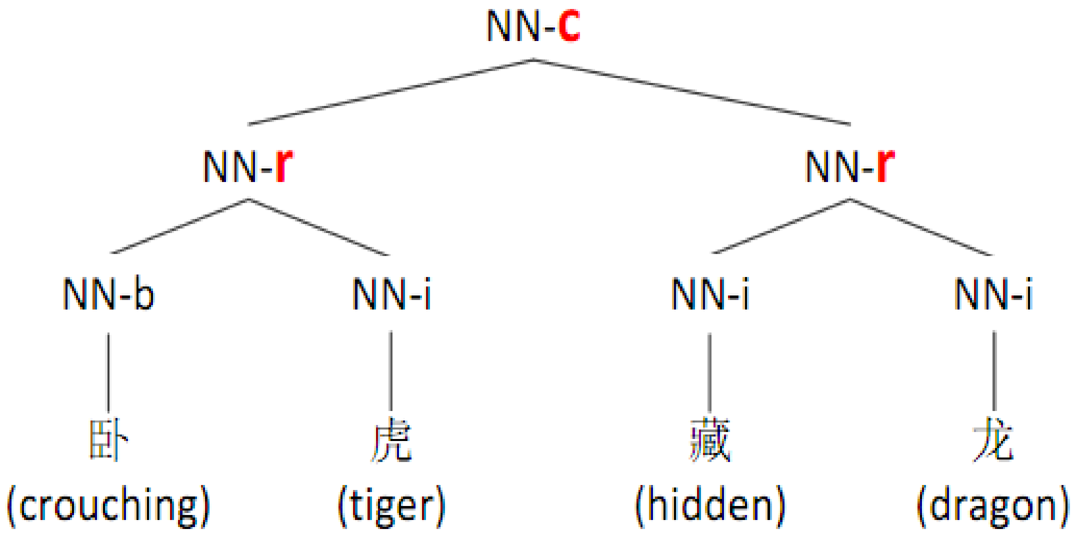
(a) subject-predicate. (b) verb-object.

(c) coordination. (d) modifier-noun.

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

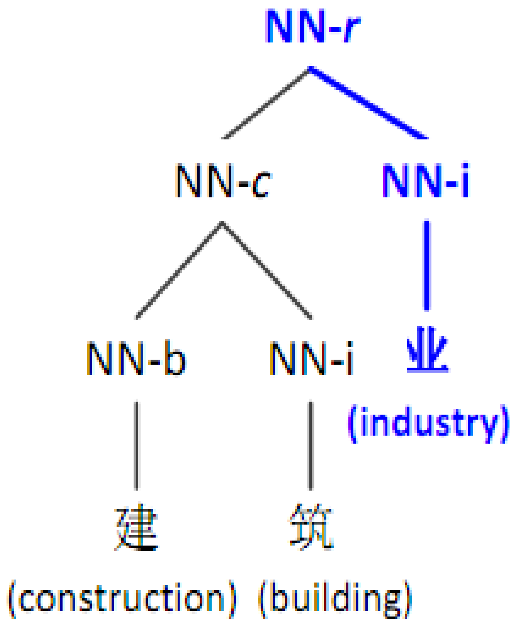
- Why character-based?
  - Chinese words have syntactic structures.



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

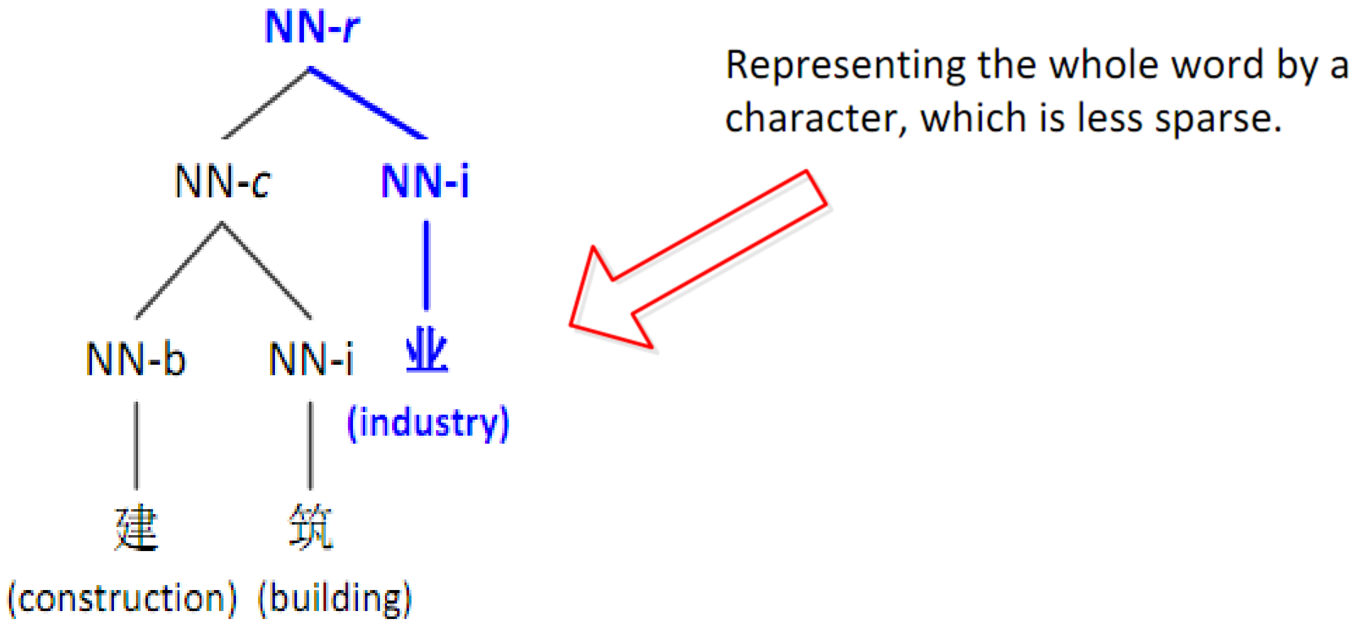
- Why character-based?
  - Deep character information of word structures.



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

- Why character-based?
  - Deep character information of word structures.



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

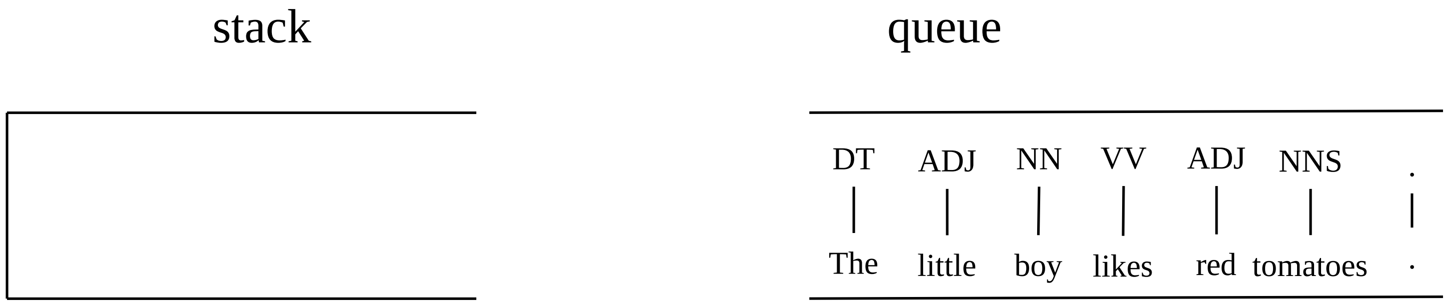
Joint Segmentation, POS-tagging and Constituent Parsing

- The character-based parsing model
  - A transition-based parser

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Transition-based Constituent Parsing

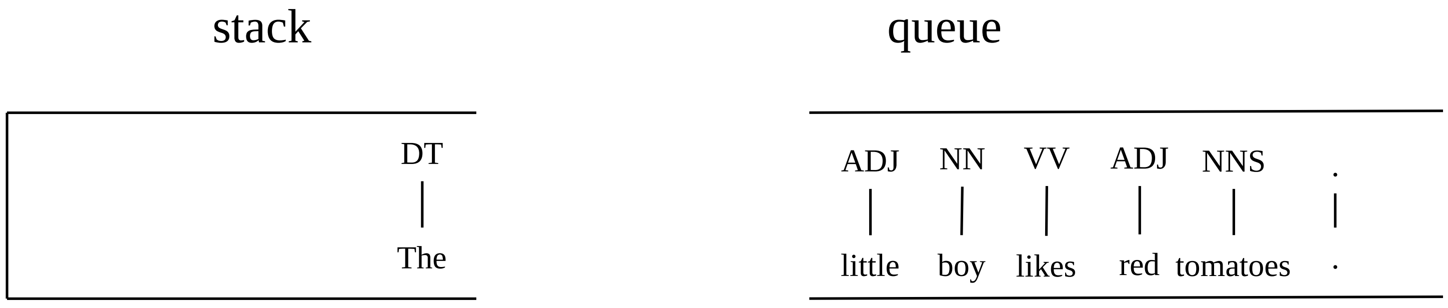
- Example
  - SHIFT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

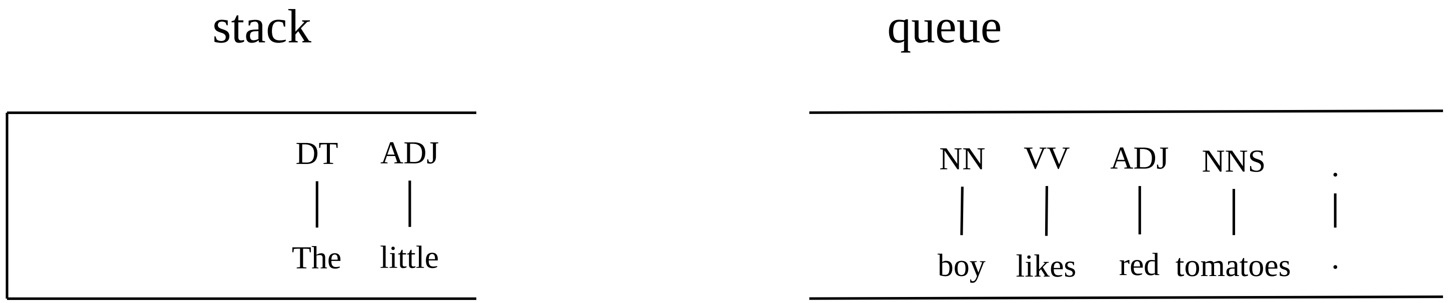
- Example
  - SHIFT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

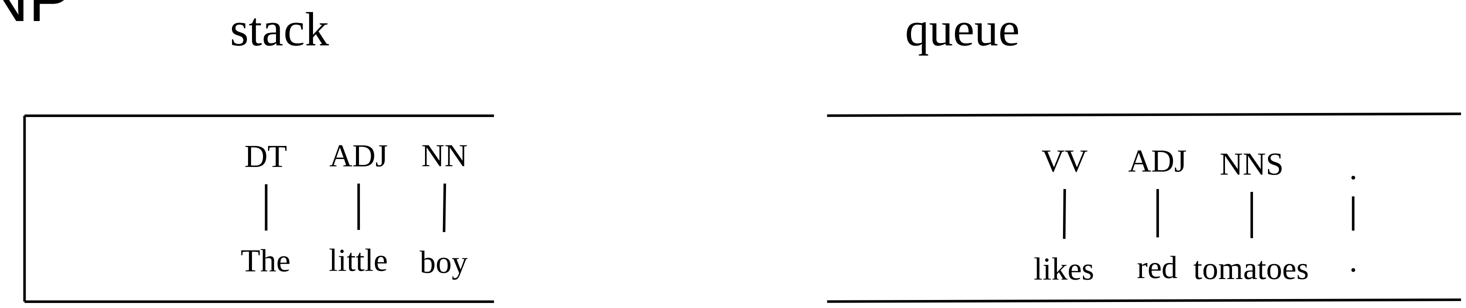
- Example
  - SHIFT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

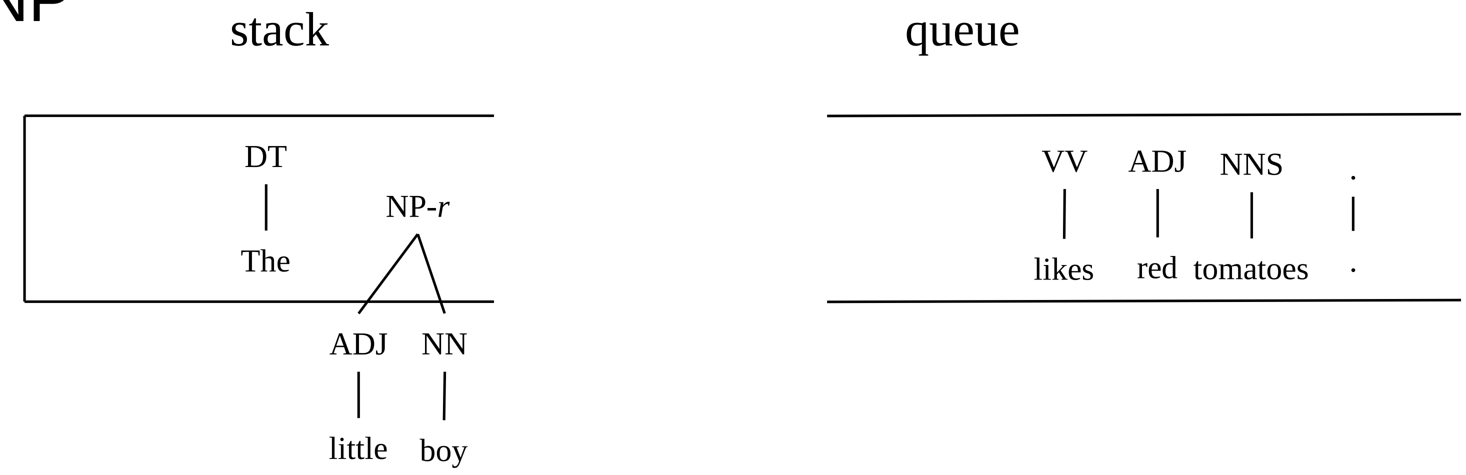
- Example
  - REDUCE-R-NP



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

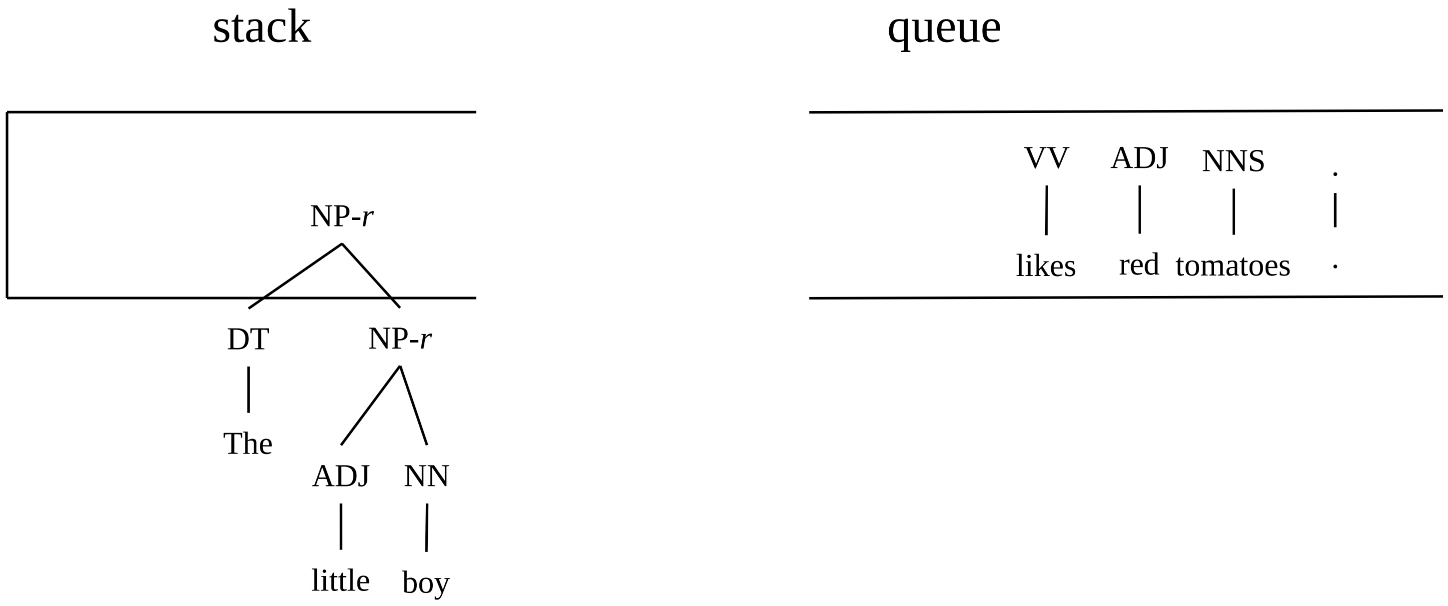
- Example
  - REDUCE-R-NP



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

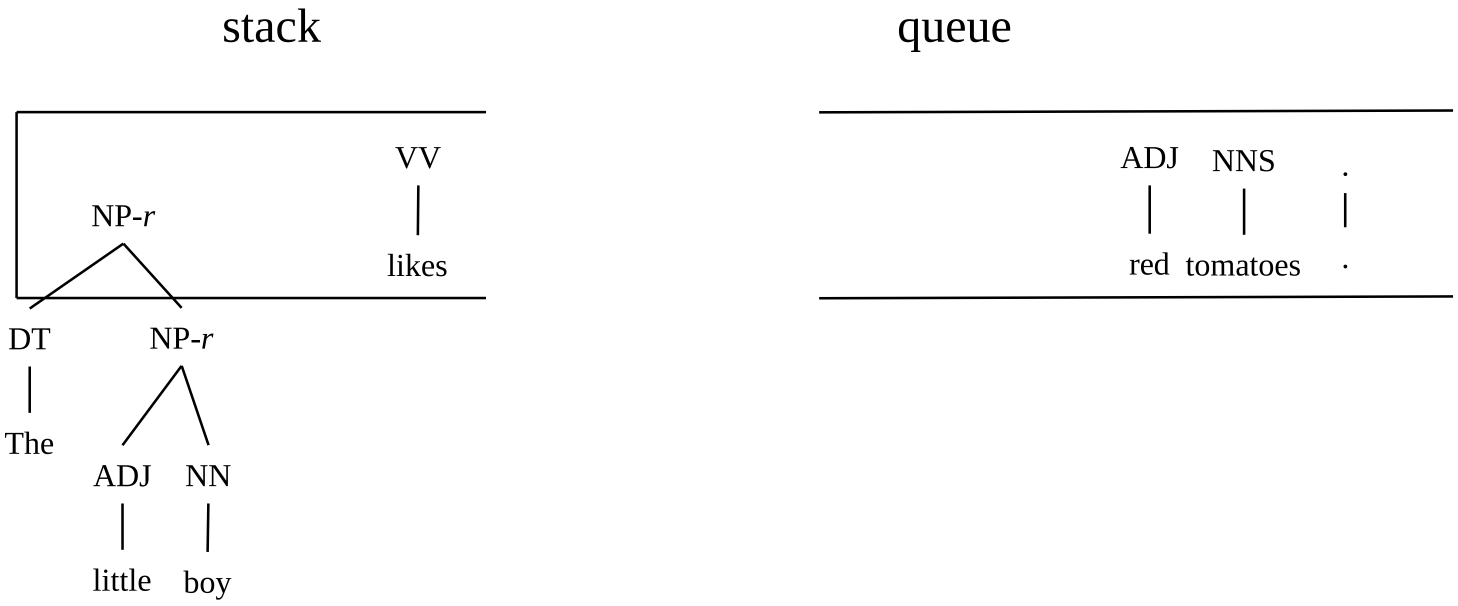
- Example
  - SHIFT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

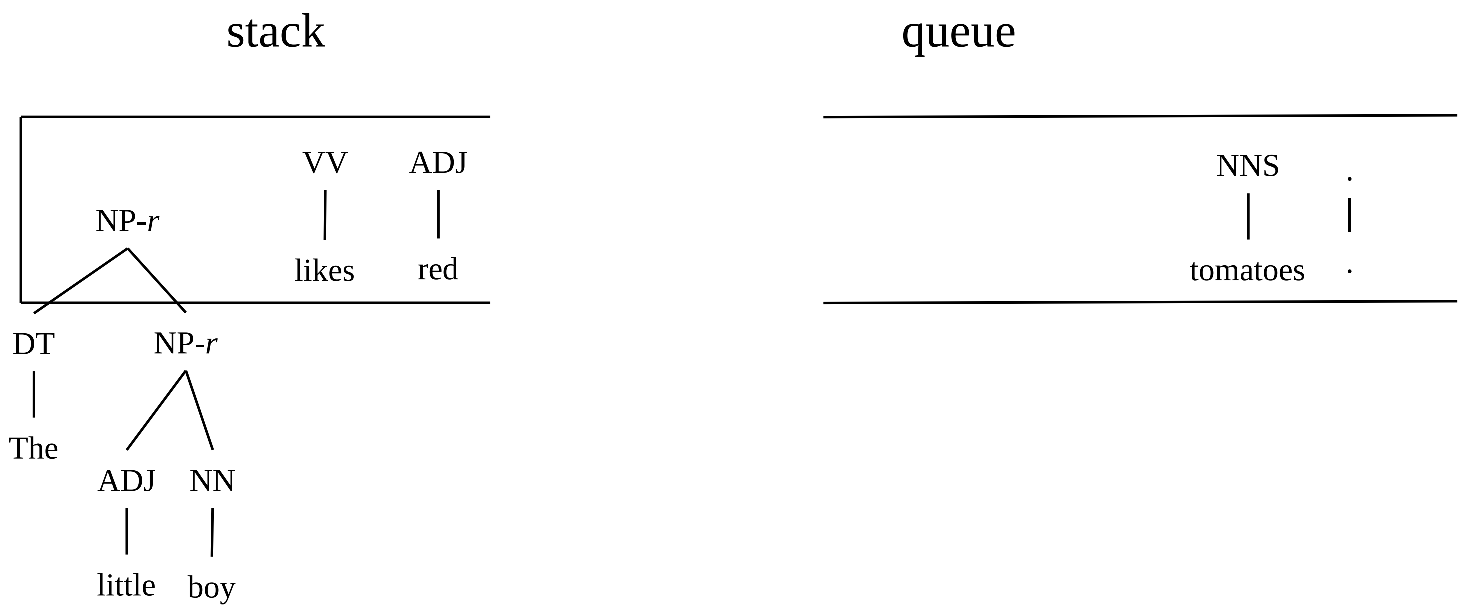
- Example
  - SHIFT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

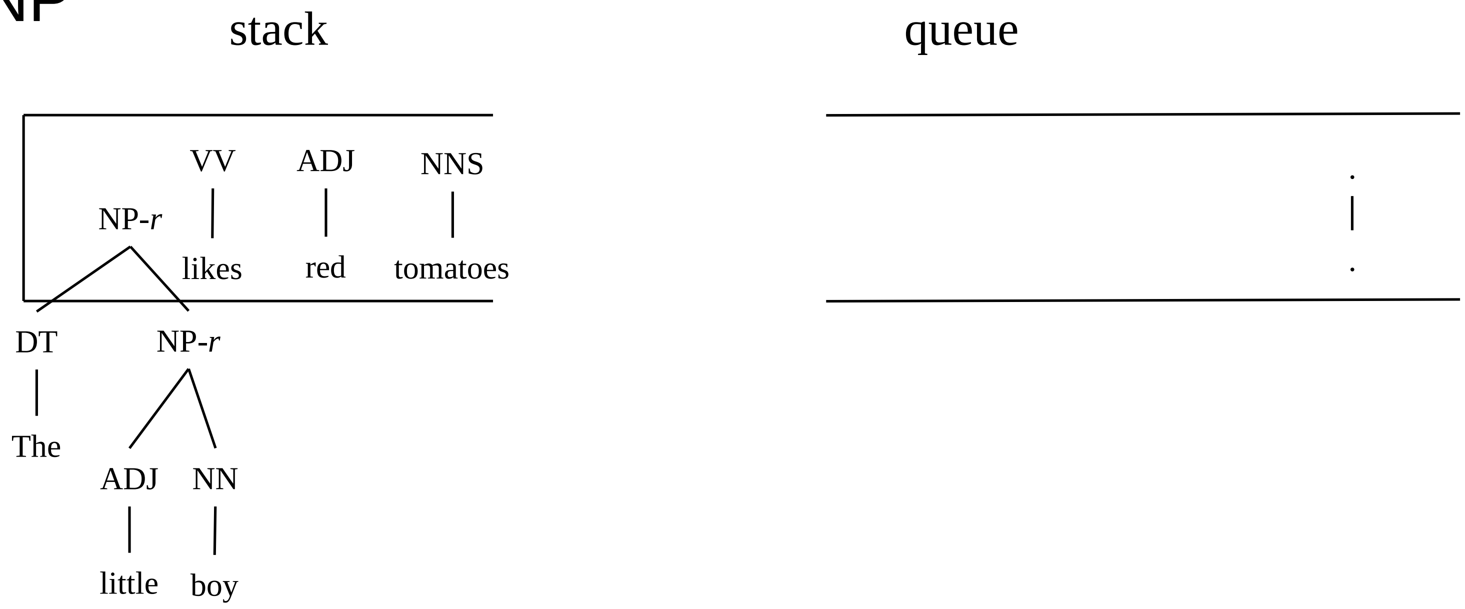
- Example
  - SHIFT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

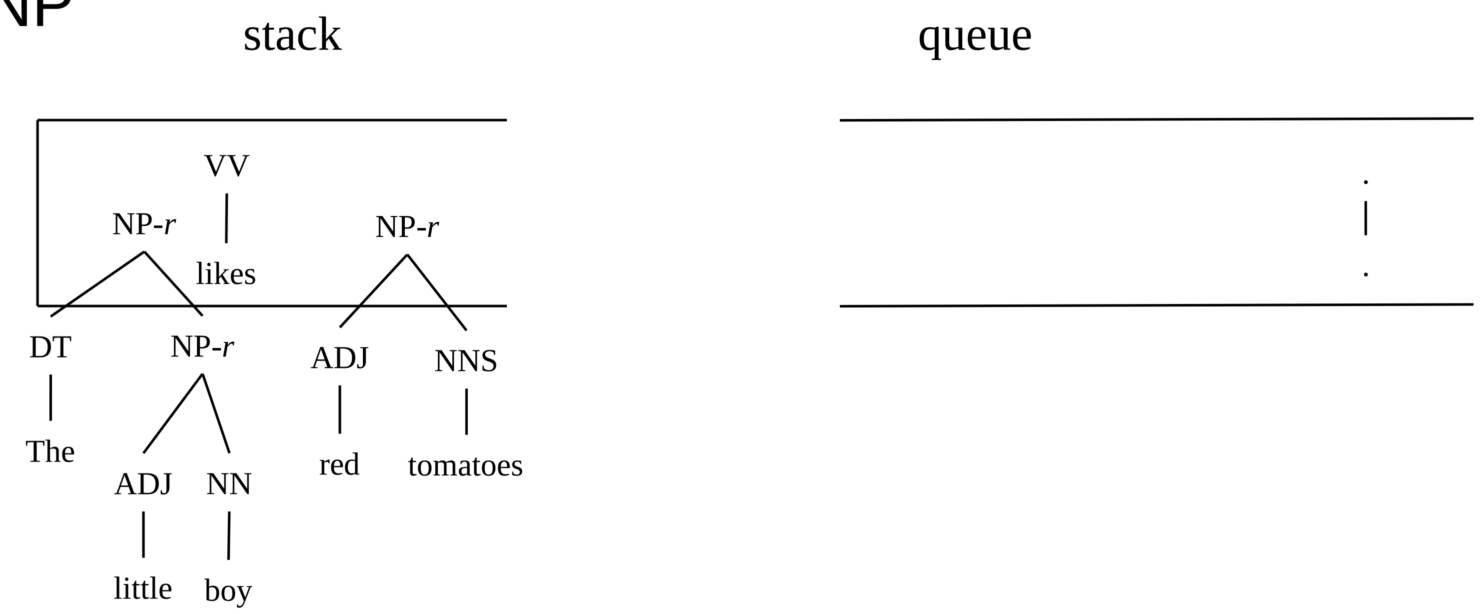
- Example
  - REDUCE-R-NP



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

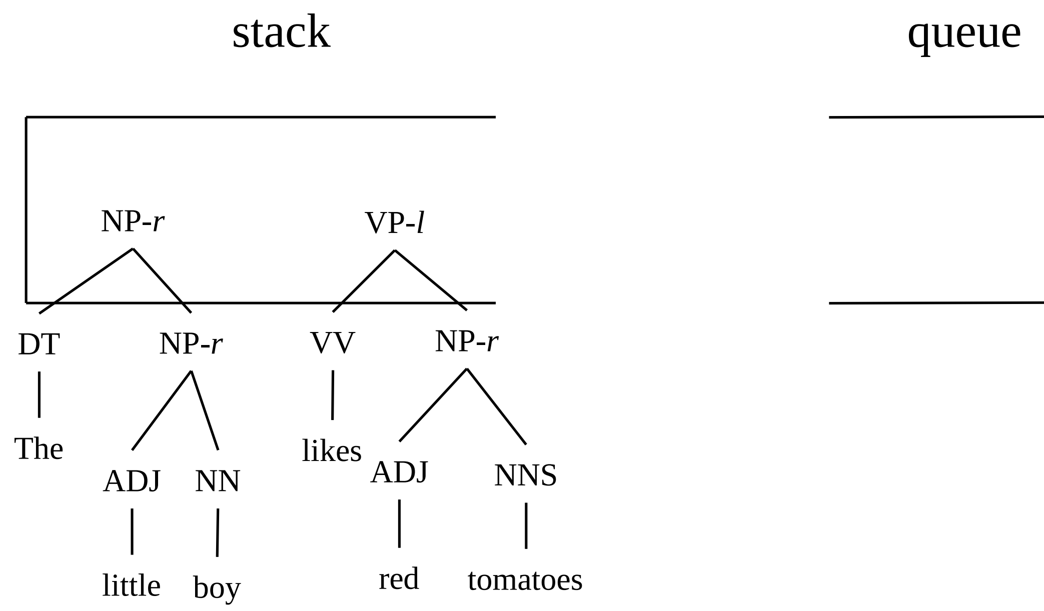
- Example
  - REDUCE-L-NP



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

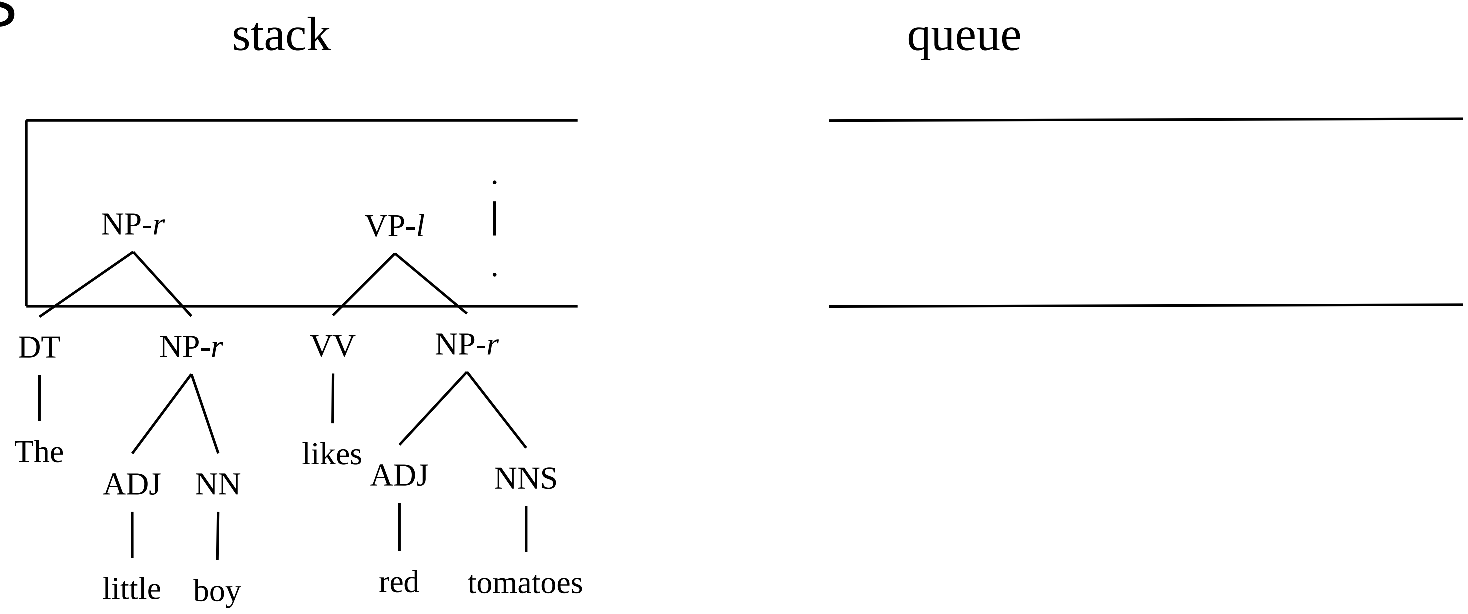
- Example
  - SHIFT



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

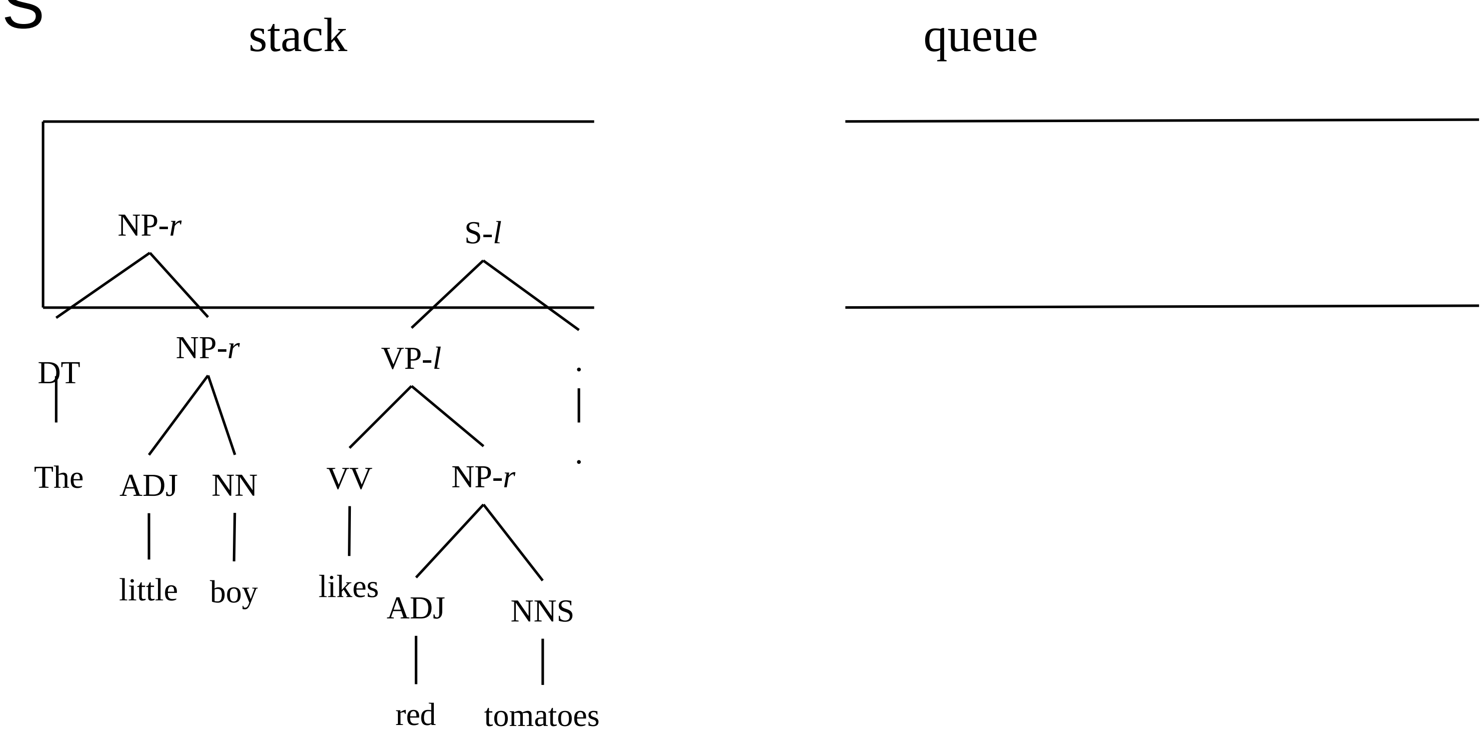
- Example
  - REDUCE-L-S



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

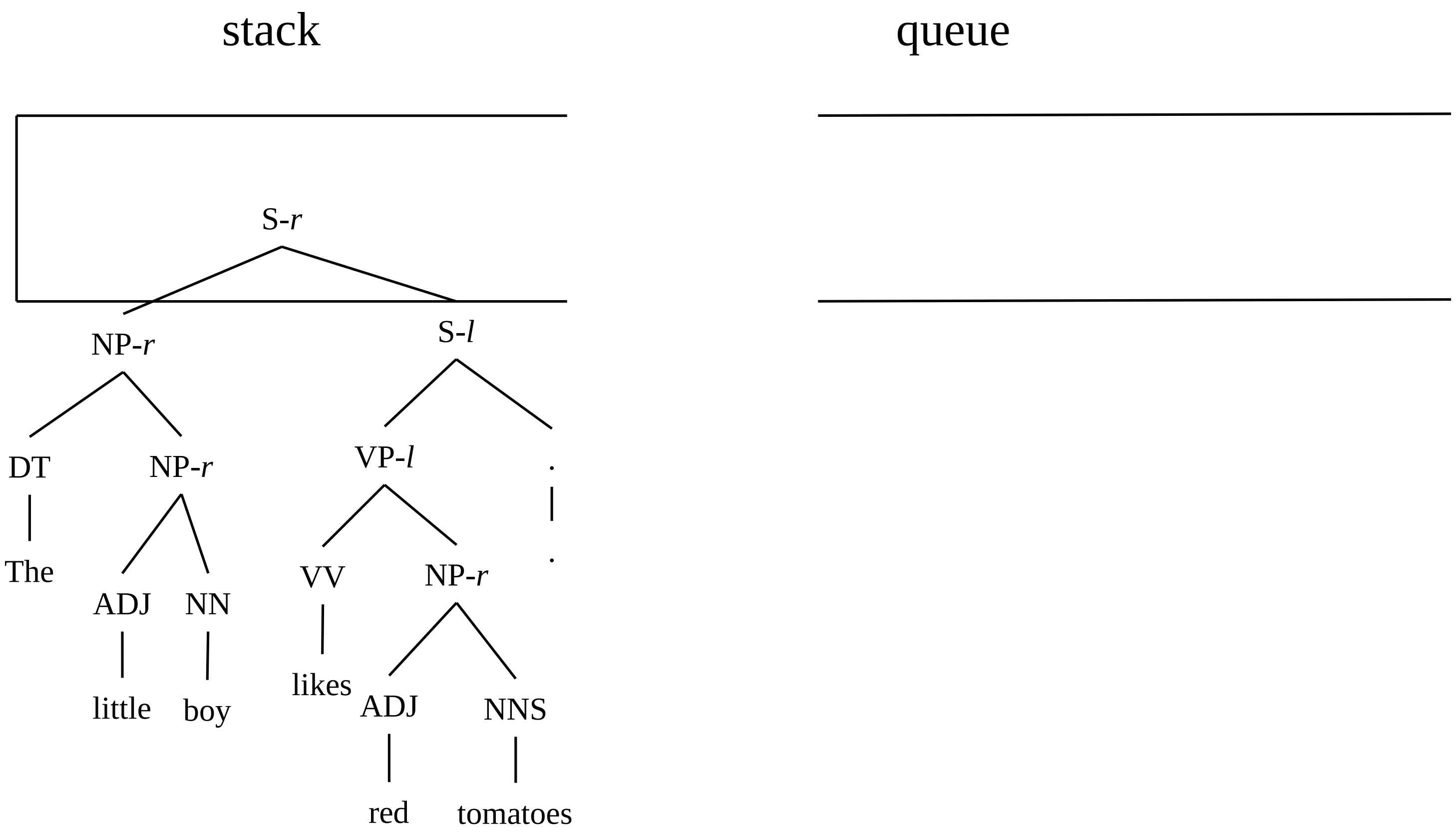
- Example
  - REDUCE-R-S



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Transition-based Constituent Parsing

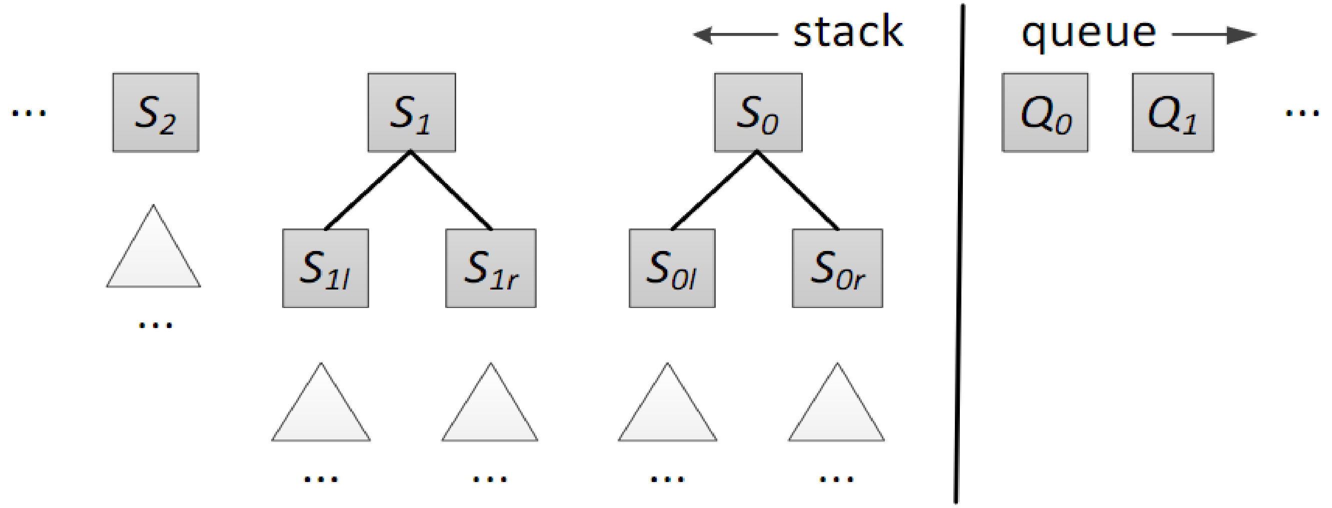
- Example
  - TERMINATE



Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Joint Segmentation, POS-tagging and Constituent Parsing

- The transition system
  - State:

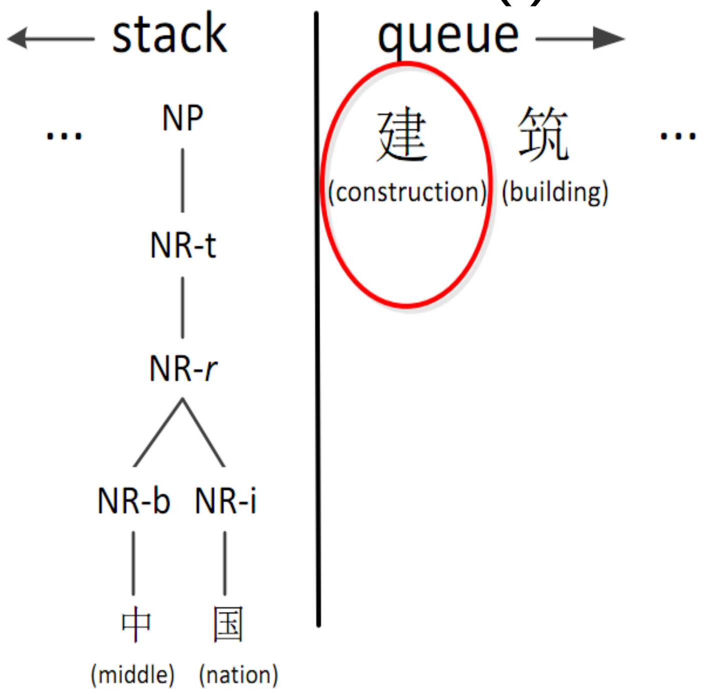


  - Actions:
    - SHIFT-SEPARATE(***t***), SHIFT-APPEND, REDUCE-SUBWORD(***d***), REDUCE-WORD, REDUCE-BINARY(***d***;***l***), REDUCE-UNARY(***l***), TERMINATE

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

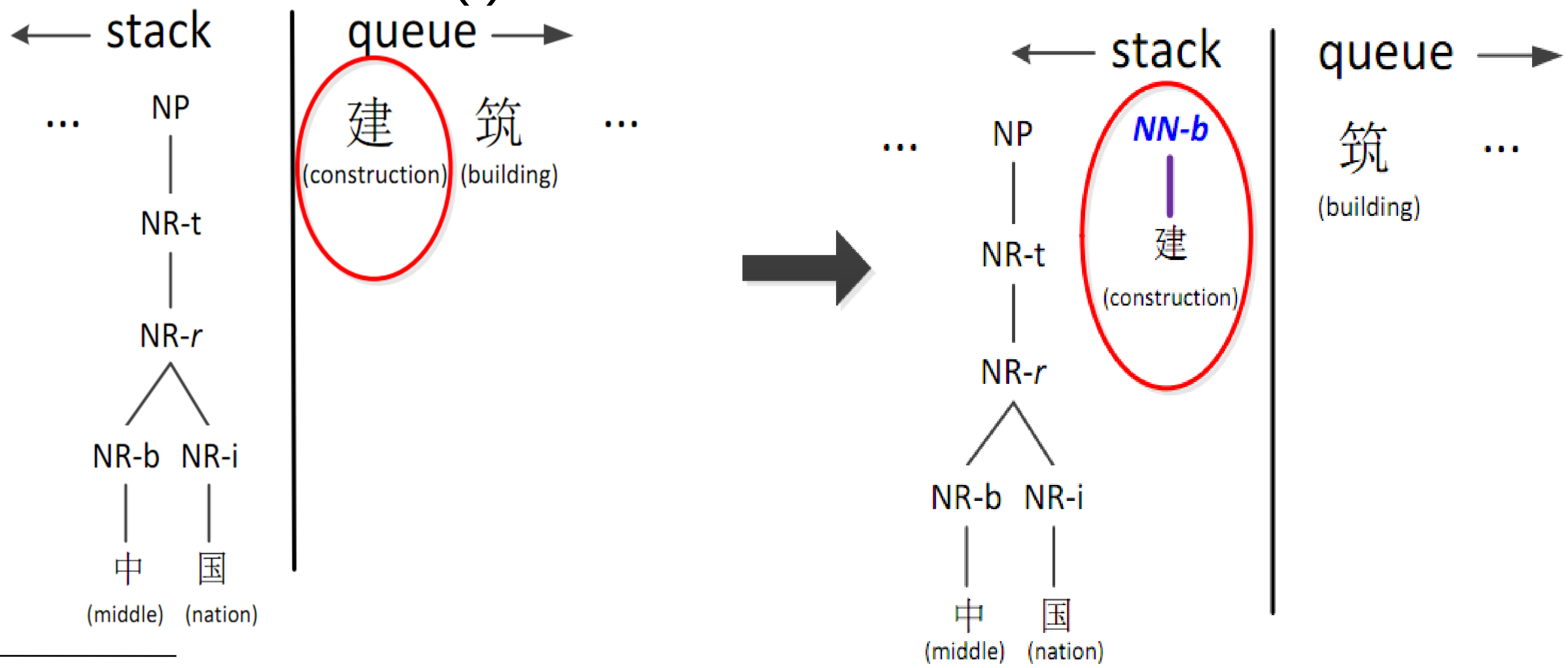
- Actions
  - SHIFT-SEPARATE(t)



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

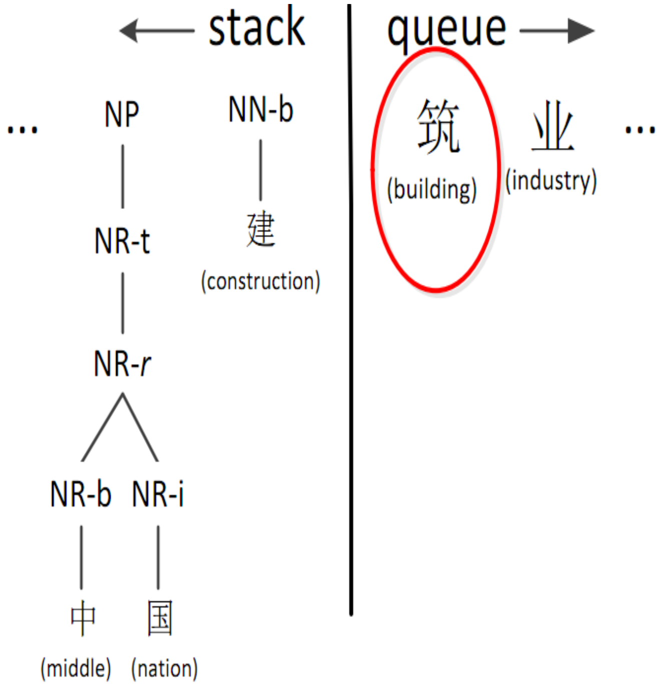
- Actions
  - SHIFT-SEPARATE(t)



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

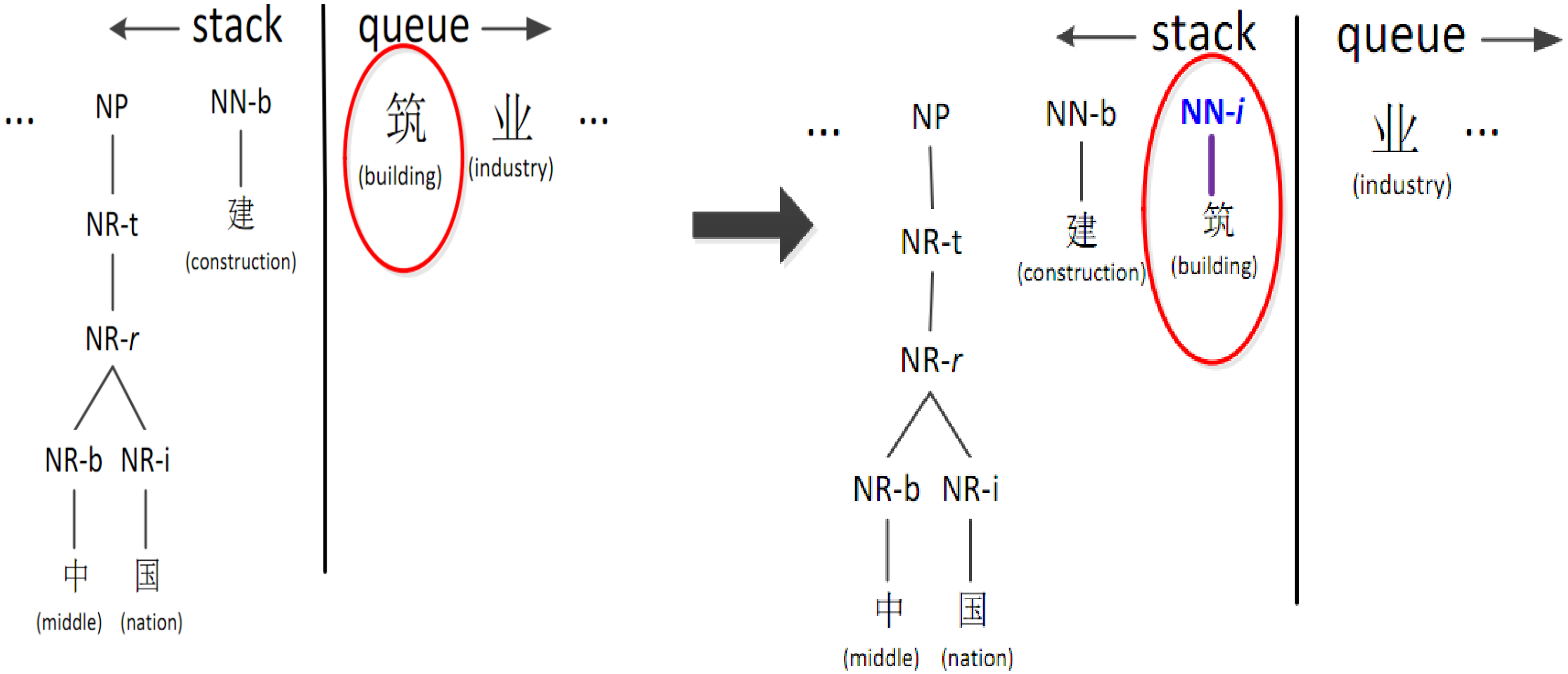
- Actions
  - SHIFT-APPEND



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

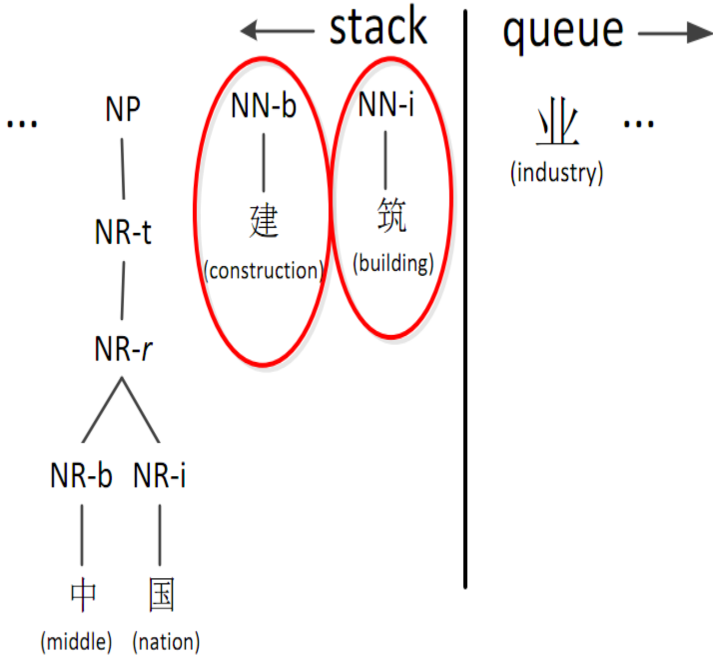
- Actions
  - SHIFT-APPEND



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

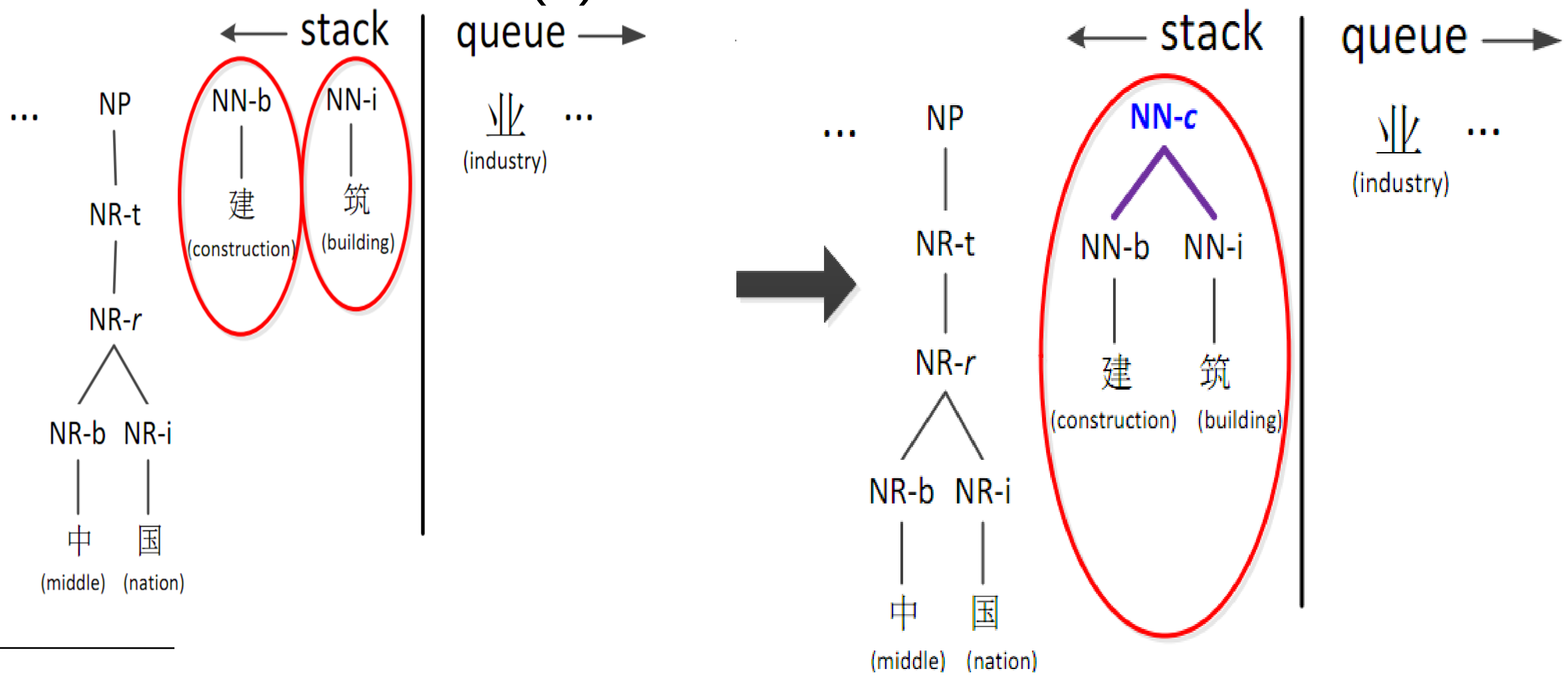
- Actions
  - REDUCE-SUBWORD(d)



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

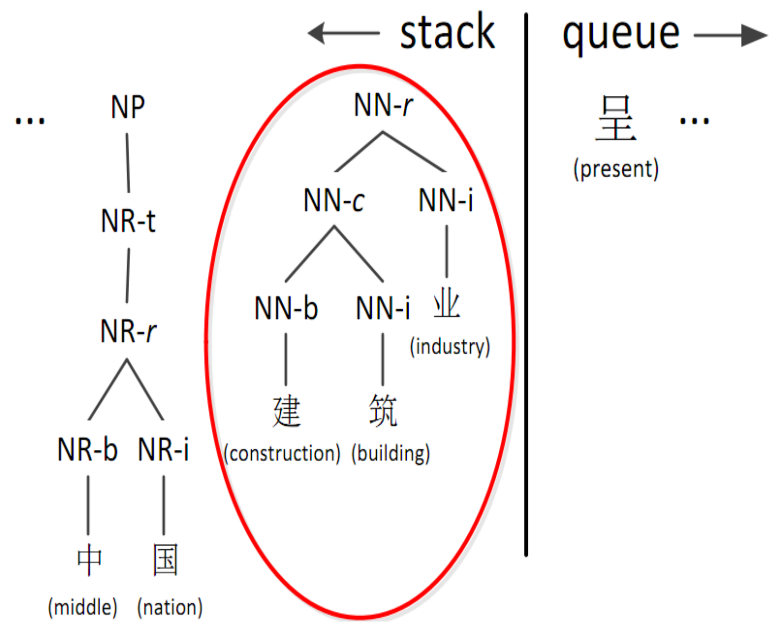
- Actions
  - REDUCE-SUBWORD(d)



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

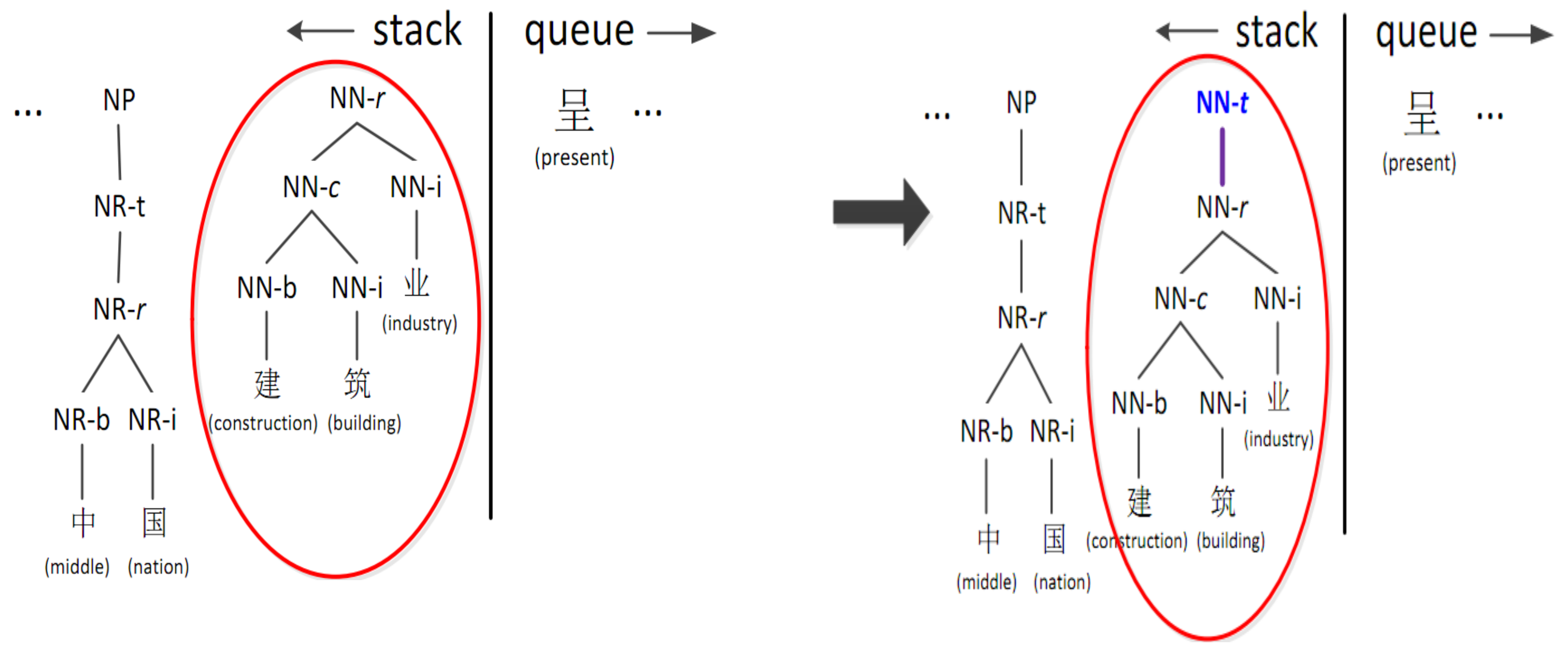
- Actions
  - REDUCE-WORD



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

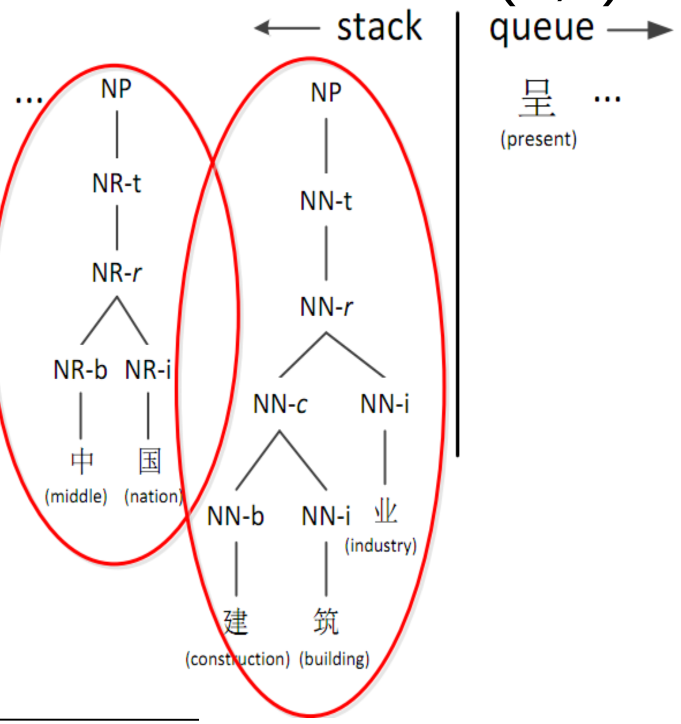
- Actions
  - REDUCE-WORD



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

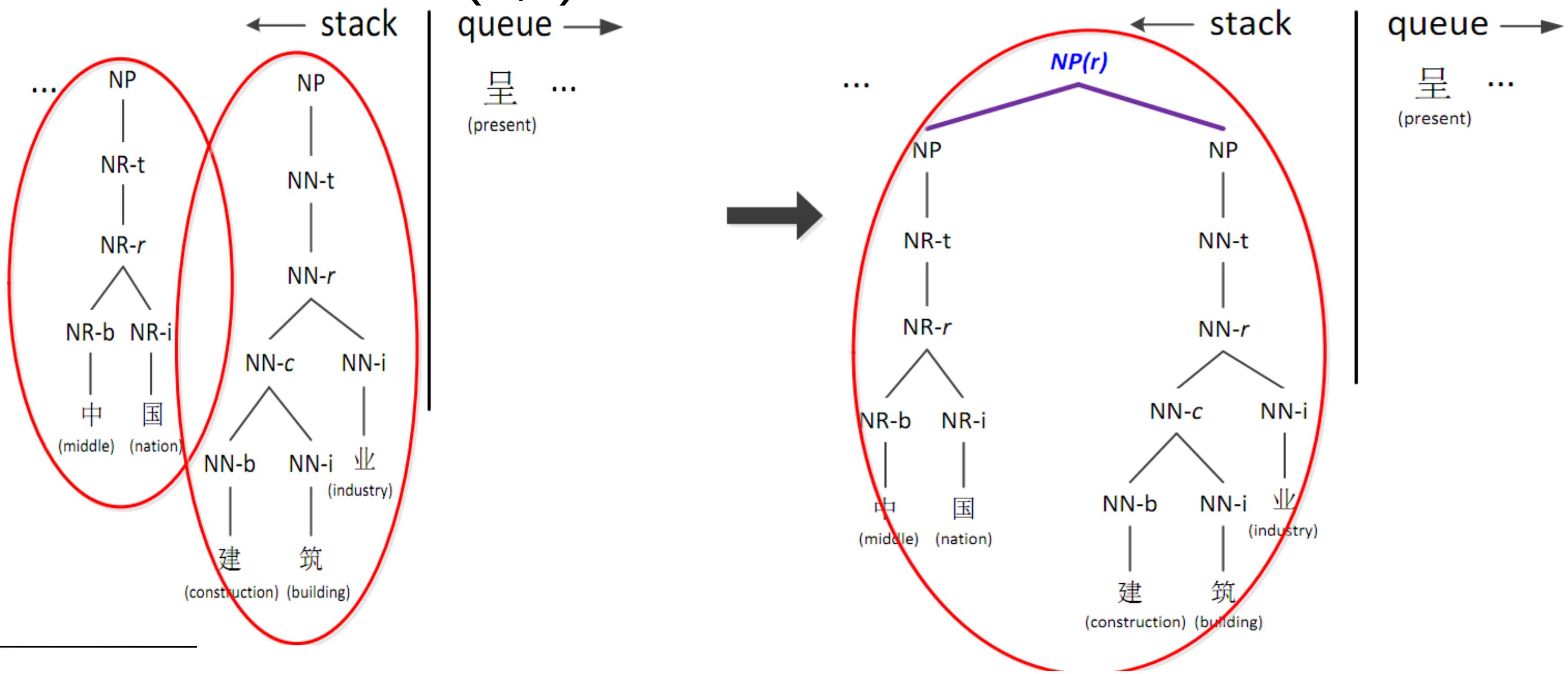
- Actions
  - REDUCE-BINARY(d; l)



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

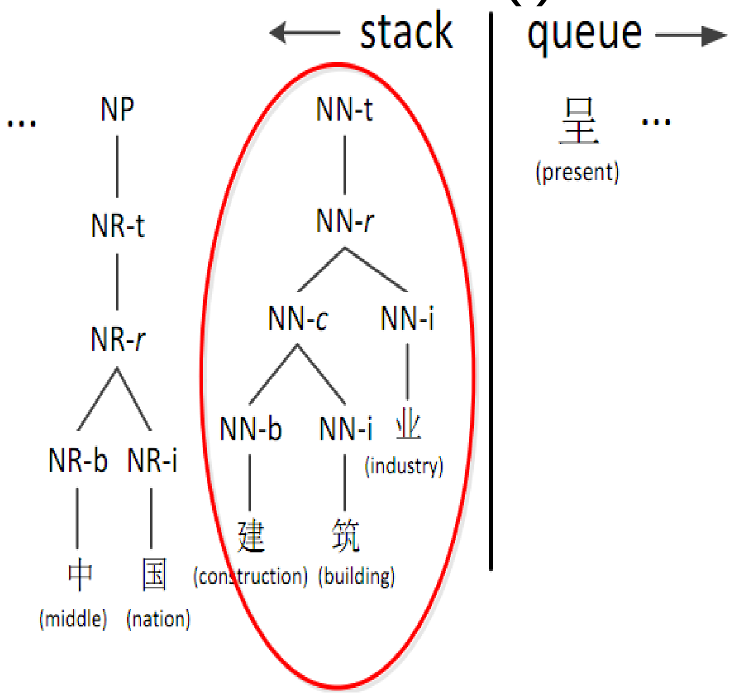
- Actions
  - REDUCE-BINARY(d; l)



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

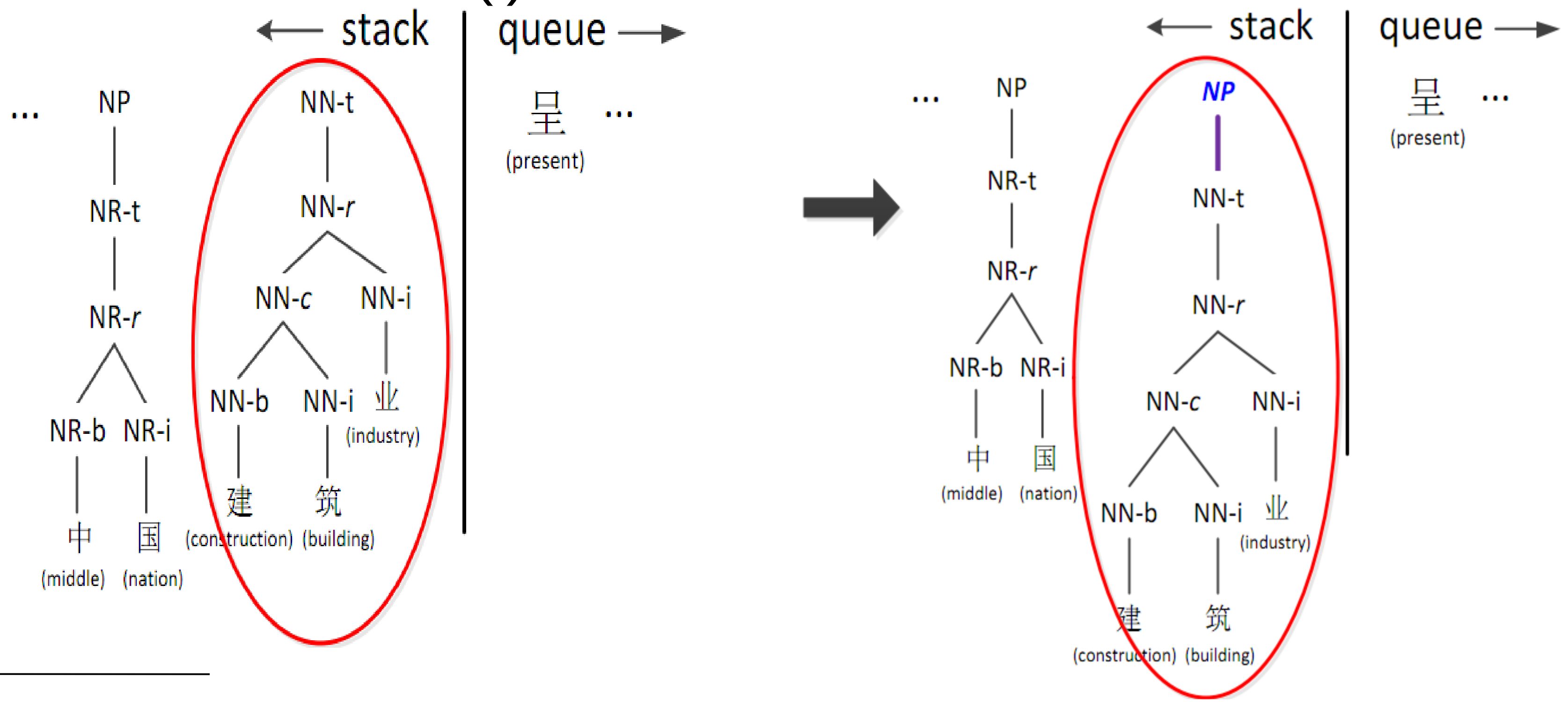
- Actions
  - REDUCE-UNARY(l)



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

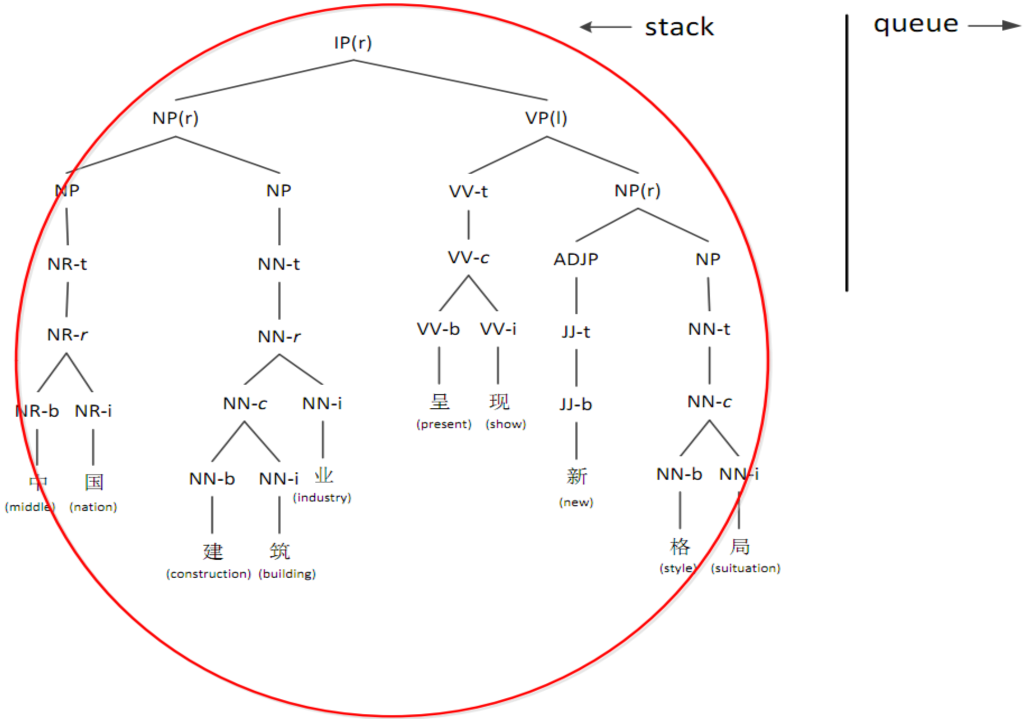
- Actions
  - REDUCE-UNARY(l)



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

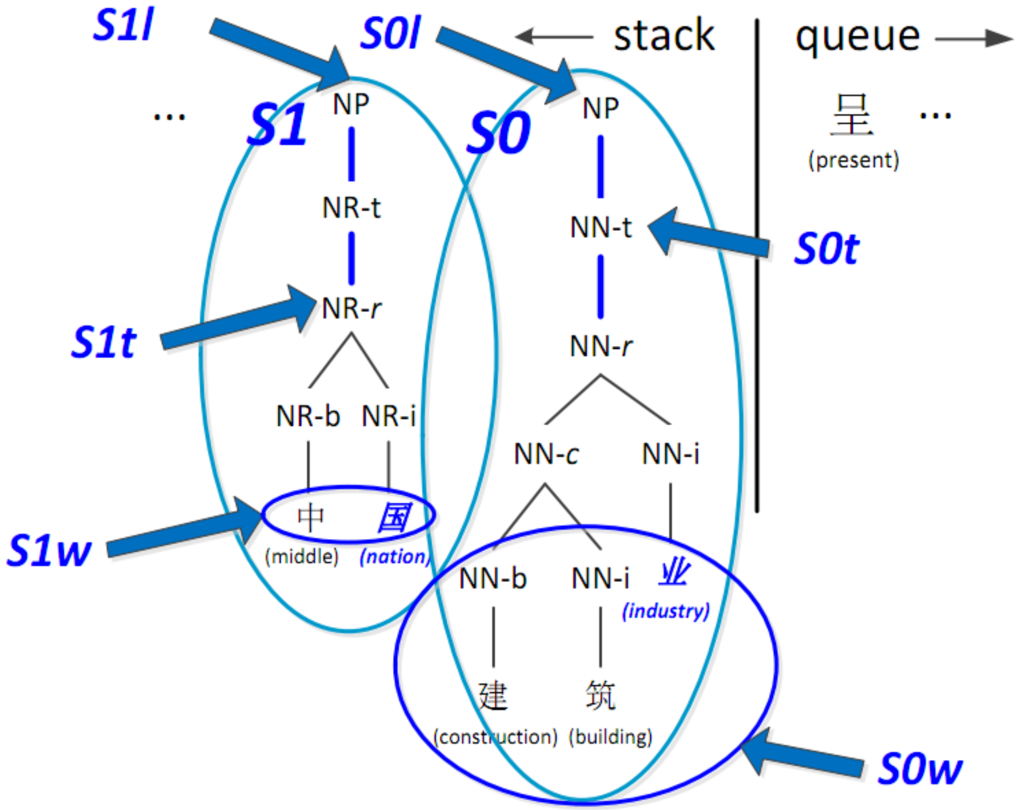
- Actions
  - TERMINATE



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

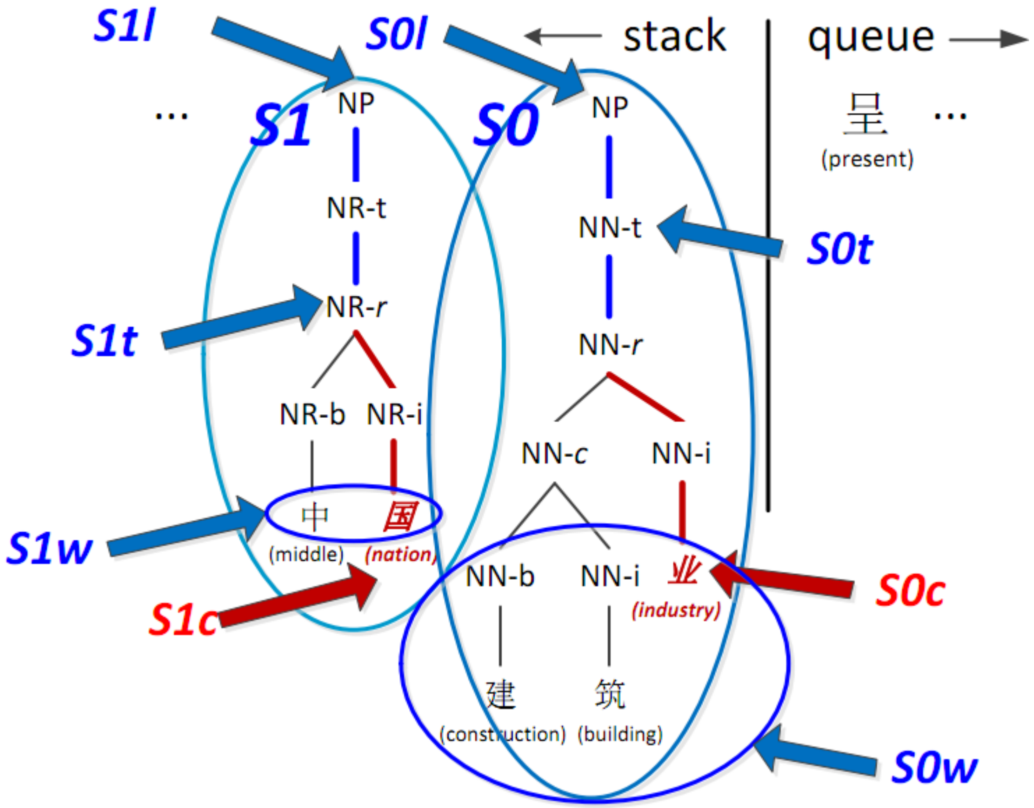
- Features



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

- Features



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

- Results on CTB

	Task	P	R	F
Pipeline	Seg	97.35	98.02	97.69
	Tag	93.51	94.15	93.83
	Parse	81.58	82.95	82.26
Flat word structures	Seg	97.32	98.13	97.73
	Tag	94.09	94.88	94.48
	Parse	83.39	83.84	83.61
Annotated word structures	Seg	97.49	98.18	97.84
	Tag	94.46	95.14	94.80
	Parse	84.42	84.43	84.43
	WS	94.02	94.69	94.35

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint Segmentation, POS-tagging and Constituent Parsing

- Results on CTB

Task	Seg	Tag	Parse
Kruengkrai+ ’09	97.87	93.67	–
Sun ’11	98.17	94.02	–
Wang+ ’11	98.11	94.18	–
Li ’11	97.3	93.5	79.7
Li+ ’12	97.50	93.31	–
Hatori+ ’12	98.26	94.64	–
Qian+ ’12	97.96	93.81	82.85
Ours pipeline	97.69	93.83	82.26
Ours joint flat	97.73	94.48	83.61
Ours joint annotated	97.84	94.80	84.43

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Chinese Parsing Exploiting Characters.* In proceedings of ACL 2013. Sophia, Bulgaria. August.

Joint POS tagging and Dependency Parsing

- Actions
  - INITIALIZATION

Stack [S]

Buffer [B]

He$_1$ won$_2$ the$_3$ game$_4$

Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

- Actions
  - SHIFT

Stack [S]

He$_1$

Buffer [B]

won$_2$ the$_3$ game$_4$

Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

- Actions
  - TAG_{PRP}

Stack [S]

He$_{1/\text{PRP}}$

Buffer [B]

won$_2$ the$_3$ game$_4$

Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

- Actions
  - SHIFT

Stack [S]

$\text{He}_{1/\text{PRP}}$ won_2

Buffer [B]

the_3 game_4

Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

- Actions
  - TAG_{VBD}

Stack [S]

$\text{He}_{1/\text{PRP}}$ $\text{won}_{2/\text{VBD}}$

Buffer [B]

the_3 game_4

Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

- Actions
  - LEFT

Stack [S]

He$_{1/PRP}$ won$_{2/VBD}$

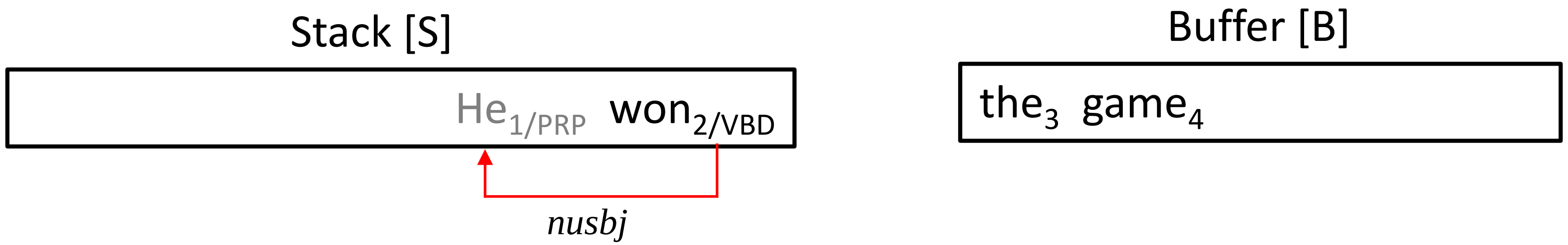
Buffer [B]

the$_3$ game$_4$

Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

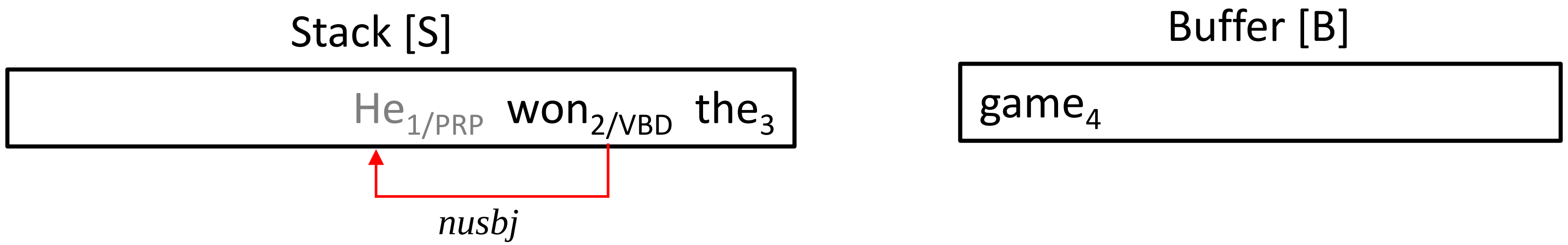
- Actions
  - $\text{LABEL}_{\text{nsubj}}$



Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

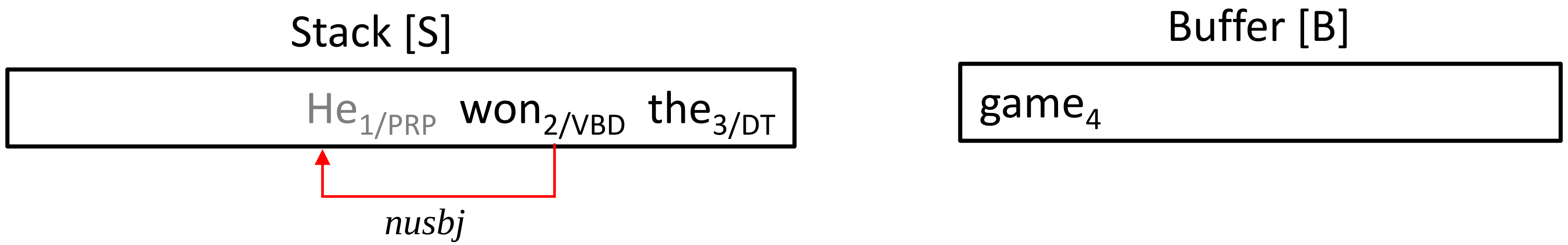
- Actions
  - SHIFT



Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

- Actions
  - TAG_{DT}



Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

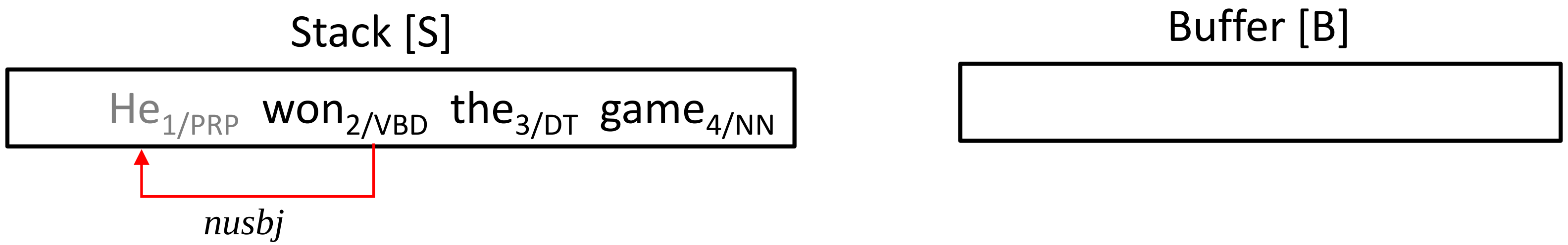
Joint POS tagging and Dependency Parsing

- Actions
  - SHIFT

Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

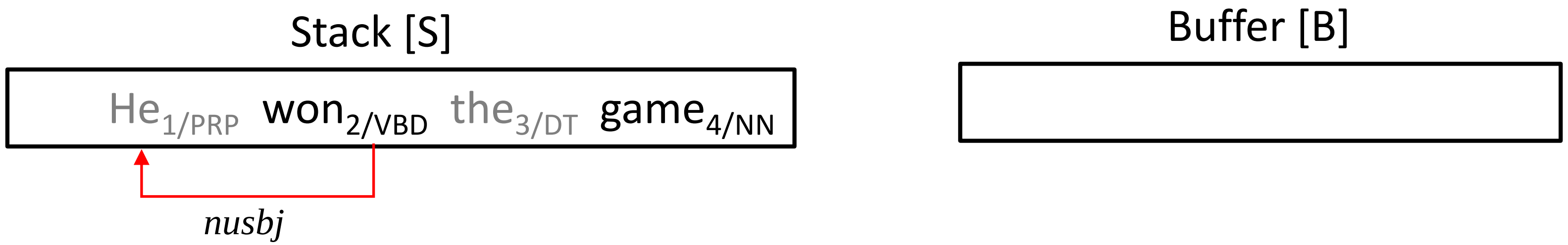
- Actions
  - TAG_{NN}



Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

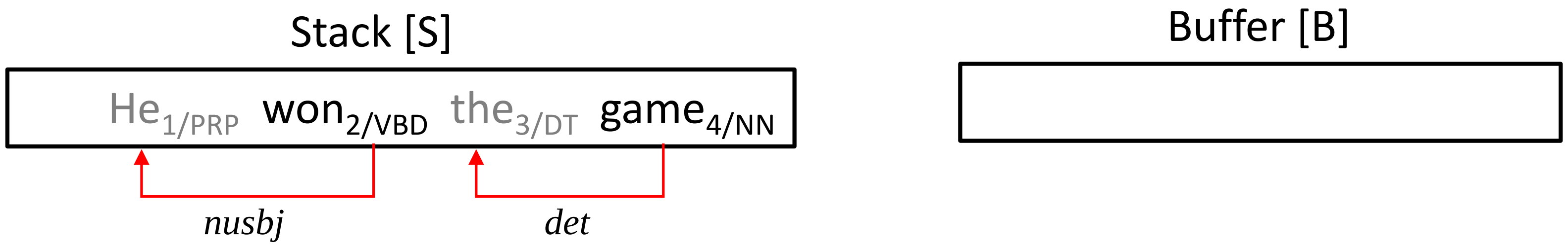
- Actions
  - LEFT



Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

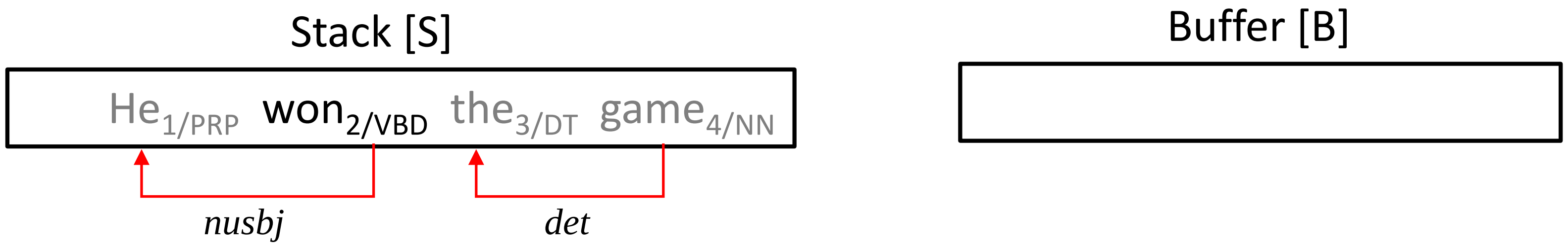
- Actions
  - $\text{LABEL}_{\text{det}}$



Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

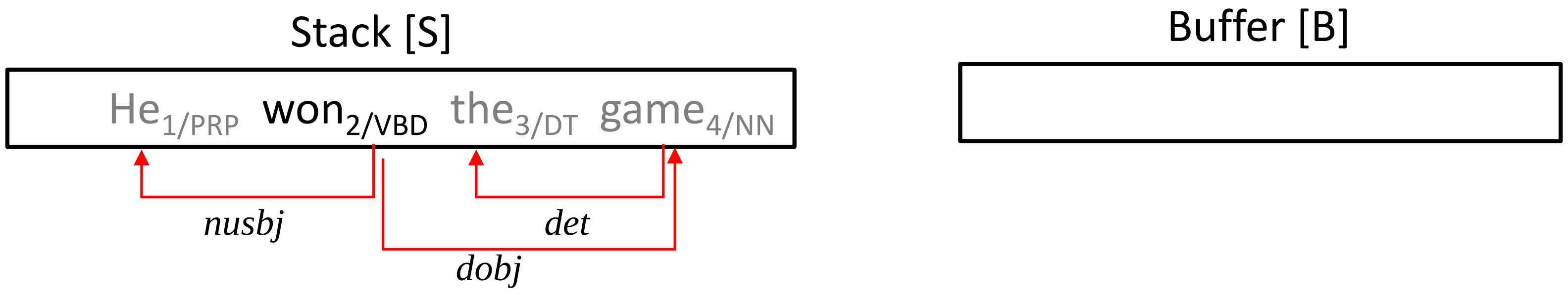
- Actions
  - RIGHT



Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint POS tagging and Dependency Parsing

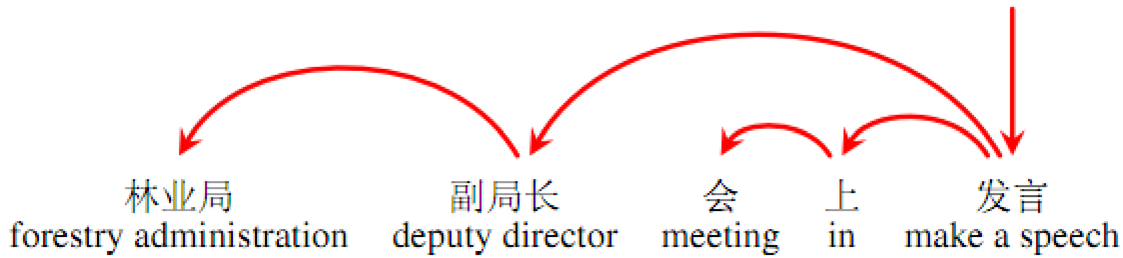
- Actions
  - $\text{LABEL}_{\text{dobj}}$



Bohnet, Bernd, and Joakim Nivre. "A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing." Proceedings of the 2012 Joint Conference on EMNLP and CoNLL. ACL, 2012.

Joint Segmentation, POS-tagging and Dependency Parsing

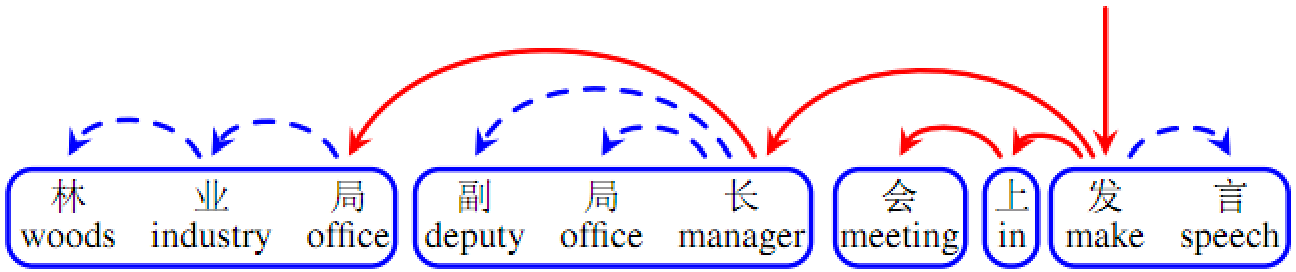
- Traditional word-based dependency parsing
  - Inter-word dependencies



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, Jun'ichi Tsujii. Incremental Joint Approach to Chinese Word Segmentation, POS Tagging, and Dependency Parsing. In the Proceedings of ACL. Jeju, Korea. 2012.

Joint Segmentation, POS-tagging and Dependency Parsing

- Character-level dependency parsing
  - Inter- and intra-word dependencies



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, Jun'ichi Tsujii. Incremental Joint Approach to Chinese Word Segmentation, POS Tagging, and Dependency Parsing. In the Proceedings of ACL. Jeju, Korea. 2012.

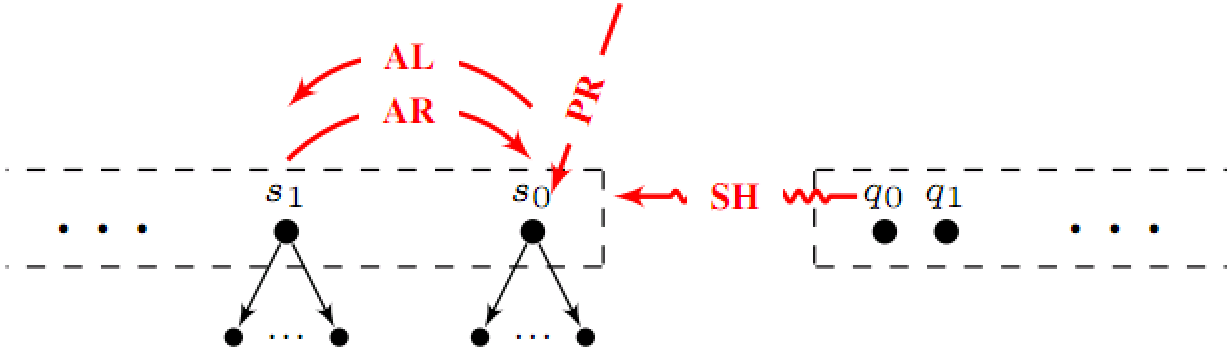
Joint Segmentation, POS-tagging and Dependency Parsing

- Extensions from word-level transition-based dependency parsing models
  - Arc-standard (Nirve 2008; Huang et al., 2009)
  - Arc-eager (Nirve 2008; Zhang and Clark, 2008)

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, Jun’ichi Tsujii. Incremental Joint Approach to Chinese Word Segmentation, POS Tagging, and Dependency Parsing. In the Proceedings of ACL. Jeju, Korea. 2012.

Joint Segmentation, POS-tagging and Dependency Parsing

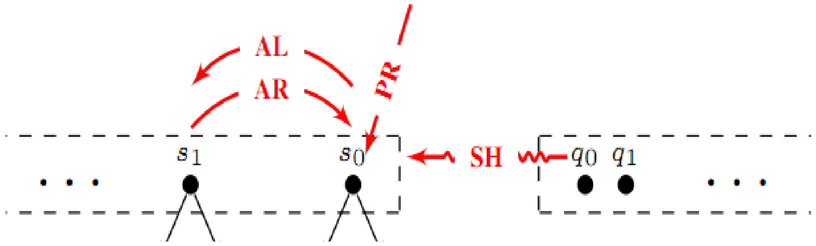
- Word-level transition-based dependency parsing
  - Arc-standard

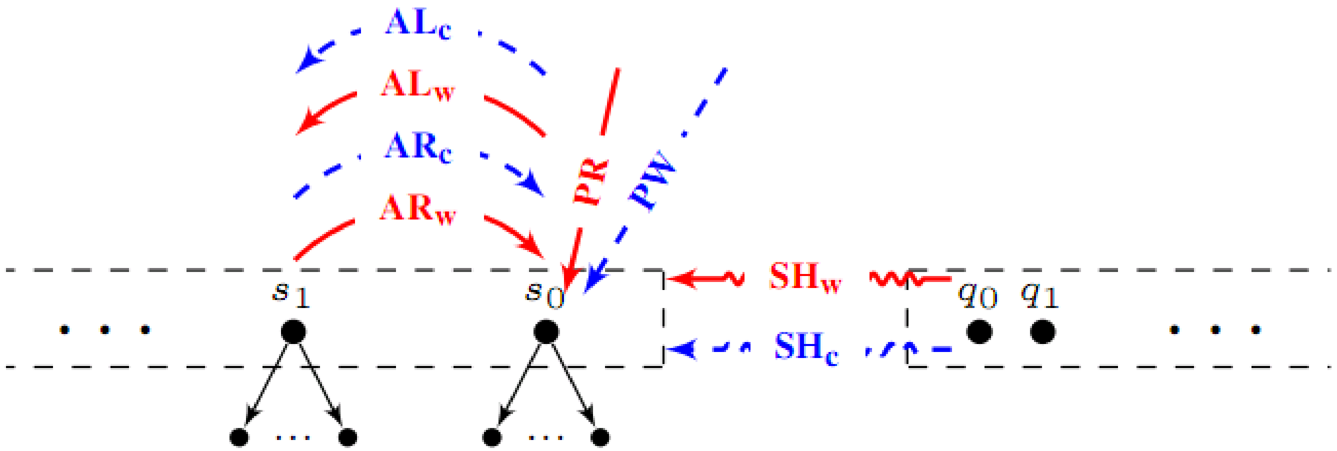


Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, Jun'ichi Tsujii. Incremental Joint Approach to Chinese Word Segmentation, POS Tagging, and Dependency Parsing. In the Proceedings of ACL. Jeju, Korea. 2012.

Joint Segmentation, POS-tagging and Dependency Parsing

- Word-level to character-level
  - Arc-standard

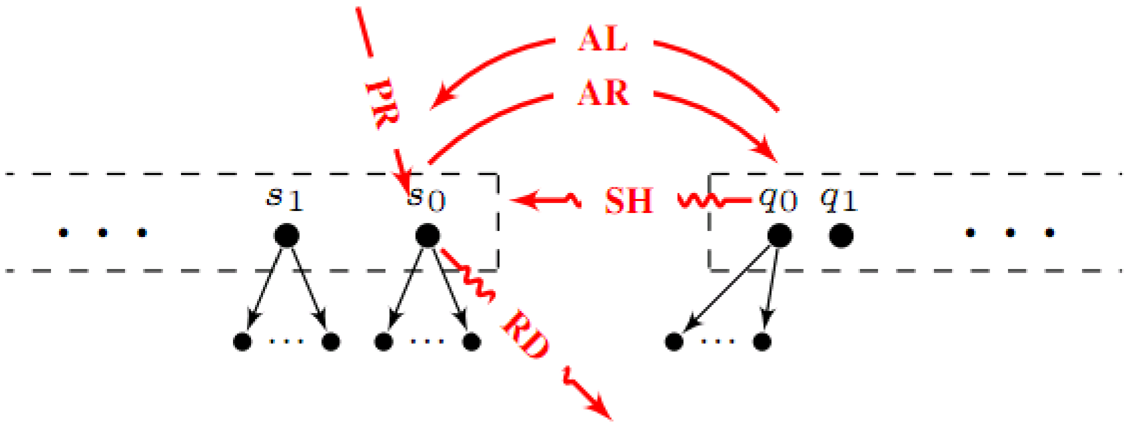




Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, Jun'ichi Tsujii. Incremental Joint Approach to Chinese Word Segmentation, POS Tagging, and Dependency Parsing. In the Proceedings of ACL. Jeju, Korea. 2012.

Joint Segmentation, POS-tagging and Dependency Parsing

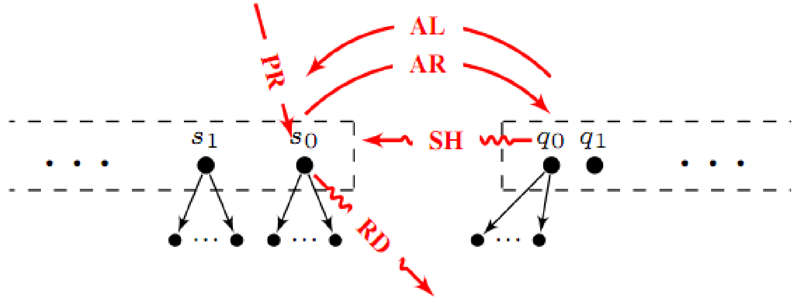
- Word-level transition-based dependency parsing
  - Arc-eager

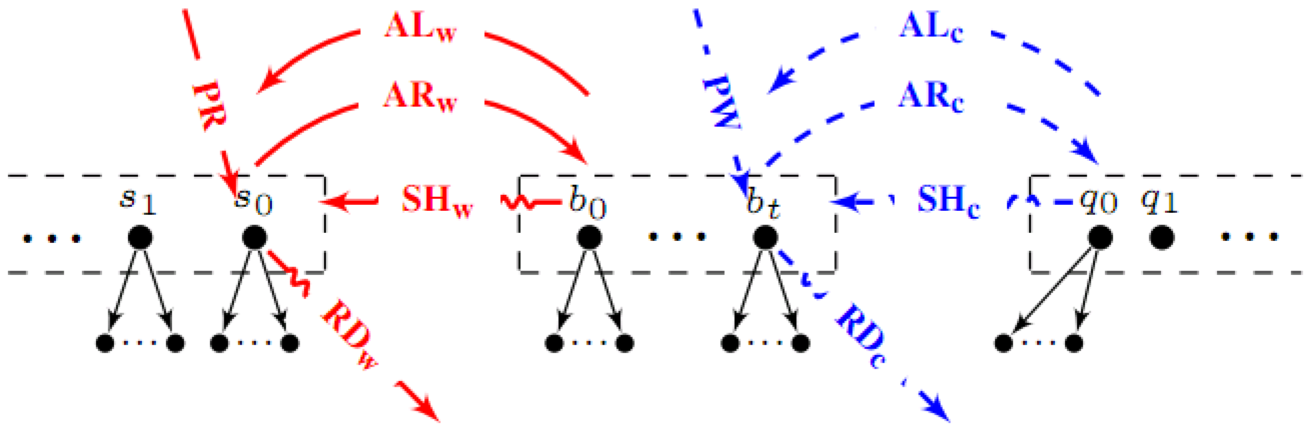


Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, Jun'ichi Tsujii. Incremental Joint Approach to Chinese Word Segmentation, POS Tagging, and Dependency Parsing. In the Proceedings of ACL. Jeju, Korea. 2012.

Joint Segmentation, POS-tagging and Dependency Parsing

- Word-level to character-level
  - Arc-eager



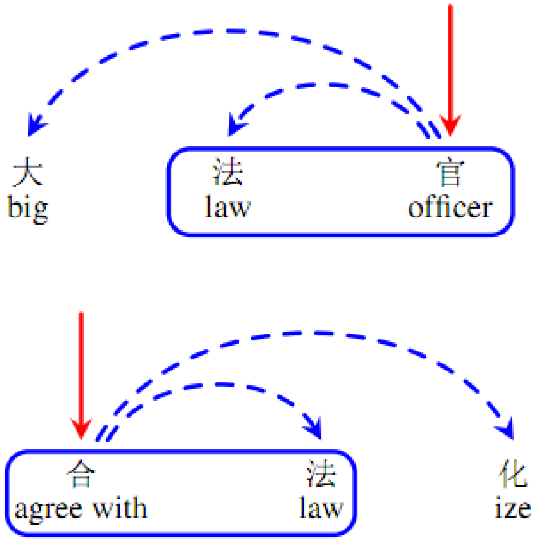


Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, Jun'ichi Tsujii. Incremental Joint Approach to Chinese Word Segmentation, POS Tagging, and Dependency Parsing. In the Proceedings of ACL. Jeju, Korea. 2012.

Joint Segmentation, POS-tagging and Dependency Parsing

- New features

Feature templates
$L\underline{c}$, $L\underline{c}\underline{t}$, $R\underline{c}$, $R\underline{c}\underline{t}$, $L_{lc1}\underline{c}$, $L_{rc1}\underline{c}$, $R_{lc1}\underline{c}$,
$L\underline{c} \cdot R\underline{c}$, $L_{lc1}\underline{c}\underline{t}$, $L_{rc1}\underline{c}\underline{t}$, $R_{lc1}\underline{c}\underline{t}$,
$L\underline{c} \cdot R\underline{w}$, $L\underline{w} \cdot R\underline{c}$, $L\underline{c}\underline{t} \cdot R\underline{w}$,
$L\underline{w}\underline{t} \cdot R\underline{c}$, $L\underline{w} \cdot R\underline{c}\underline{t}$, $L\underline{c} \cdot R\underline{w}\underline{t}$,
$L\underline{c} \cdot R\underline{c} \cdot L_{lc1}\underline{c}$, $L\underline{c} \cdot R\underline{c} \cdot L_{rc1}\underline{c}$,
$L\underline{c} \cdot R\underline{c} \cdot L_{lc2}\underline{c}$, $L\underline{c} \cdot R\underline{c} \cdot L_{rc2}\underline{c}$,
$L\underline{c} \cdot R\underline{c} \cdot R_{lc1}\underline{c}$, $L\underline{c} \cdot R\underline{c} \cdot R_{lc2}\underline{c}$,
$L\underline{lsw}$, $L\underline{rsw}$, $R\underline{lsw}$, $R\underline{rsw}$, $L\underline{lsw}\underline{t}$,
$L\underline{rsw}\underline{t}$, $R\underline{lsw}\underline{t}$, $R\underline{rsw}\underline{t}$, $L\underline{lsw} \cdot R\underline{w}$,
$L\underline{rsw} \cdot R\underline{w}$, $L\underline{w} \cdot R\underline{lsw}$, $L\underline{w} \cdot R\underline{rsw}$



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, Jun'ichi Tsujii. Incremental Joint Approach to Chinese Word Segmentation, POS Tagging, and Dependency Parsing. In the Proceedings of ACL. Jeju, Korea. 2012.

Joint Segmentation, POS-tagging and Dependency Parsing

- Results on CTB

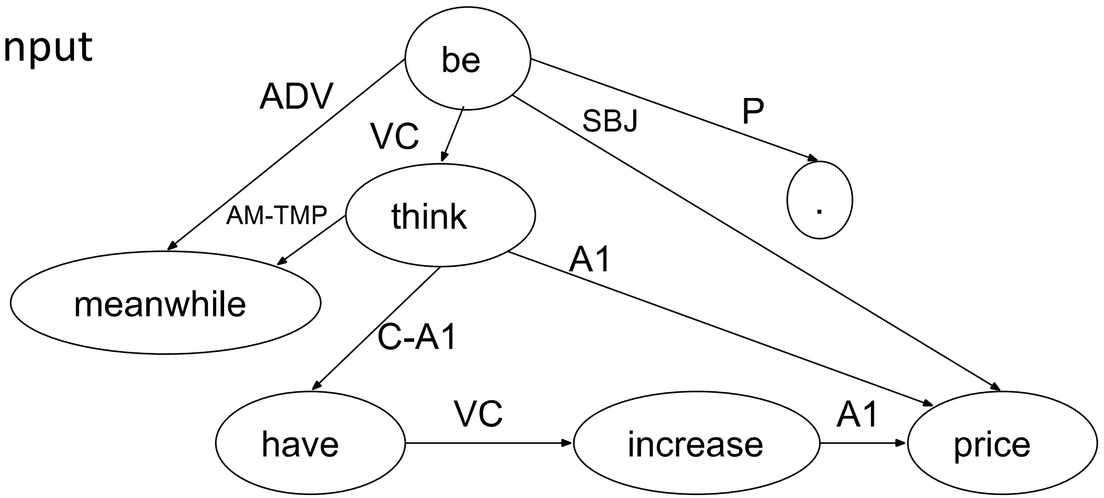
Model	CTB50				CTB60				CTB70			
	SEG	POS	DEP	WS	SEG	POS	DEP	WS	SEG	POS	DEP	WS
The arc-standard models												
STD (pipe)	97.53	93.28	79.72	–	95.32	90.65	75.35	–	95.23	89.92	73.93	–
STD (real, pseudo)	97.78	93.74	–	**97.40**	**95.77**‡	91.24‡	–	**95.08**	**95.59**‡	90.49‡	–	**94.97**
STD (pseudo, real)	97.67	94.28‡	81.63‡	–	95.63‡	**91.40**‡	76.75‡	–	95.53‡	90.75‡	75.63‡	–
STD (real, real)	**97.84**	**94.62**‡	**82.14**‡	97.30	95.56‡	91.39‡	**77.09**‡	94.80	95.51‡	**90.76**‡	**75.70**‡	94.78
Hatori+ '12	97.75	94.33	81.56	–	95.26	91.06	75.93	–	95.27	90.53	74.73	–
The arc-eager models												
EAG (pipe)	97.53	93.28	79.59	–	95.32	90.65	74.98	–	95.23	89.92	73.46	–
EAG (real, pseudo)	97.75	93.88	–	97.45	95.63‡	91.07‡	–	95.06	**95.50**‡	90.36‡	–	**95.00**
EAG (pseudo, real)	97.76	**94.36**‡	81.70‡	–	95.63‡	91.34‡	76.87‡	–	95.39‡	90.56‡	75.56‡	–
EAG (real, real)	**97.84**	**94.36**‡	**82.07**‡	**97.49**	**95.71**‡	**91.51**‡	**76.99**‡	**95.16**	95.47‡	**90.72**‡	**75.76**‡	94.94

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, Jun'ichi Tsujii. Incremental Joint Approach to Chinese Word Segmentation, POS Tagging, and Dependency Parsing. In the Proceedings of ACL. Jeju, Korea. 2012.

Joint Morphology and Linearization

- Task

Input

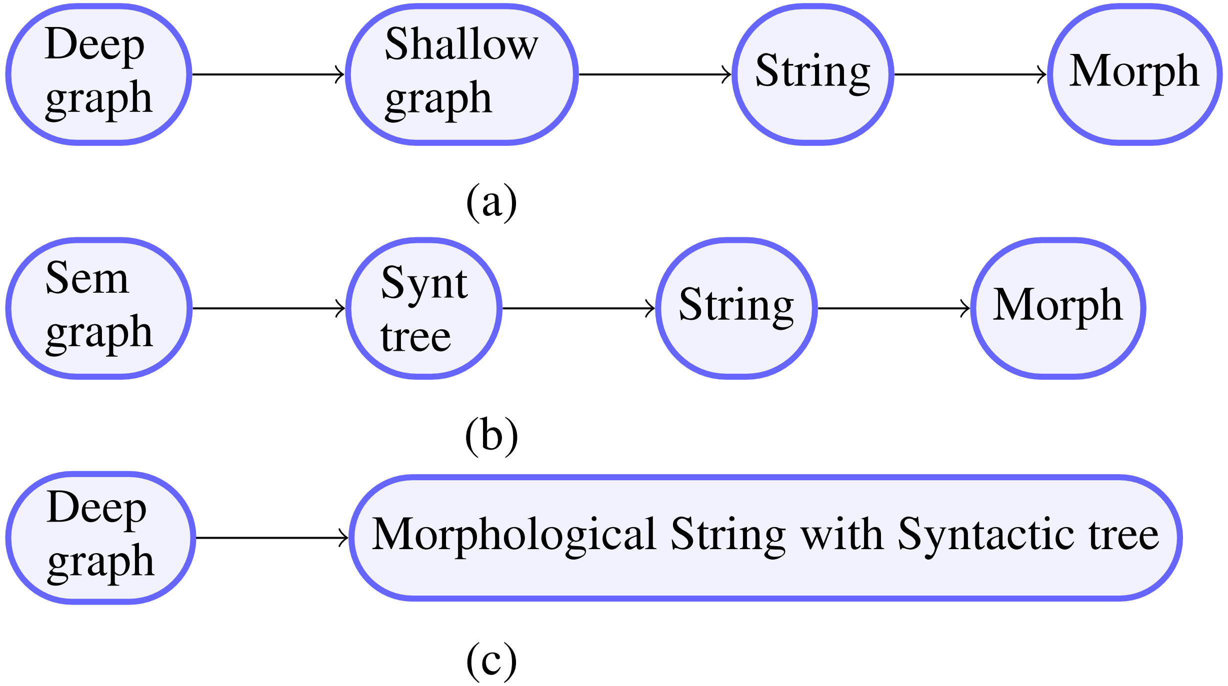


Output: meanwhile, prices are thought to have increased.

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Model



(a) NLG pipeline with deep input graph
(b) Pipeline based on the meaning text theory
(c) This paper

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Actions
  - SHIFT-Word-POS [SH]
    - Shifts *Word* from ρ, as- signs POS to it and pushes it to top of stack as S_0;
  - LEFTARC-LABEL [LA]
    - Constructs dependency arc $S_1 \xleftarrow{LABEL} S_0$ and pops out second element from top of stack S_1
  - RIGHTARD-LABEL [RA]
    - Constructs dependency arc $S_1 \xrightarrow{LABEL} S_0$ and pops out top of stack S_0
  - INSERT [IN]
    - Inserts comma at the present position
  - SPLITARC-Word [SP]
    - splits an arc in the input graph *C*, inserting a function word between the words connected by the arc.

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Example

Input Lemmas:

think[1] price[2] .[3] increase[4] be[5] have[6] meanwhile[7]

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action

think[1]	price[2]	.[3]	increase[4]	be[5]	have[6]	meanwhile[7]

σ

ρ

1	2	3	4	5	6	7

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - SH-meanwhile

think$_{[1]}$ price$_{[2]}$.$_{[3]}$ increase$_{[4]}$ be$_{[5]}$ have$_{[6]}$ meanwhile$_{[7]}$

Meanwhile

σ

7

ρ

1	2	3	4	5	6

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - INSERT

$\text{think}_{[1]}$	$\text{price}_{[2]}$	$._{[3]}$	$\text{increase}_{[4]}$	$\text{be}_{[5]}$	$\text{have}_{[6]}$	$\text{meanwhile}_{[7]}$

Meanwhile ,

σ

7

ρ

1	2	3	4	5	6

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - SH-prices

think[1]	price[2]	.[3]	increase[4]	be[5]	have[6]	meanwhile[7]

Meanwhile	,	prices

σ

7	2

ρ

1	3	4	5	6

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - SH-are

think[1]	price[2]	.[3]	increase[4]	be[5]	have[6]	meanwhile[7]

Meanwhile , prices are

σ

7 2 5

ρ

1 3 4 6

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - SH-thought

think[1]	price[2]	.[3]	increase[4]	be[5]	have[6]	meanwhile[7]

Meanwhile	,	prices	are	thought

σ

7	2	5	1

ρ

3	4	6

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - SH-to

think$_{[1]}$ price$_{[2]}$.$_{[3]}$ increase$_{[4]}$ be$_{[5]}$ have$_{[6]}$ meanwhile$_{[7]}$

Meanwhile , prices are thought to

σ

7 2 5 1

ρ

3 4 6

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - SH-have

think$_{[1]}$ price$_{[2]}$.$_{[3]}$ increase$_{[4]}$ be$_{[5]}$ have$_{[6]}$ meanwhile$_{[7]}$

Meanwhile , prices are thought to have

σ

7 2 5 1 6

ρ

3 4

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
    - SH-increased

think[1]	price[2]	.[3]	increase[4]	be[5]	have[6]	meanwhile[7]

Meanwhile	,	prices	are	thought	to	have	increased

σ

7	2	5	1	6	4

ρ

3

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - RA (6 → 4) [VC]

think$_{[1]}$ price$_{[2]}$.$_{[3]}$ increase$_{[4]}$ be$_{[5]}$ have$_{[6]}$ meanwhile$_{[7]}$

VC

Meanwhile , prices are thought to $\boxed{\text{have}}$ $\boxed{\text{increased}}$

σ

7	2	5	1	6	4

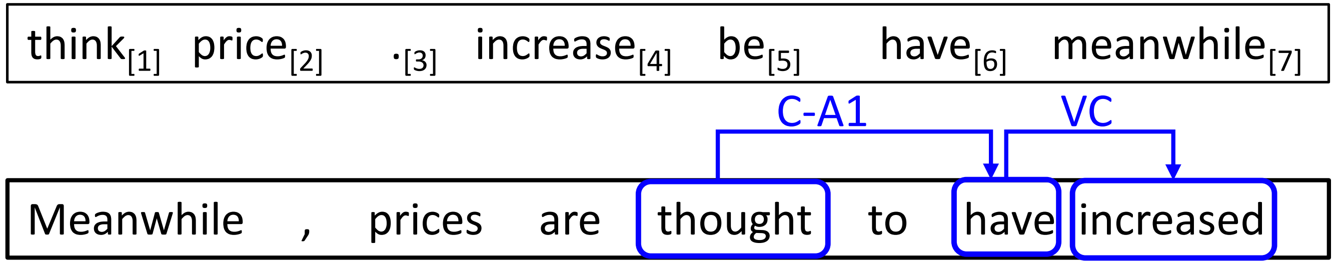
ρ

3

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - RA (1 → 6) [C-A1]



σ

7	2	5	1	6

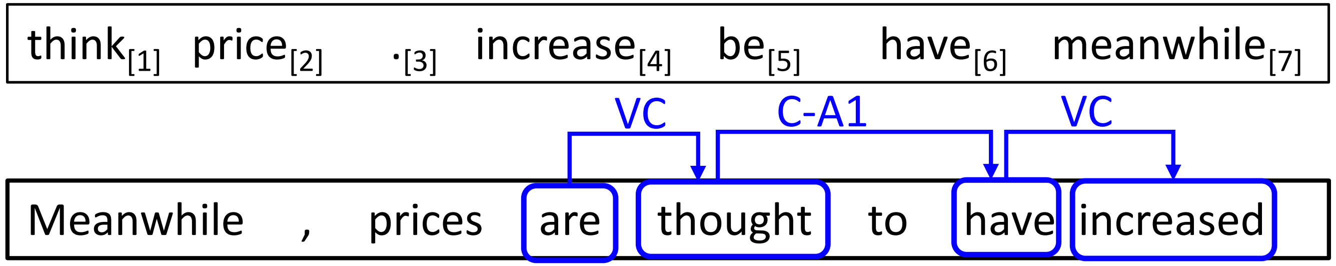
ρ

3

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - RA (5 → 1) [VC]



σ

7	2	5	1

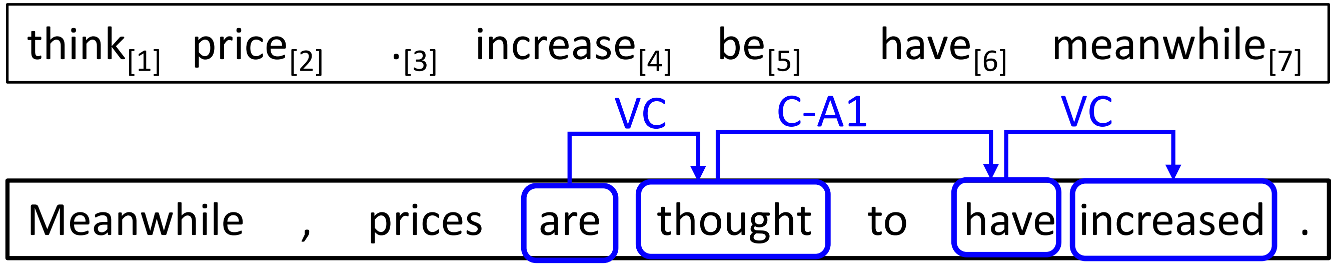
ρ

3

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

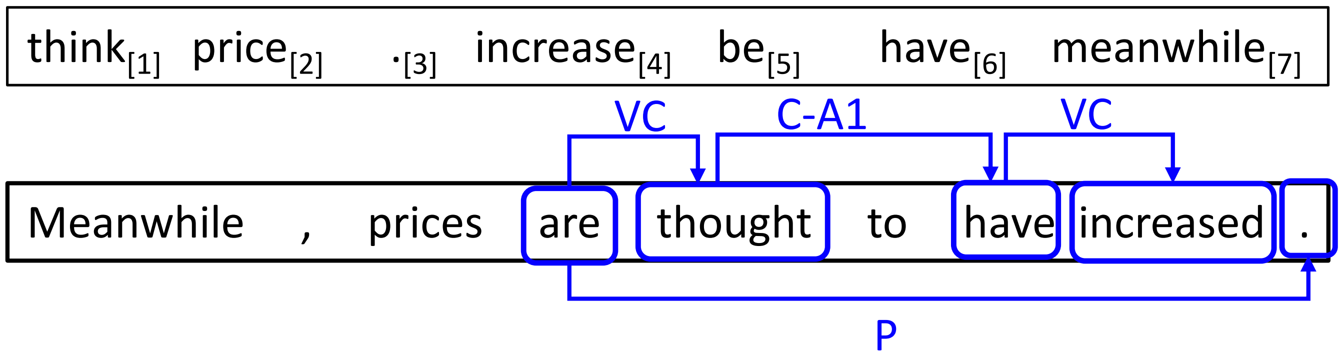
- Transition Action
  - SH-.



Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

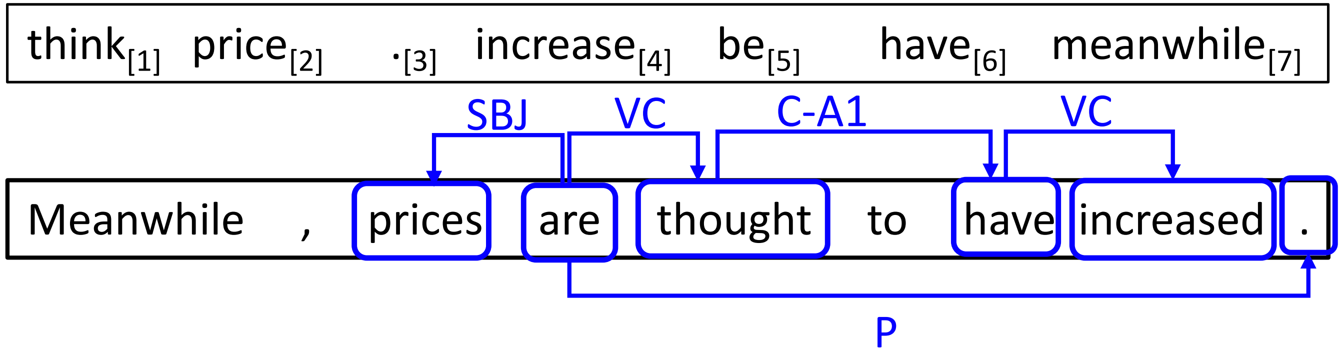
- Transition Action
  - RA (5 → 3) [P]



Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

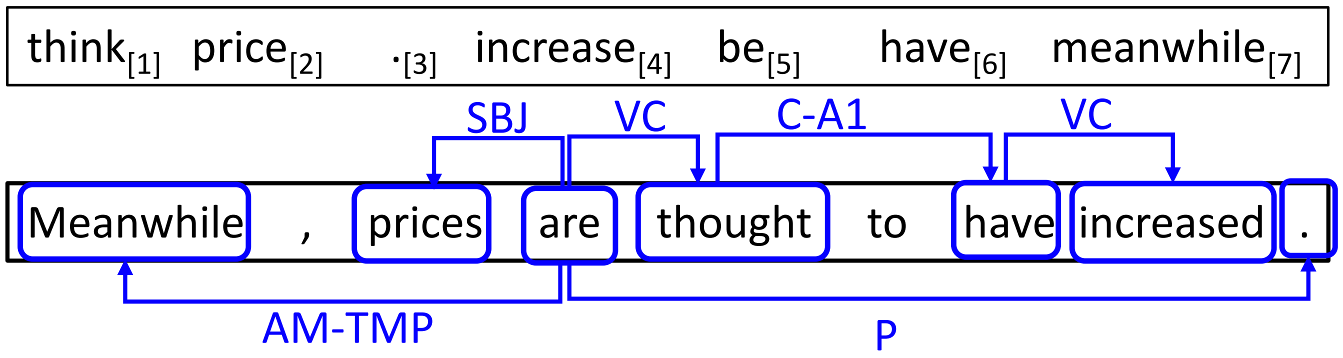
- Transition Action
  - LA (2 ← 5) [SBJ]



Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Transition Action
  - LA (7 ← 5) [AM-TMP]



Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Morphology and Linearization

- Results on dataset of the Surface Realisation Shared Task

System	BLEU Score
STUMABA-D	79.43
Pipeline	70.99
TBDIL	**80.49**

Ratish Puduppully, Yue Zhang, Manish Shrivastava. Transition-Based Deep Input Linearization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain, April.

Joint Entity and Relation Extraction

- Task: simultaneously extracting drugs, diseases and adverse drug events.

Gliclazide$_{\text{drug}}$-induced **acute hepatitis**$_{\text{disease}}$

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- Entity actions:
  - O, which marks the current word as not belong to either a drug or disease mention.
  - BC, which marks the current word as the beginning of a drug mention.
  - BD, which marks the current word as the beginning of a disease mention.
  - I, which marks the current word as part of a drug or disease mention but not the beginning.
- For example
  - Given a sentence: Gliclazide-induced acute hepatitis.
  - The action sequence: “BC O O BD I O “ yields the result ”**Gliclazide**$_{\text{drug}}$-induced **acute hepatitis**$_{\text{disease}}$.”

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- Relation actions
  - N, which indicates that a pair of entities does not have an ADE relation
  - Y, which indicates that a pair of entities has an ADE relation

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- The state of the joint model as a tuple <*labels, disease, drugs, s ADEs*>
  - *labels* is a label sequence
  - *disease* is a list of readily-recognized disease entity mentions
  - *drugs* is a list of readily-recognized drug entity mentions
  - *ADEs* is a set of ADEs

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis	caused	by	methotrexate	and	etretinate	.
BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

next action

state	next action
<[],[],[],[]>	BD

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis caused by methotrexate and etretinate .

BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

<[BD],[],[],[]>

next action

O

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis	caused	by	methotrexate	and	etretinate	.
BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

state	next action
<[BD,O],[Hepatitis],[],[]>	O

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis	caused	by	methotrexate	and	etretinate	.
BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

<[BD,O,O],[Hepatitis],[],[]>

next action

BC

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis	caused	by	methotrexate	and	etretinate	.
BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

<[BD,O,O,BC],[Hepatitis],[],[]>

next action

O

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis	caused	by	methotrexate	and	etretinate	.
BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

<[BD,O,O,BC,O],[Hepatitis],[methotrexate],[]>

next action

Y

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis	caused	by	methotrexate	and	etretinate	.
BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

<[BD,O,O,BC,O,Y],[Hepatitis],[methotrexate],[(Hepatitis,methotrexate)]>

next action

BC

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis caused by methotrexate and etretinate .

BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

<[BD,O,O,BC,O,Y,BC],[Hepatitis],[methotrexate],[(Hepatitis,methotrexate)]>

next action

O

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis	caused	by	methotrexate	and	etretinate	.
BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

<[BD,O,O,BC,O,Y,BC,O],[Hepatitis],[methotrexate,etretinate],[(Hepatitis, methotrexate)]>

next action

Y

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- State transition examples

Hepatitis	caused	by	methotrexate	and	etretinate	.
BD	O	O	BC	O	BC	O

state <labels, disease, drugs, relations>

<[BD,O,O,BC,O,Y,BC,O,Y],[Hepatitis],[methotrexate,etretinate],[(Hepatitis,methotrexate),(Hepatitis,etretinate)]>

next action

<EOS>

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- Results on ADE data

Method	Entity Recognition			ADE extraction		
	P	**R**	**F_1**	**P**	**R**	**F_1**
Li *et al.* [2015]	75.9	71.6	73.6	55.2	47.9	51.1
Baseline	77.8	72.0	74.8	60.7	51.5	55.7
Discrete Joint	**80.0**	75.1	77.5	**65.1**	56.7	60.6

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Outline

- Motivation
- Statistical Models
- Deep Learning Models

Deep Learning Models

- Neural Transition-based Models
- Neural Graph-based Models (Multi-task Learning)
  - Cross Task
  - Cross Domain
  - Cross Lingual
  - Cross Standard

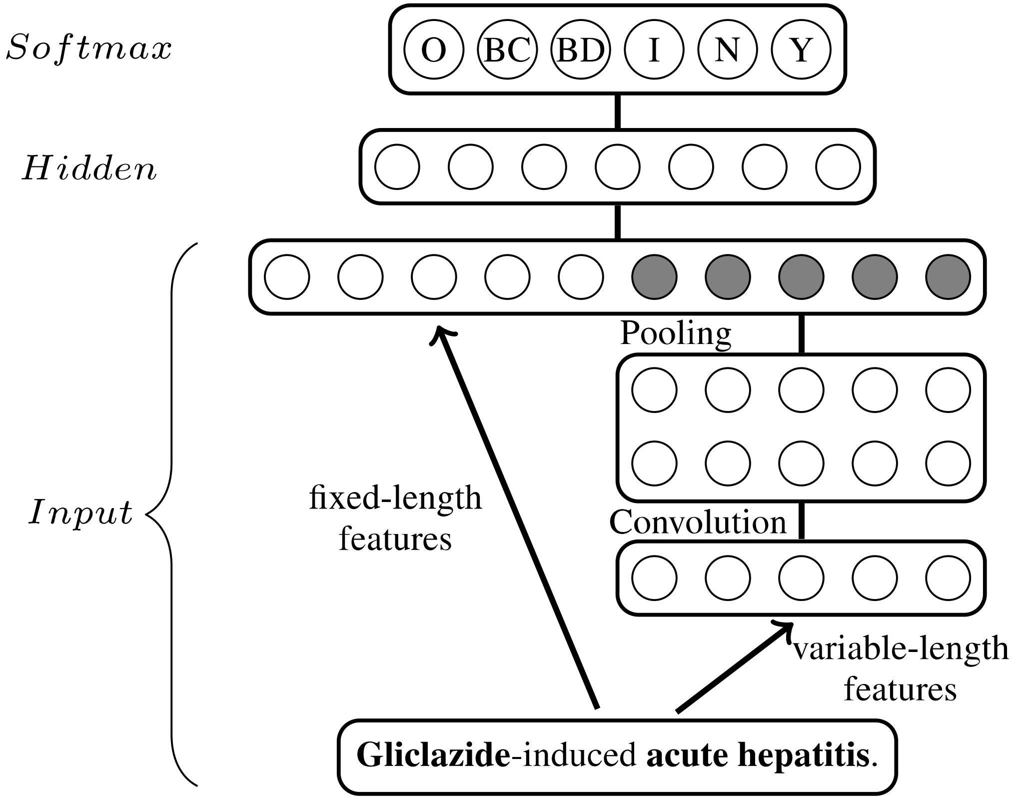
Deep Learning Models

- Neural Transition-based Models
- Neural Graph-based Models (Multi-
  - Cross Task
  - Cross Domain
  - Cross Ling
  - Cr

Joint Learning **Joint Search**

Joint Entity and Relation Extraction

- Model



Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Entity and Relation Extraction

- Results

Method	Entity Recognition			ADE extraction		
	P	**R**	**F_1**	**P**	**R**	**F_1**
Li *et al.* [2015]	75.9	71.6	73.6	55.2	47.9	51.1
Baseline	77.8	72.0	74.8	60.7	51.5	55.7
Discrete Joint	**80.0**	75.1	77.5	**65.1**	56.7	60.6
Neural Joint	79.5	**79.6**	**79.5**	64.0	**62.9**	**63.4**

Fei Li, Yue Zhang, Meishan Zhang and Donghong Ji. *Joint Models for Extracting Adverse Drug Events from Biomedical Text*. In Proceddings of IJCAI 2016. New York City, USA, July.

Joint Parsing and SRL

- Transition Action

all are expected to reopen soon root

Stack [S]

Buffer [M]

Queue [B]

all, are, expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-SHIFT

all are expected to reopen soon root

Stack [S]	Buffer [M]	Queue [B]
all		are, expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-SHIFT

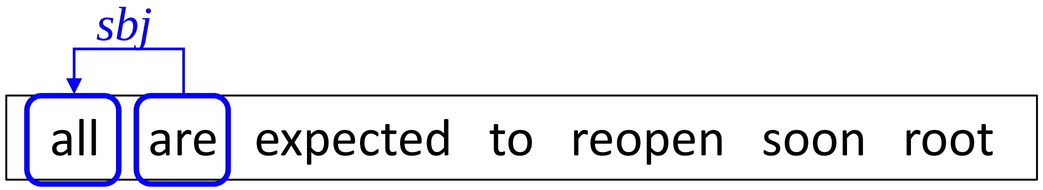
all	are	expected	to	reopen	soon	root

Stack [S]	Buffer [M]	Queue [B]
all	all	are, expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-LEFT (*sbj*)



Stack [S]

Buffer [M]

all

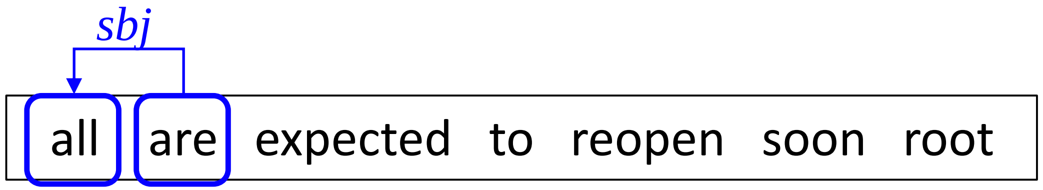
Queue [B]

are, expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-SHIFT



Stack [S]	Buffer [M]	Queue [B]
are	all	are, expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-SHIFT

sbj

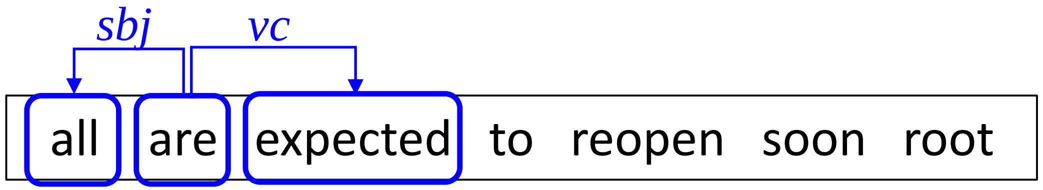
all are expected to reopen soon root

Stack [S]	Buffer [M]	Queue [B]
are	all, are	expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-RIGHT (*vc*)



Stack [S]

are, expected

Buffer [M]

all, are

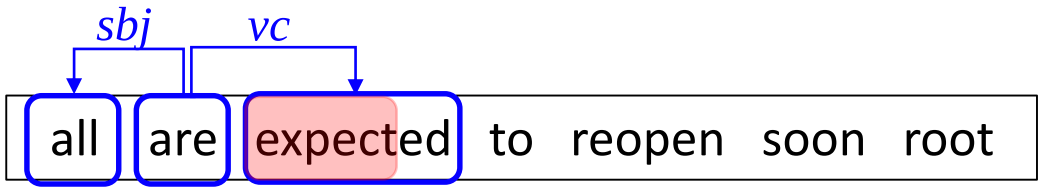
Queue [B]

expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-PRED (***expect.01***)



Stack [S]

are, expected

Buffer [M]

all, are

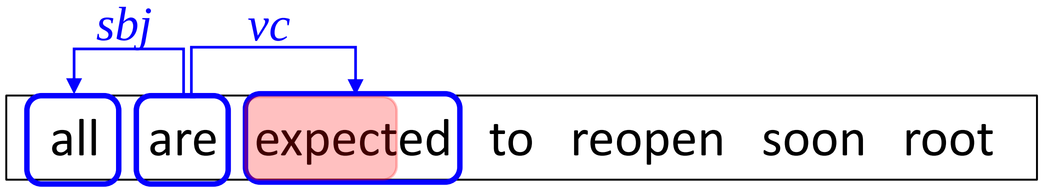
Queue [B]

expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-REDUCE

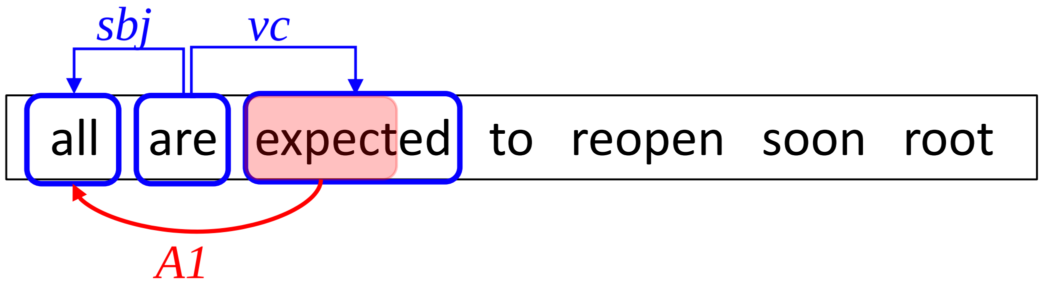


Stack [S]	Buffer [M]	Queue [B]
are, expected	all	expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-LEFT(*A1*)



Stack [S]: are, expected

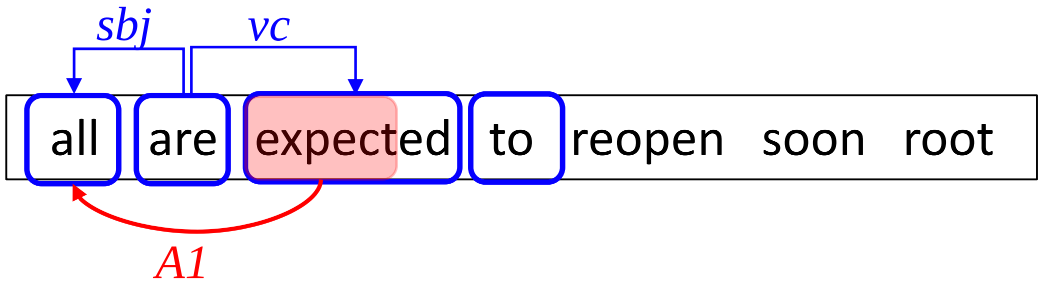
Buffer [M]: all

Queue [B]: expected, to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-SHIFT

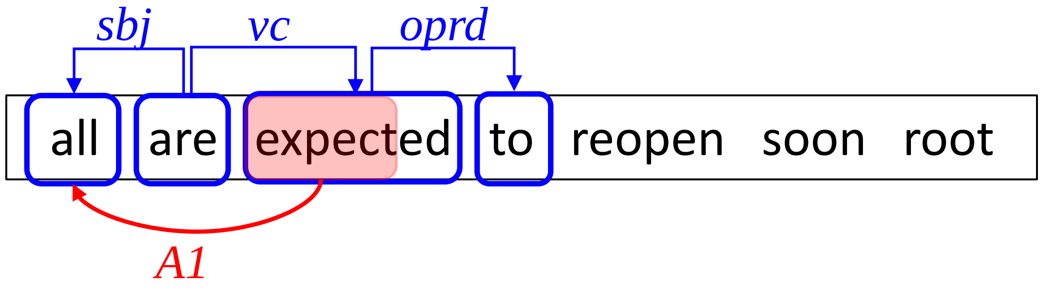


Stack [S]	Buffer [M]	Queue [B]
are, expected	all, expected	to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
    - S-RIGHT (*oprd*)



Stack [S]

are, expected, to

Buffer [M]

all, expected

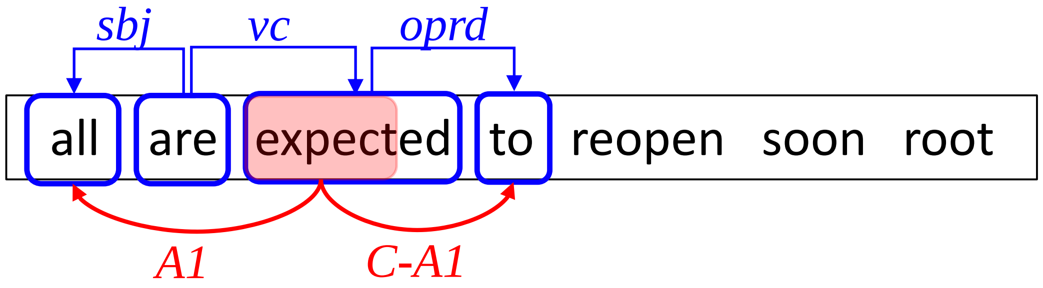
Queue [B]

to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-RIGHT (*C-A1*)

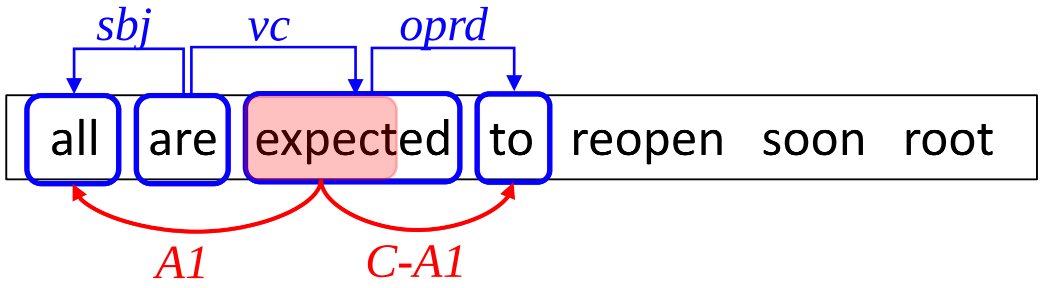


Stack [S]	Buffer [M]	Queue [B]
are, expected, to	all, expected	to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-REDUCE



Stack [S]

are, expected, to

Buffer [M]

all

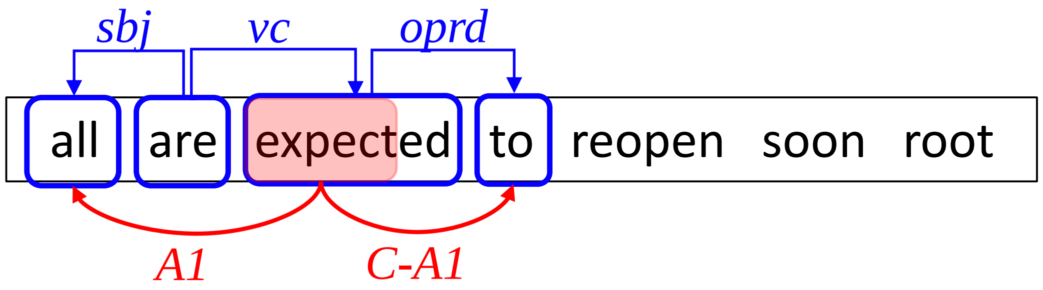
Queue [B]

to, reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
    - M-SHIFT

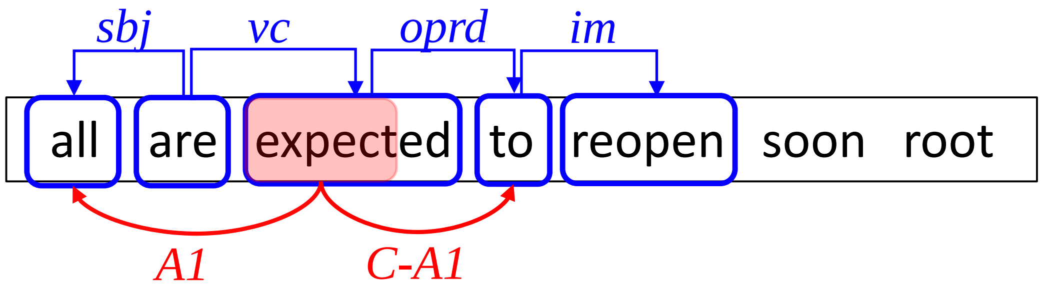


Stack [S]	Buffer [M]	Queue [B]
are, expected, to	all, to	reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-RIGHT (*im*)



Stack [S]

are, expected, to, reopen

Buffer [M]

all, to

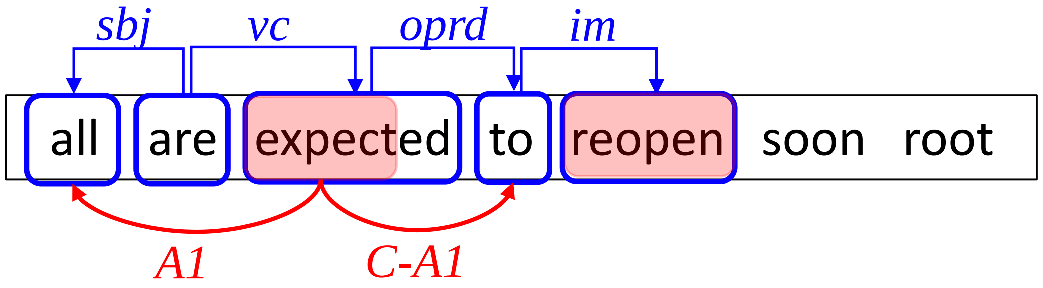
Queue [B]

reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-PRED (***reopen.01***)



Stack [S]

are, expected, to, reopen

Buffer [M]

all, to

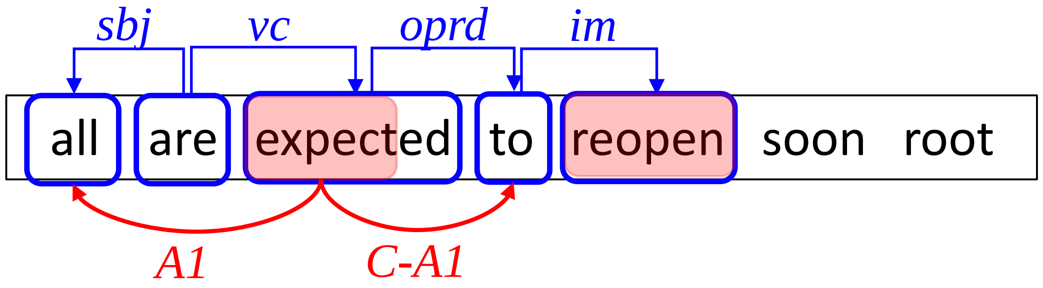
Queue [B]

reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-REDUCE



Stack [S]

are, expected, to, reopen

Buffer [M]

all

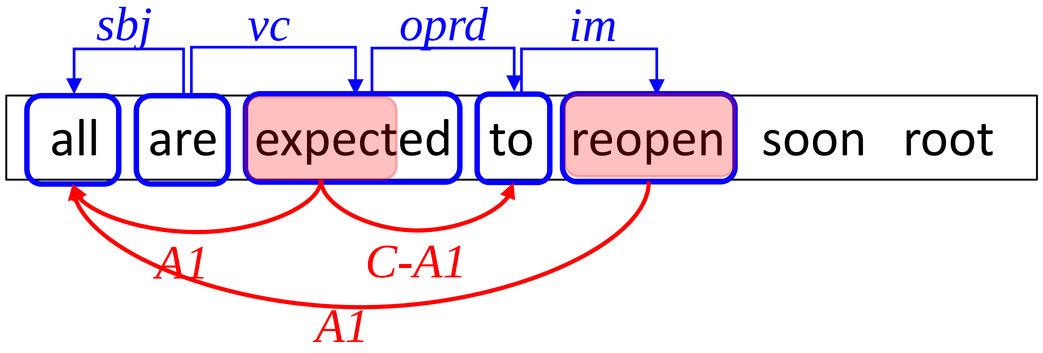
Queue [B]

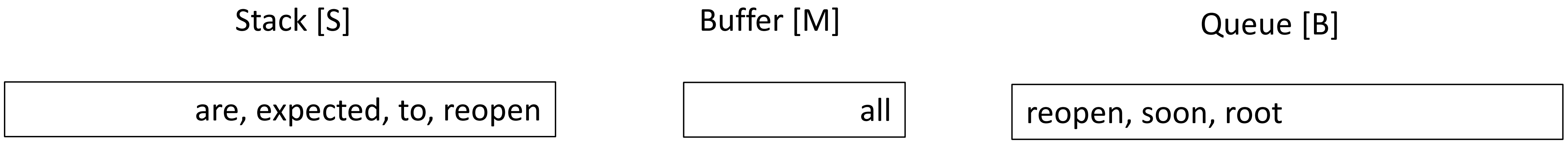
reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-LEFT (*A1*)

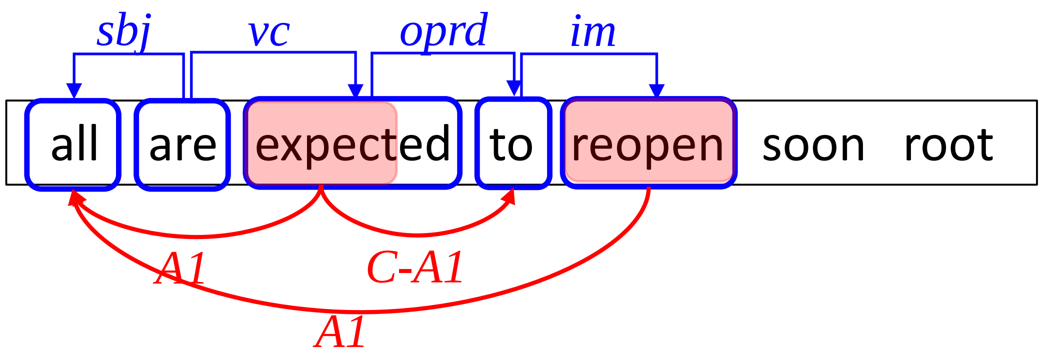




Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-REDUCE



Stack [S]

are, expected, to, reopen

Buffer [M]

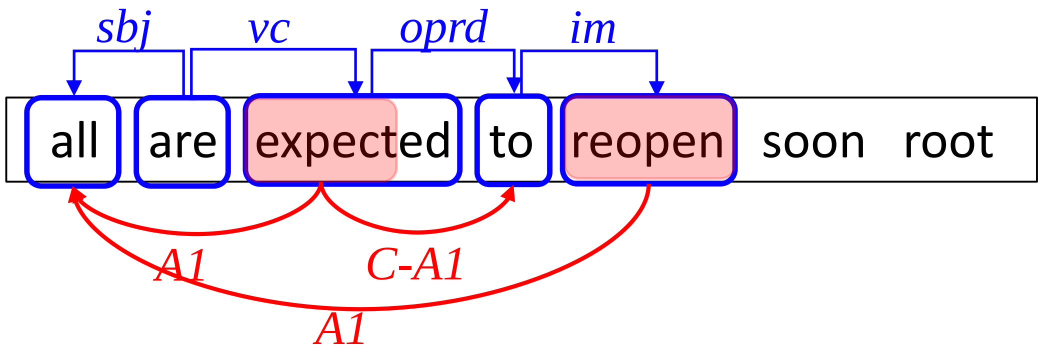
Queue [B]

reopen, soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-SHIFT

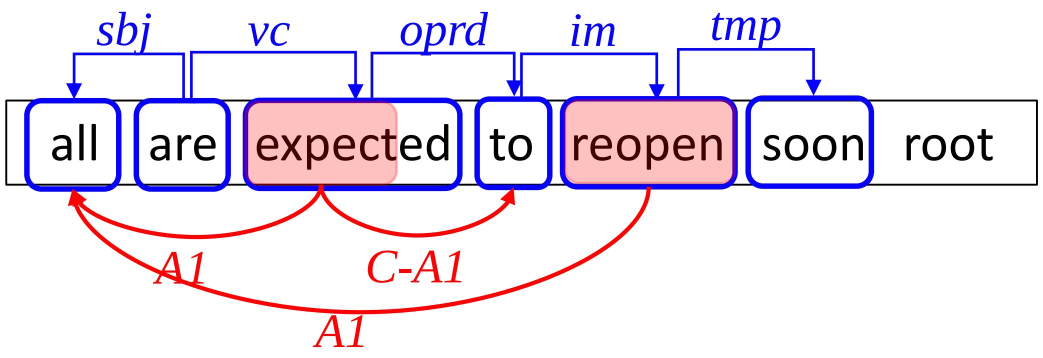


Stack [S]	Buffer [M]	Queue [B]
are, expected, to, reopen	reopen	soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-RIGHT (*tmp*)



Stack [S]

are, expected, to, reopen, soon

Buffer [M]

reopen

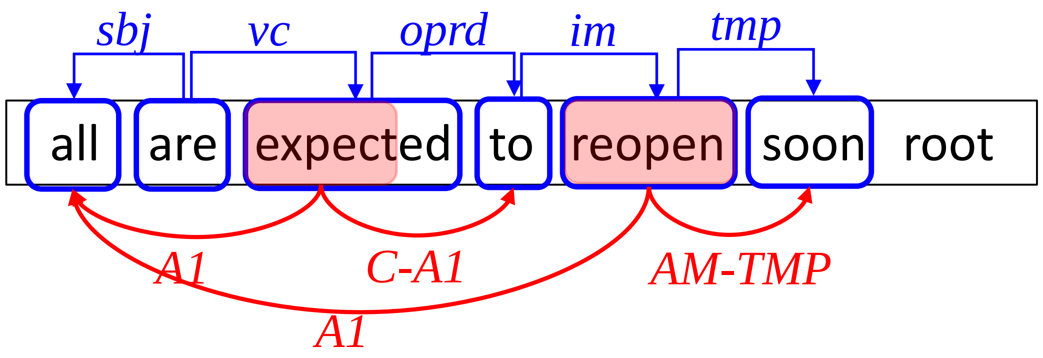
Queue [B]

soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-RIGHT (*AM-TMP*)



Stack [S]

are, expected, to, reopen, soon

Buffer [M]

reopen

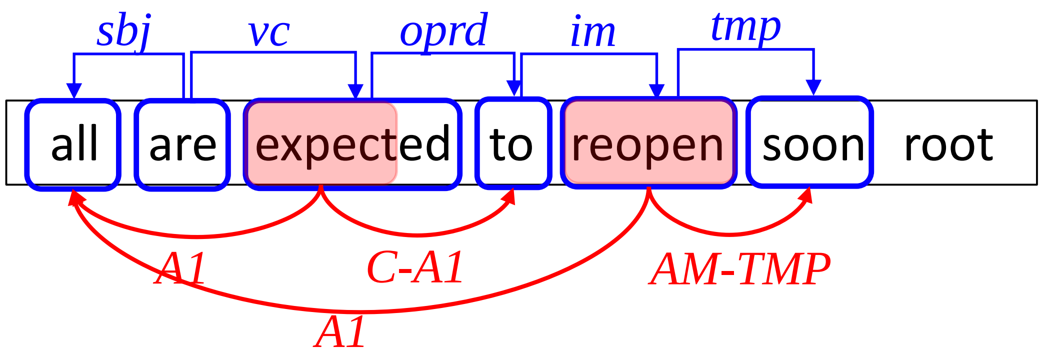
Queue [B]

soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-REDUCE



Stack [S]

are, expected, to, reopen, soon

Buffer [M]

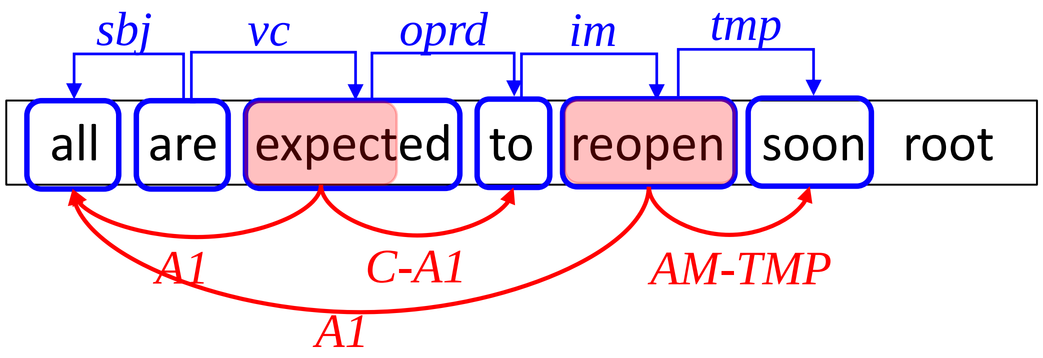
Queue [B]

soon, root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-SHIFT

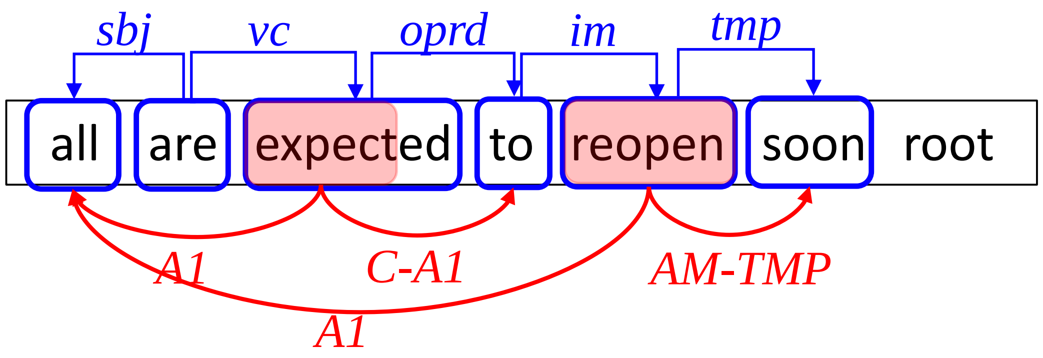


Stack [S]	Buffer [M]	Queue [B]
are, expected, to, reopen	soon	root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-REDUCE



Stack [S]

are, expected, to, reopen

Buffer [M]

soon

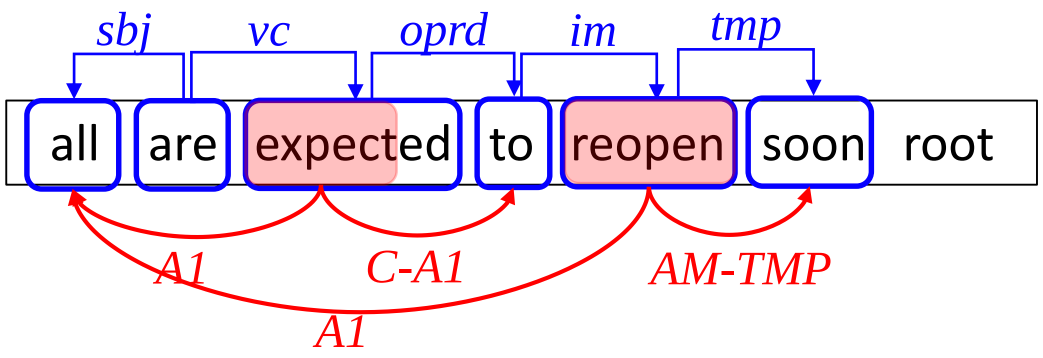
Queue [B]

root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-REDUCE



Stack [S]

are, expected, to

Buffer [M]

soon

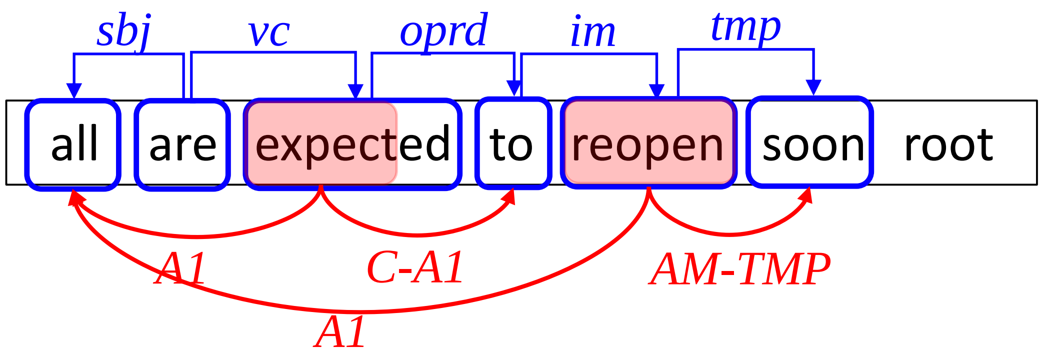
Queue [B]

root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-REDUCE

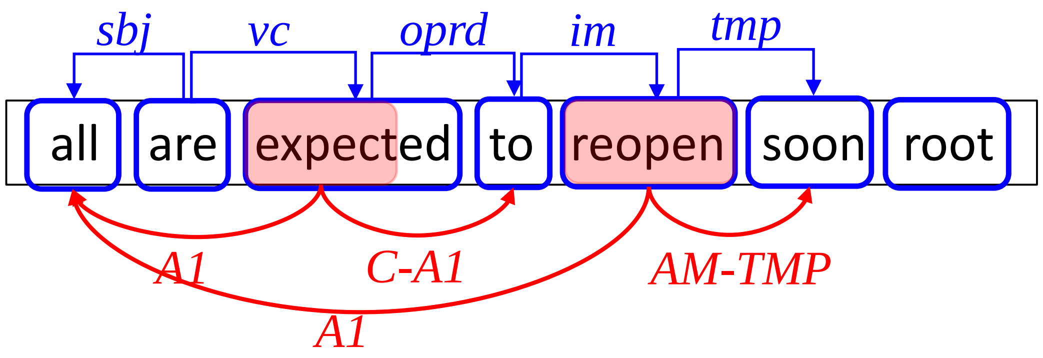


Stack [S]	Buffer [M]	Queue [B]
are, expected	soon	root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-REDUCE

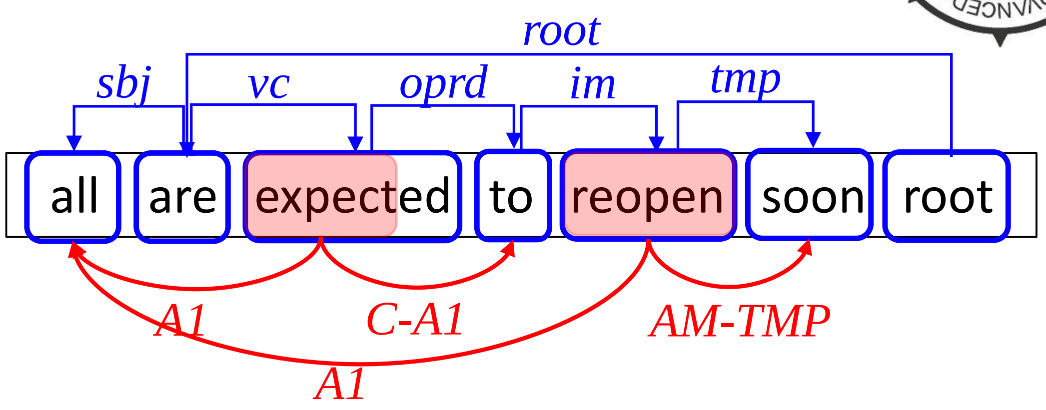


Stack [S]	Buffer [M]	Queue [B]
are	soon	root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-LEFT (*root*)



Stack [S]

Buffer [M]

soon

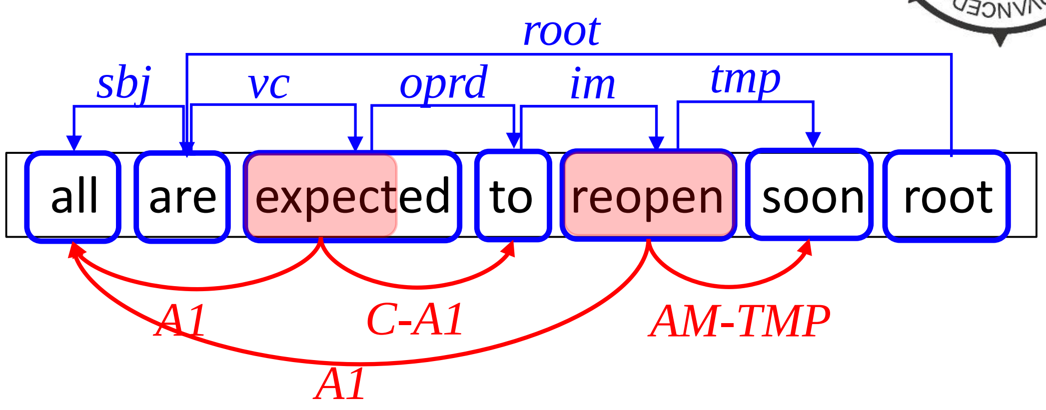
Queue [B]

root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - S-SHIFT

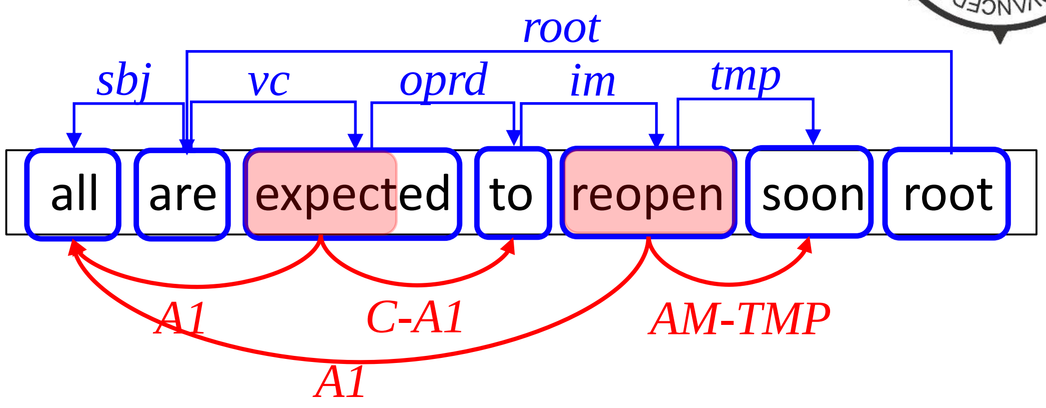


Stack [S]	Buffer [M]	Queue [B]
root	soon	root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-REDUCE



Stack [S]

root

Buffer [M]

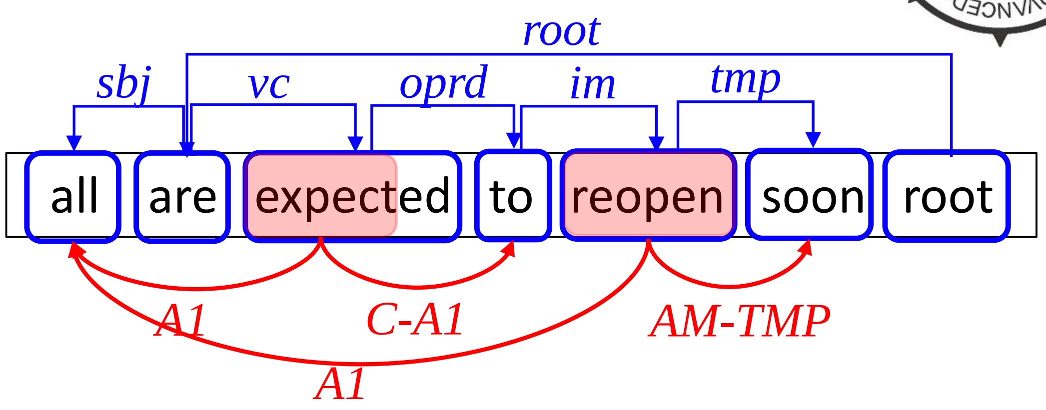
Queue [B]

root

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Transition Action
  - M-SHIFT



Stack [S]

root

Buffer [M]

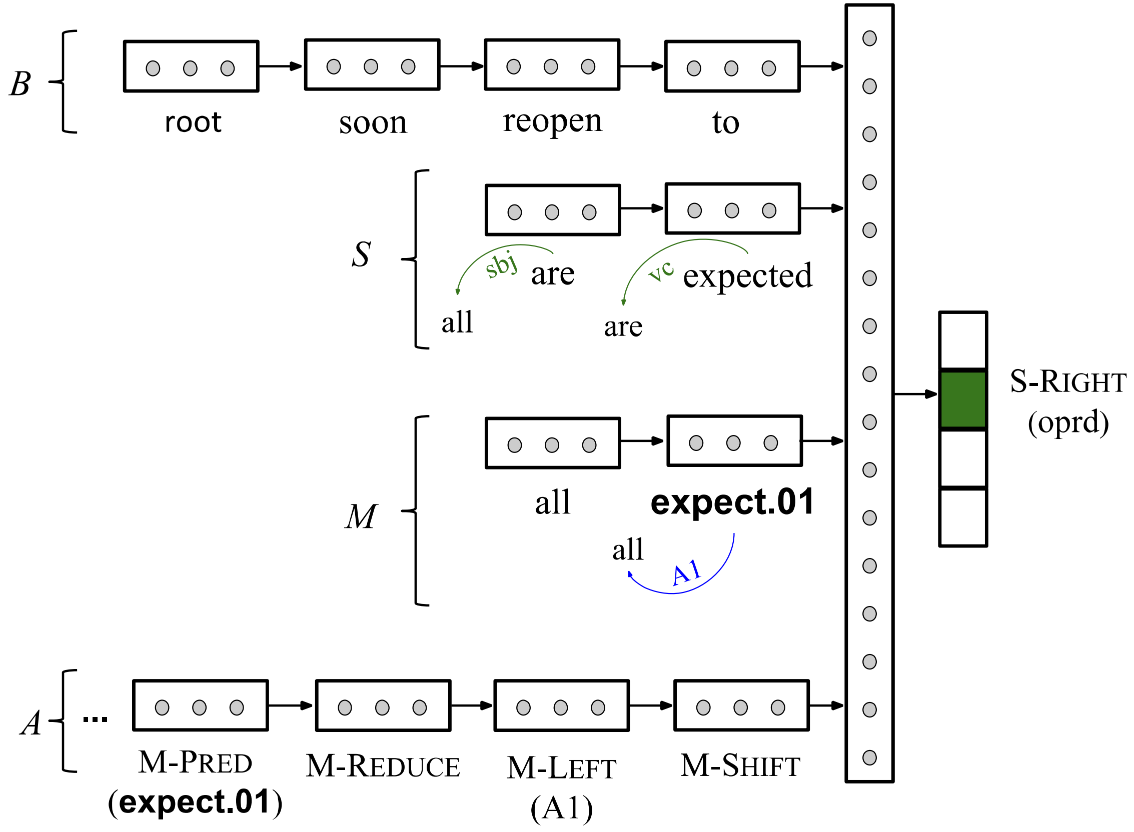
root

Queue [B]

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Model



Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

- Results on CONLL

Model	**LAS**	**Sem.** F_1	**Macro** F_1
joint models:			
Lluís and Màrquez (2008)	85.8	70.3	78.1
Henderson et al. (2008)	87.6	73.1	80.5
Johansson (2009)	86.6	77.1	81.8
Titov et al. (2009)	87.5	76.1	81.8
CoNLL 2008 best:			
#3: Zhao and Kit (2008)	87.7	76.7	82.2
#2: Che et al. (2008)	86.7	78.5	82.7
#2: Ciaramita et al. (2008)	87.4	78.0	82.7
#1: J&N (2008)	89.3	81.6	85.5
Joint (this work)	89.1	80.5	84.9

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Parsing and SRL

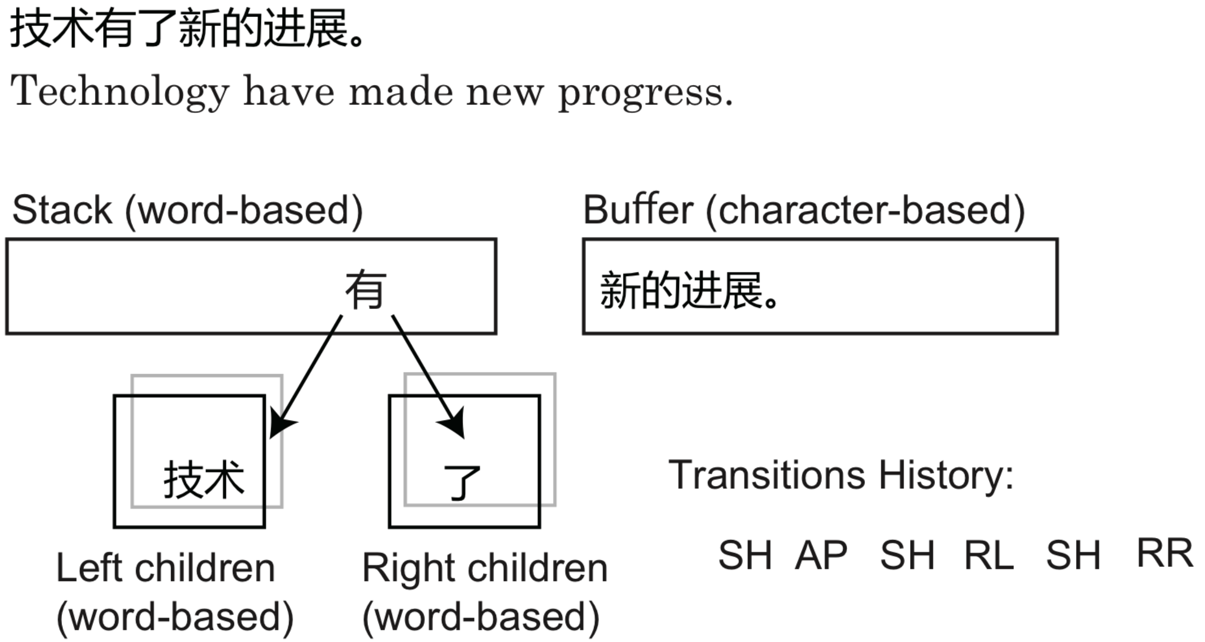
- Joint VS Pipeline

Model	**LAS**	**Sem. F_1 (WSJ)**	**Sem. F_1 (Brown)**	**Macro F_1**
CoNLL'09 best:				
#3 G+ '09	88.79	83.24	70.65	86.03
#2 C+ '09	88.48	85.51	73.82	87.00
#1 Z+ '09a	89.19	86.15	74.58	87.69
this work:				
Syntax-only	89.83			
Sem.-only		84.39	73.87	
Hybrid	89.83	84.58	75.64	87.20
Joint	89.94	84.97	74.48	87.45
pipelines:				
R&W '14		86.34	75.90	
L+ '15		86.58	75.57	
T+ '15		87.30	75.50	
F+ '15		87.80	75.50	

Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer and Noah A. Smith. Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs In proceedings of CoNLL (CoNLL 2016).

Joint Word Segmentation, POS Tagging and Dependency Parsing

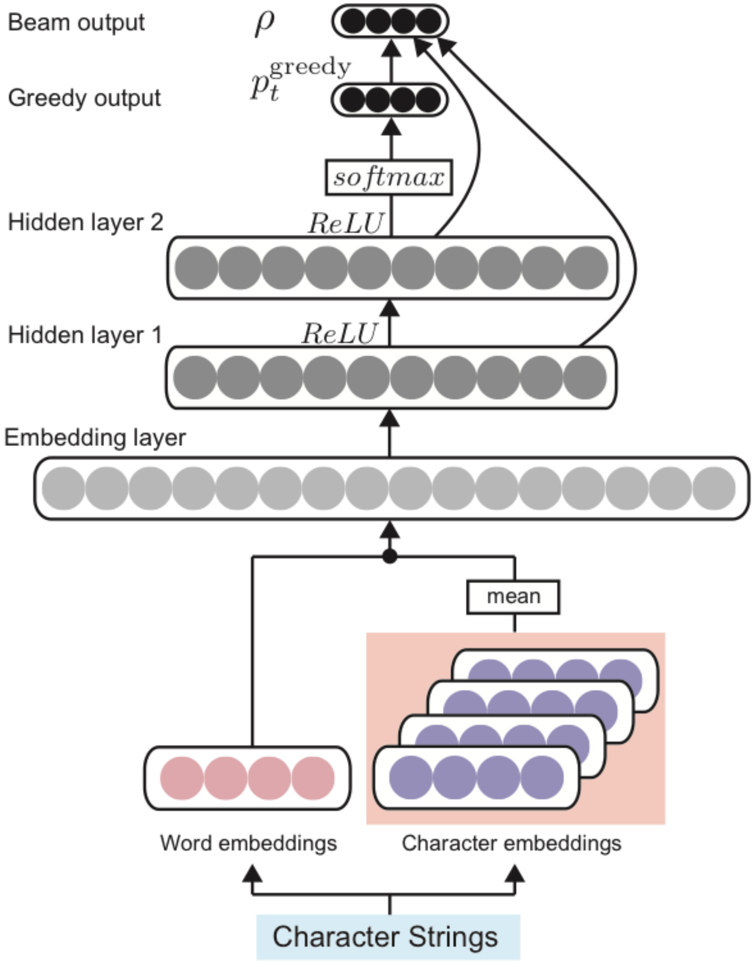
- Model



Kurita, Shuhei, Daisuke Kawahara, and Sadao Kurohashi. "Neural Joint Model for Transition-based Chinese Syntactic Analysis." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics.* Vol. 1. 2017.

Joint Word Segmentation, POS Tagging and Dependency Parsing

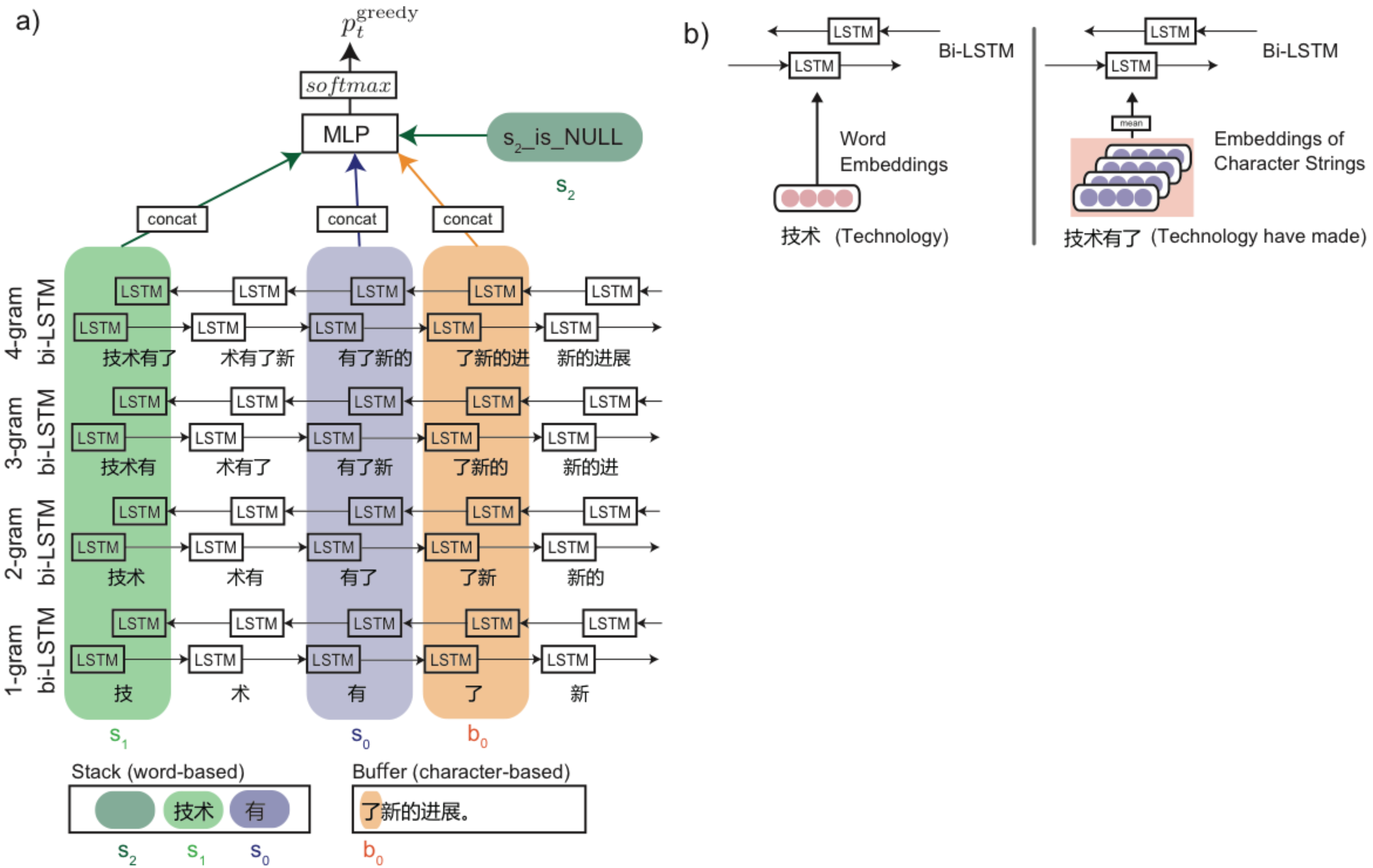
- Feed-forward NN model



Kurita, Shuhei, Daisuke Kawahara, and Sadao Kurohashi. "Neural Joint Model for Transition-based Chinese Syntactic Analysis." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics.* Vol. 1. 2017.

Joint Word Segmentation, POS Tagging and Dependency Parsing

- The bi-LSTM model



Kurita, Shuhei, Daisuke Kawahara, and Sadao Kurohashi. "Neural Joint Model for Transition-based Chinese Syntactic Analysis." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics.* Vol. 1. 2017.

Joint Word Segmentation, POS Tagging and Dependency Parsing

- Results on PTB

Model	Seg	POS	Dep
Hatori+12	97.75	94.33	81.56
M. Zhang+14 STD	97.67	94.28	81.63
M. Zhang+14 EAG	97.76	94.36	81.70
Y. Zhang+15	98.04	94.47	82.01
SegTagDep(g)	98.24	94.49	80.15
SegTagDep	98.37	**94.83**‡	81.42‡
SegTag+Dep	**98.60**‡	94.76‡	**82.60**‡

Kurita, Shuhei, Daisuke Kawahara, and Sadao Kurohashi. "Neural Joint Model for Transition-based Chinese Syntactic Analysis." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics.* Vol. 1. 2017.

Joint Extraction of Entities and Relations

- Transition Actions
  - Initialization

$\text{John}_{[1]}$ $\text{lives}_{[2]}$ $\text{in}_{[3]}$ $\text{Los}_{[4]}$ $\text{Angeles}_{[5]}$ $\text{California}_{[6]}$

σ	δ	e	β
			1, . . . , 6

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - GEN-SHIFT

$\boxed{\text{John}_{[1]}}$ lives$_{[2]}$ in$_{[3]}$ Los$_{[4]}$ Angeles$_{[5]}$ California$_{[6]}$

σ	δ	e	β
		1	2, . . . , 6

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - GEN-NER

Per

John$_{[1]}$ lives$_{[2]}$ in$_{[3]}$ Los$_{[4]}$ Angeles$_{[5]}$ California$_{[6]}$

σ	δ	e	β
			$1^*, \ldots, 6$

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - NO-SHIFT

Per

$\boxed{\text{John}_{[1]}}$ lives$_{[2]}$ in$_{[3]}$ Los$_{[4]}$ Angeles$_{[5]}$ California$_{[6]}$

σ	δ	e	β
1^*			2, . . . , 6

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - O-DELETE

Per

John$_{[1]}$ lives$_{[2]}$ in$_{[3]}$ Los$_{[4]}$ Angeles$_{[5]}$ California$_{[6]}$

σ	δ	e	β
1*			3, . . . , 6

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - O-DELETE

Per

John[1] lives[2] in[3] Los[4] Angeles[5] California[6]

σ	δ	e	β
1*			4, . . . , 6

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - GEN-SHIFT

Per

John$_{[1]}$ lives$_{[2]}$ in$_{[3]}$ Los$_{[4]}$ Angeles$_{[5]}$ California$_{[6]}$

σ	δ	e	β
1^*		4	5 , 6

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - GEN-SHIFT

Per

John[1] lives[2] in[3] Los[4] Angeles[5] California[6]

σ	δ	e	β
1*		4, 5	6

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - GEN-NER

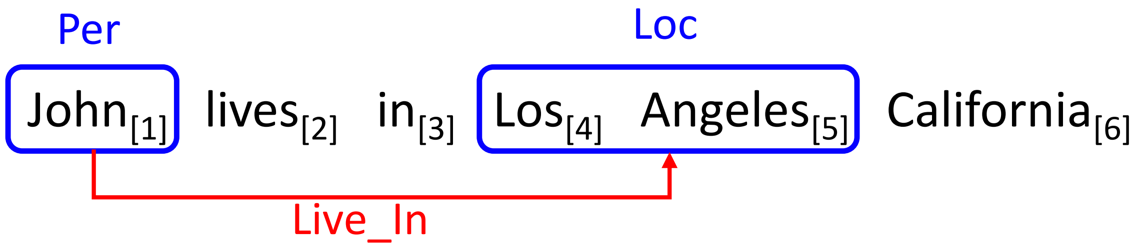
Per: John$_{[1]}$ lives$_{[2]}$ in$_{[3]}$ Loc: Los$_{[4]}$ Angeles$_{[5]}$ California$_{[6]}$

σ	δ	e	β
1^*			5^*, 6

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - RIGHT-SHIFT

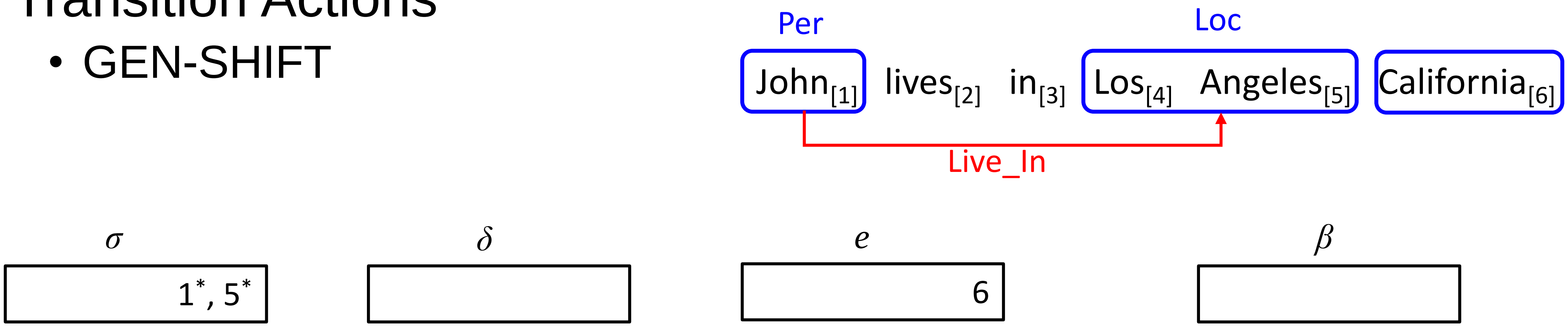


σ	δ	e	β
$1^*, 5^*$			6

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

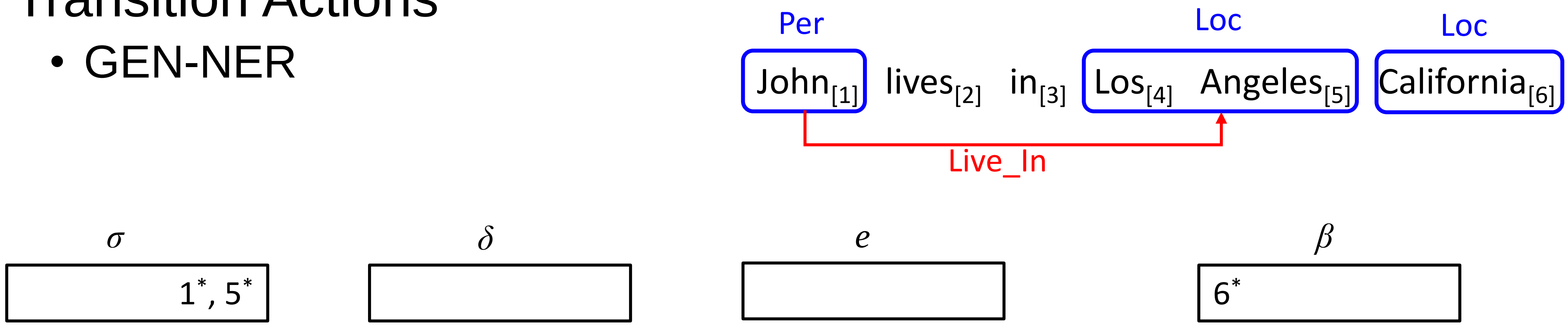
- Transition Actions
  - GEN-SHIFT



Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

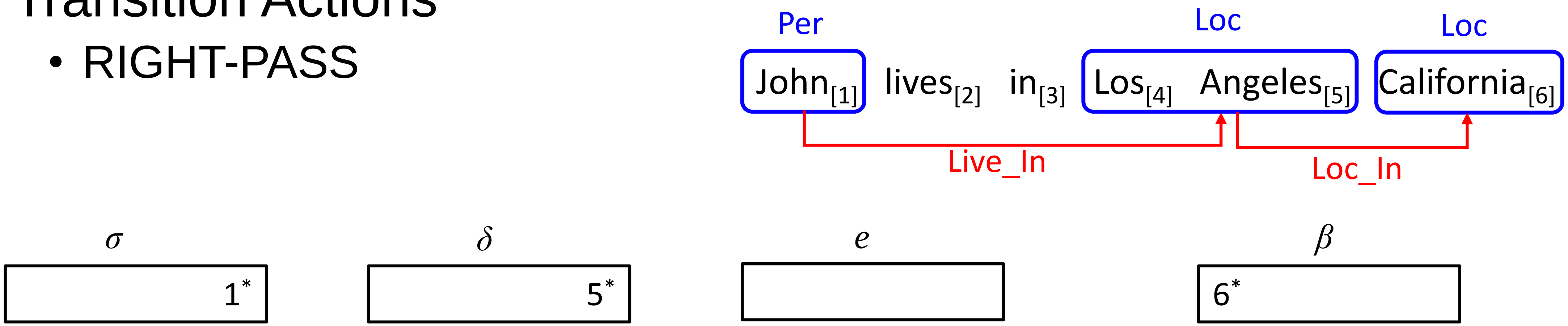
- Transition Actions
  - GEN-NER



Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

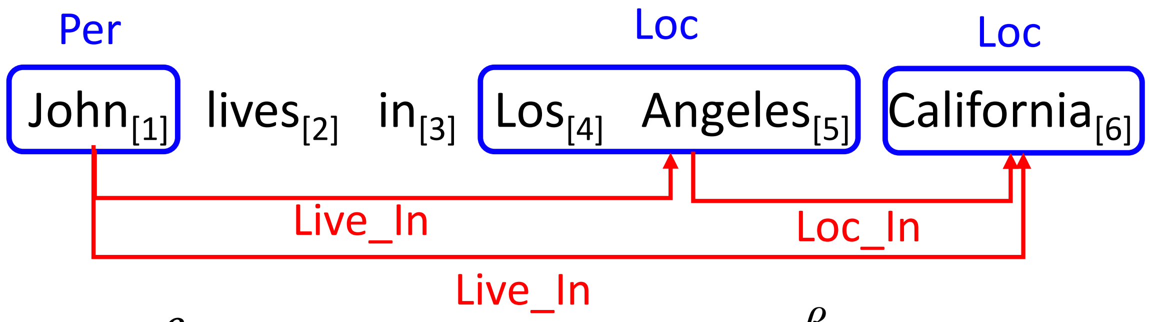
- Transition Actions
  - RIGHT-PASS

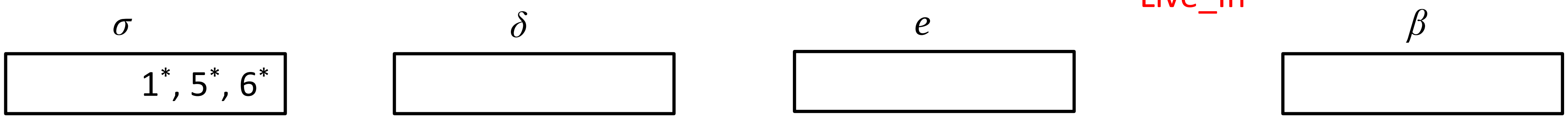


Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Transition Actions
  - RIGHT-SHIFT

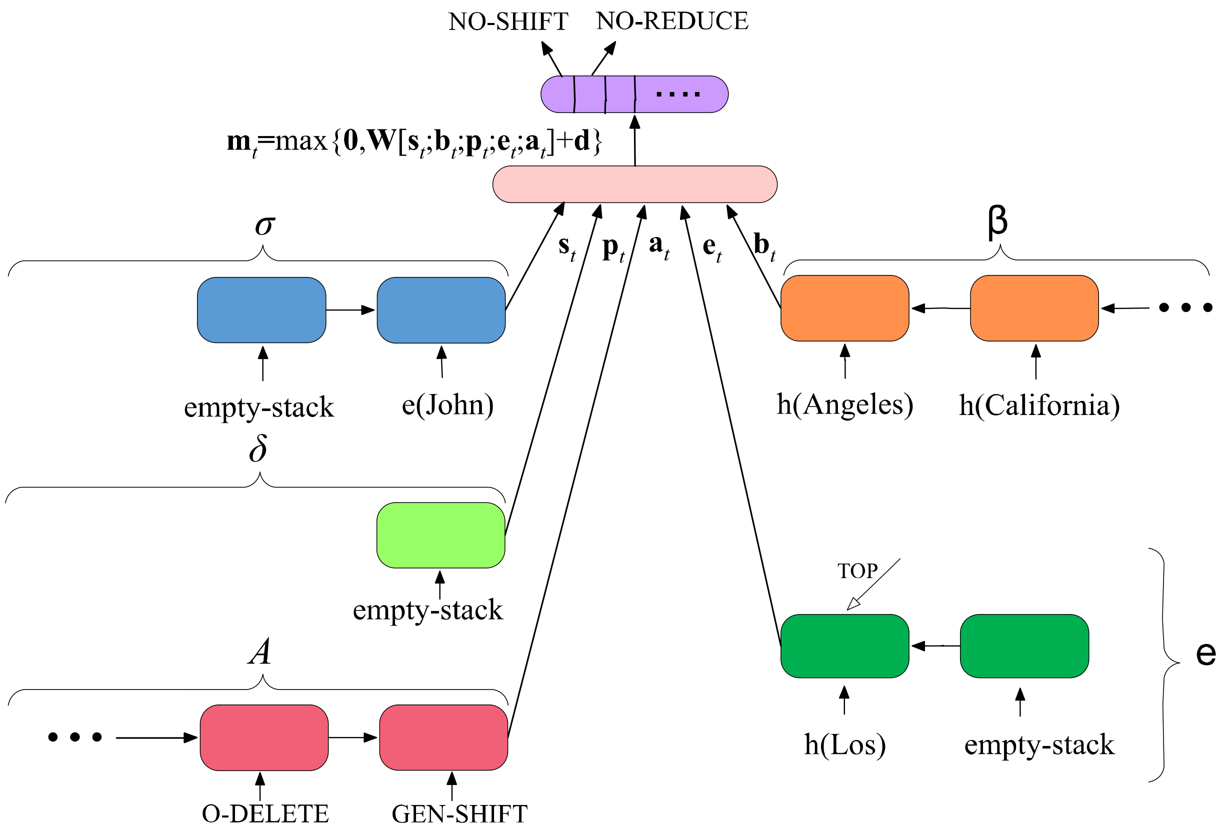




Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Model



Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Joint Extraction of Entities and Relations

- Results on NYT

Method	Prec.	Rec.	F1
FCM [Gormley *et al.*, 2015]	55.3	15.4	24.0
DS+logistic [Mintz *et al.*, 2009]	25.8	39.3	31.1
LINE [Tang *et al.*, 2015]	33.5	32.9	33.2
MultiR [Hoffmann *et al.*, 2011]	33.8	32.7	33.3
DS-Joint [Li and Ji, 2014]	57.4	25.6	35.4
CoType [Ren *et al.*, 2017]	42.3	51.1	46.3
LSTM-LSTM-Bias	61.5	41.4	49.5
LSTM-LSTM-Bias*	60.8	41.3	49.1
Our Method	**64.3**	**42.1**	**50.9**

Shaolei Wang, Yue Zhang, Wanxiang Che, Ting Liu. 2018. Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. In Proceedings of 27th IJCAI-ECAI

Deep Learning Models

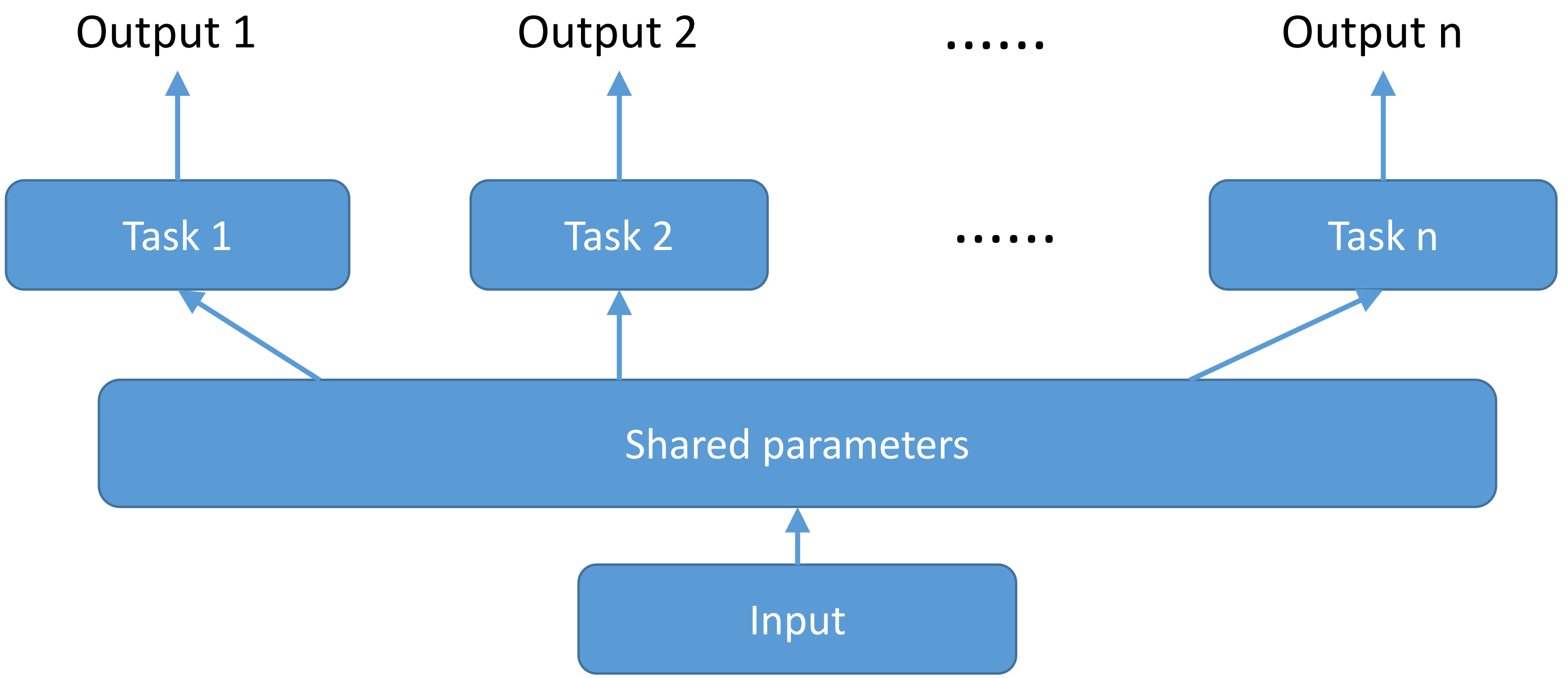
- Neural Transition-based Models
- Neural Graph-based Models (Multi-task Learning)
  - Cross Task
  - Cross Lingual
  - Cross Domain
  - Cross Standard

Deep Learning Models

- Neural Transition-based Models
- Neural Graph-based Models (Multi-
  - Cross Task
  - Cross Lingual
  - Cross Dom
  - Cr

Joint Learning Separate Search

Neural Graph-based Models (Multi-task Learning)



Neural Graph-based Models (Multi-task Learning)

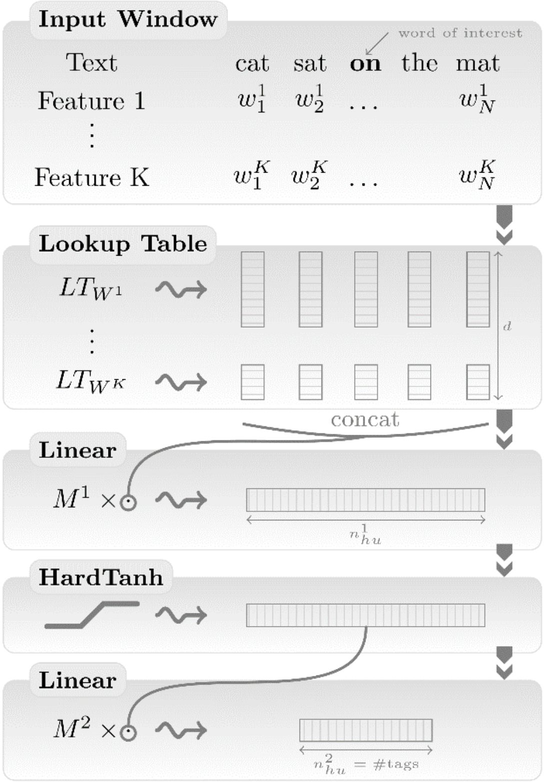
- Cross Task
- Cross Lingual
- Cross Domain
- Cross Standard

Neural Graph-based Models (Multi-task Learning)

- Cross Task
- Cross Lingual
- Cross Domain
- Cross Standard

Joint Tagging, Chunking and NER

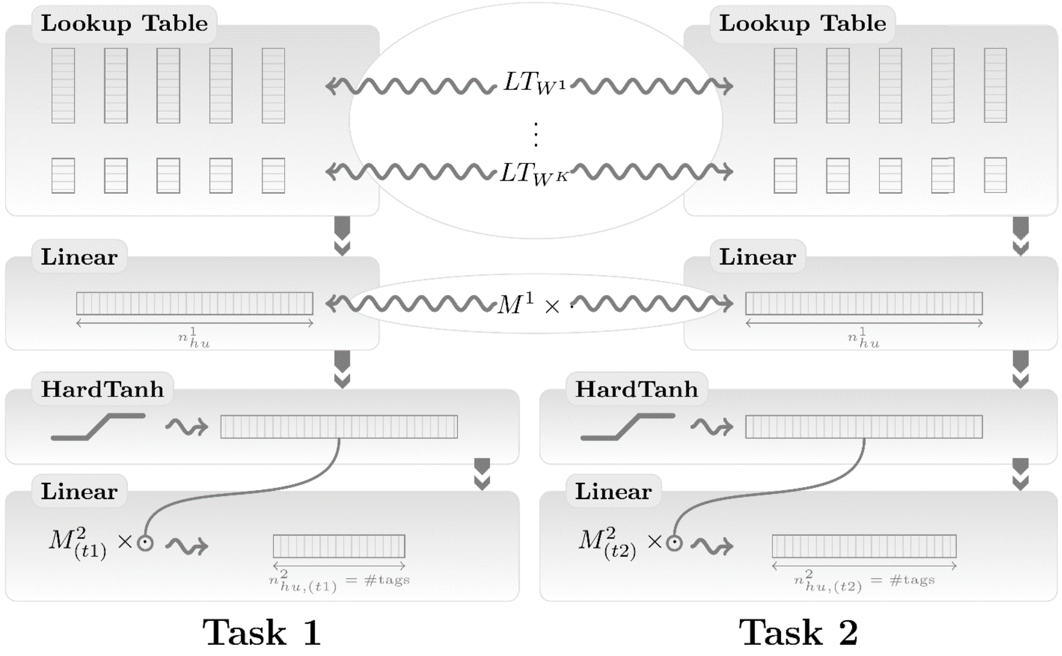
- Seminal work in NLP



Collobert, Ronan, et al. "Natural language processing (almost) from scratch." *Journal of Machine Learning Research* 12.Aug (2011): 2493-2537.

Joint Tagging, Chunking and NER

- Multitasking between Tagging, Chunking and NER
  - Share lookup table
  - Share first linear layers



Collobert, Ronan, et al. "Natural language processing (almost) from scratch." *Journal of Machine Learning Research* 12.Aug (2011): 2493-2537.

Joint Tagging, Chunking and NER

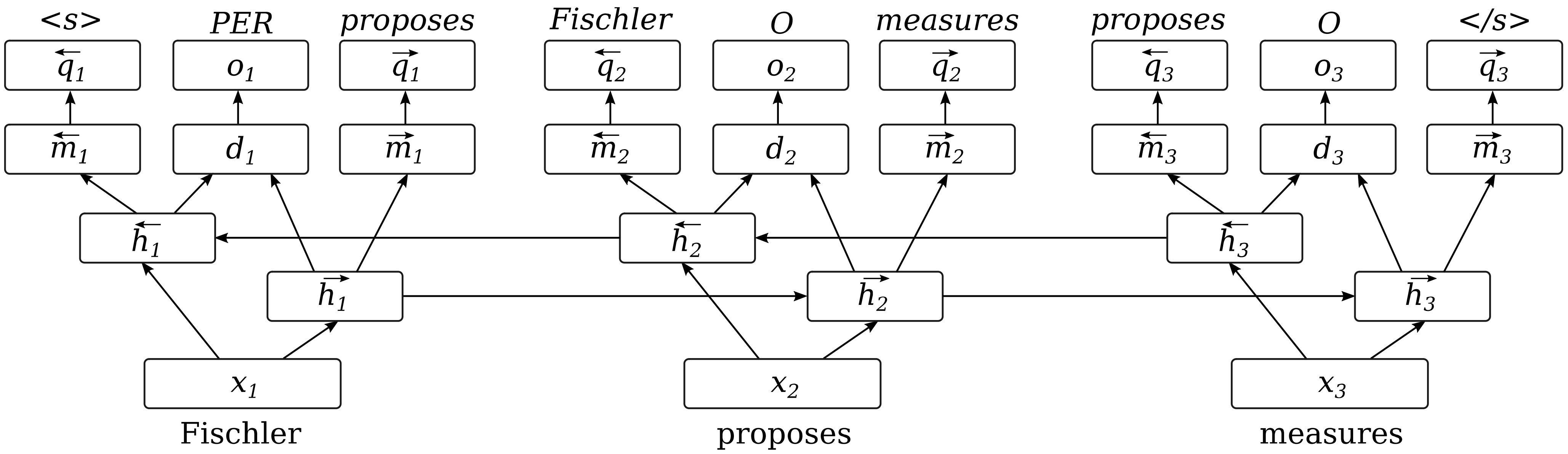
- Results

Approach	POS (PWA)	CHUNK (F1)	NER (F1)
Benchmark Systems	97.24	94.29	89.31
	Window Approach		
NN+SLL+LM2	97.20	93.63	88.67
NN+SLL+LM2+MTL	97.22	94.10	88.62

Collobert, Ronan, et al. "Natural language processing (almost) from scratch." *Journal of Machine Learning Research* 12.Aug (2011): 2493-2537.

NER and Language Modelling

- Model



Rei, Marek. “Semi-supervised Multitask Learning for Sequence Labeling.”, In proceedings of ACL (2017).

NER and Language Modelling

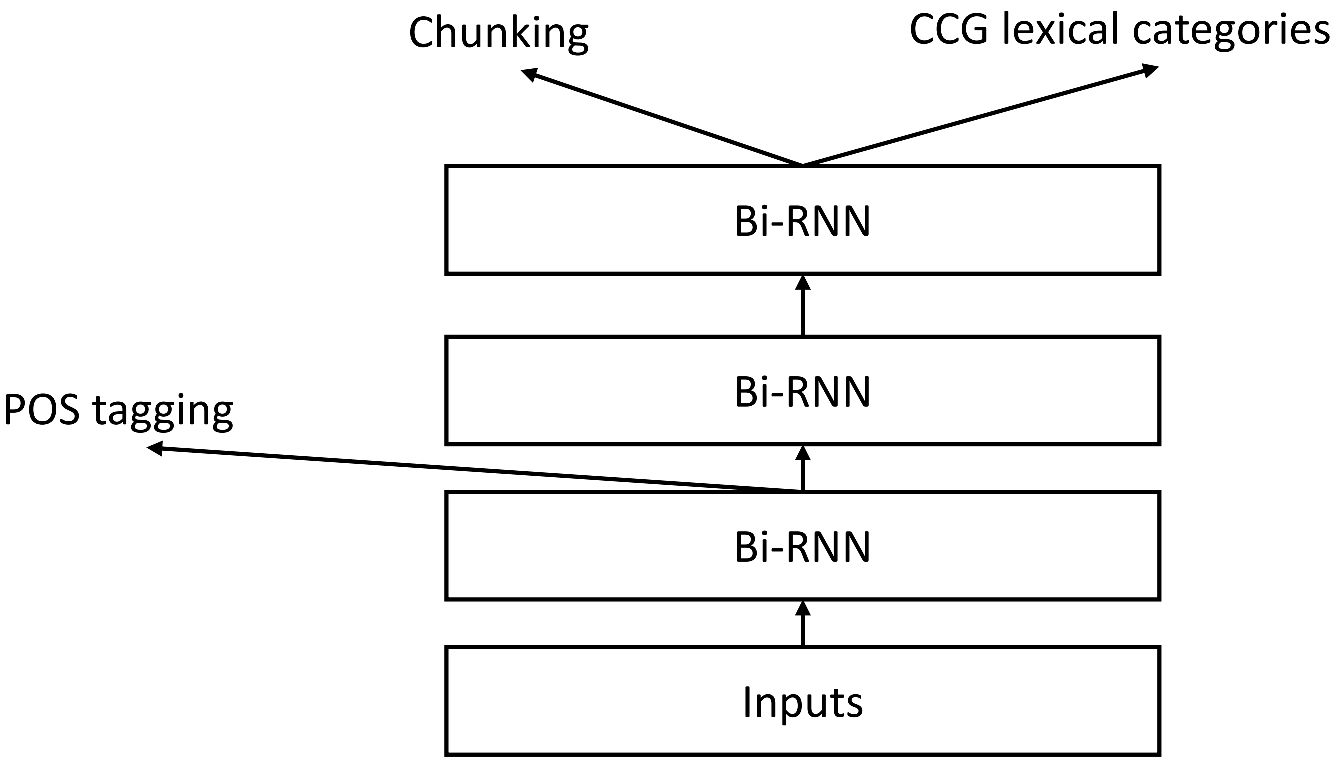
- Results

	CoNLL-00		CoNLL-03		CHEMDNER		JNLPBA	
	DEV	TEST	DEV	TEST	DEV	TEST	DEV	TEST
Baseline	92.92	92.67	90.85	85.63	83.63	84.51	77.13	72.79
+ dropout	93.40	93.15	91.14	86.00	84.78	85.67	77.61	73.16
+ LMcost	**94.22**	**93.88**	**91.48**	**86.26**	**85.45**	**86.27**	**78.51**	**73.83**

Rei, Marek. "Semi-supervised Multitask Learning for Sequence Labeling.", In proceedings of ACL (2017).

Joint POS tagging/Chunking and CCG Super Tagging

- Model



Søgaard, Anders, and Yoav Goldberg. "Deep multi-task learning with low level tasks supervised at lower layers." Proceedings *of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vol. 2. 2016.

Joint POS tagging/Chunking and CCG Super Tagging

- Results

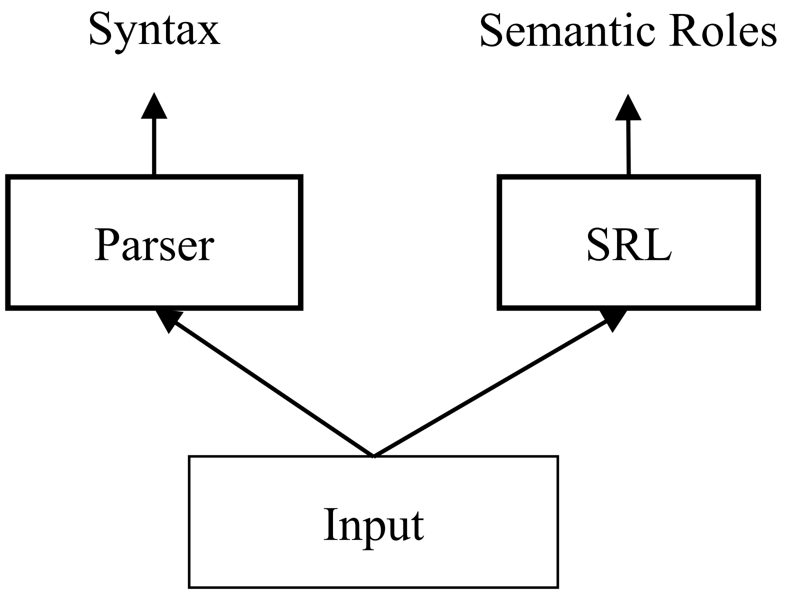
	POS	CHUNKS	CCG
BI-LSTM	-	95.28	91.04
	3	95.30	92.94
	1	**95.56**	**93.26**

- Additional tasks such as NER do not benefit from multi-task learning

Søgaard, Anders, and Yoav Goldberg. "Deep multi-task learning with low level tasks supervised at lower layers." Proceedings *of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vol. 2. 2016.

Joint Parsing and SRL

- Share only the embedding layer



Peng Shi, Zhiyang Teng and Yue Zhang. *Exploiting Mutual Benefits between Syntax and Semantic Roles using Neural Network*. In Proceeddings of EMNLP 2016.

Joint Parsing and SRL

- Results on CONLL

Model	$\mathbf{F_1}$	**UAS**	**LAS**
Bi-LSTM	72.71	-	-
S-LSTM	-	84.33	82.10
DEP→SRL(*lab*/*lstm*)	73.00/**74.18**	84.33	82.10
SRL→DEP	72.71	84.75	82.62
Joint	73.84	**85.15**	**82.91**

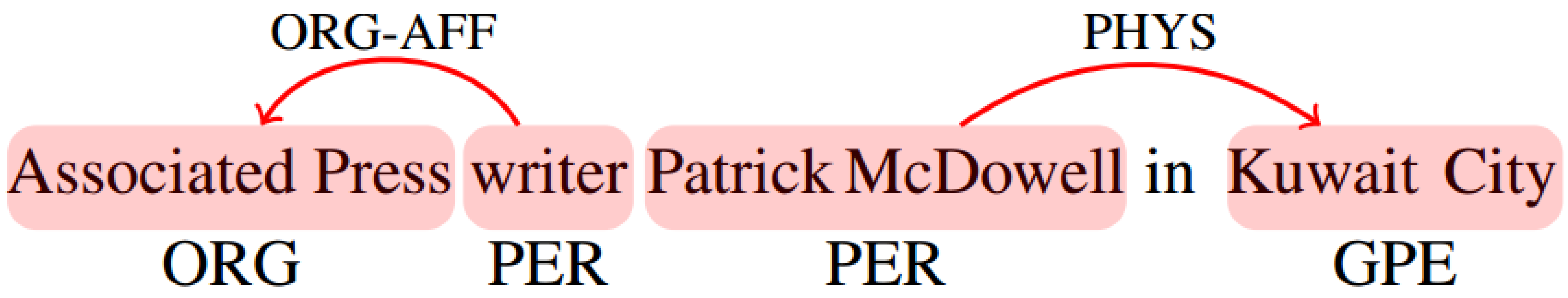
- Sharing more layers have mixed results

Peng Shi, Zhiyang Teng and Yue Zhang. *Exploiting Mutual Benefits between Syntax and Semantic Roles using Neural Network*. In Proceeddings of EMNLP 2016.

Peng Shi and Yue Zhang, Joint Bi-Affine Parsing and Semantic Role Labeling, IALP 2017, Best Paper

Joint Entity and Relation Extraction

- Task



Miwa, Makoto, and Mohit Bansal. "End-to-end relation extraction using lstms on sequences and tree structures." *In proceedings of ACL* (2016).

Joint Entity and Relation Extraction

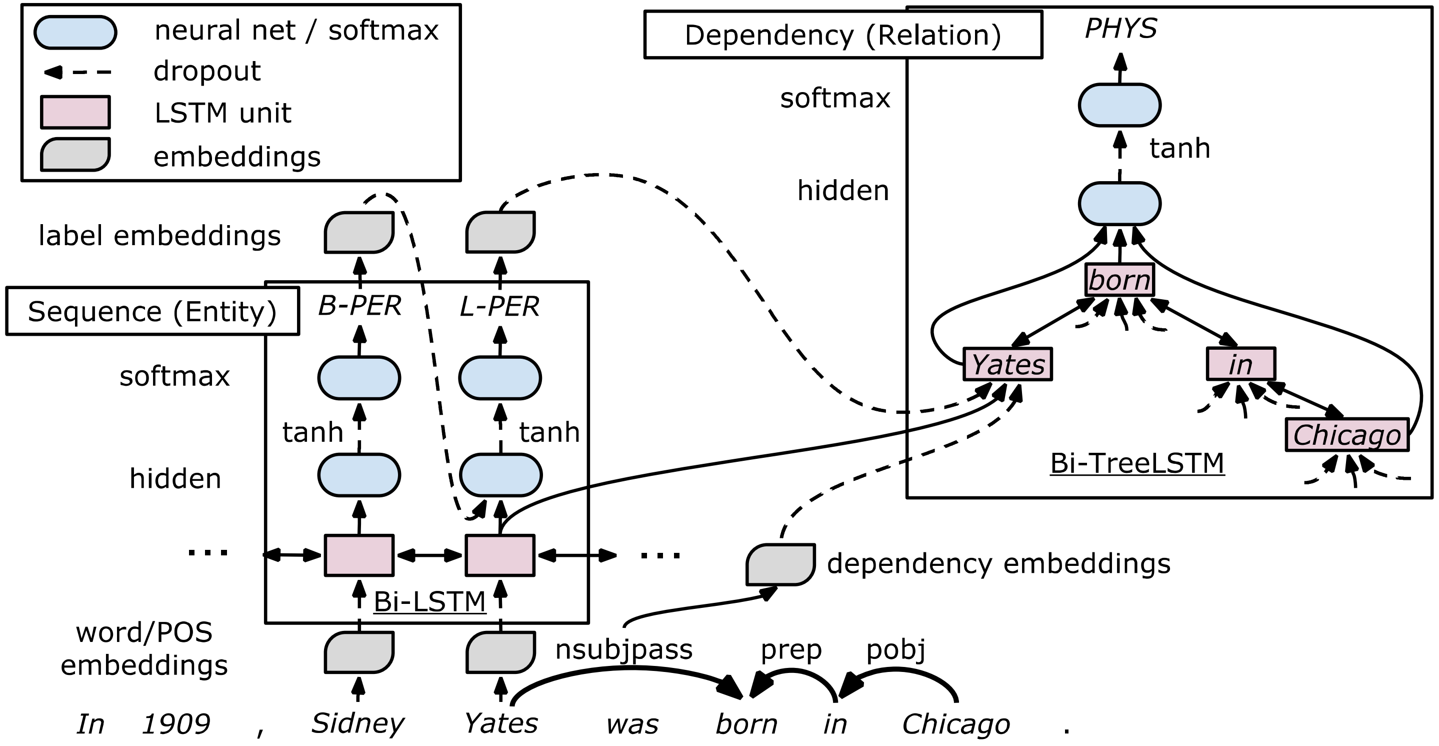
- Table-Filling

	Associated	Press	writer	Patrick	McDowell	in	Kuwait	City
Associated	1 B-ORG	9 ⊥	16 ⊥	22 ⊥	27 ⊥	31 ⊥	34 ⊥	36 ⊥
Press		2 L-ORG	10 $\overleftarrow{\text{ORG-AFF}}$	17 ⊥	23 ⊥	28 ⊥	32 ⊥	35 ⊥
writer			3 U-PER	11 ⊥	18 ⊥	24 ⊥	29 ⊥	33 ⊥
Patrick				4 B-PER	12 ⊥	19 ⊥	25 ⊥	30 ⊥
McDowell					5 L-PER	13 ⊥	20 ⊥	26 $\overrightarrow{\text{PHYS}}$
in						6 O	14 ⊥	21 ⊥
Kuwait							7 B-GPE	15 ⊥
City								8 L-GPE

Miwa, Makoto, and Mohit Bansal. “End-to-end relation extraction using lstms on sequences and tree structures.” *In proceedings of ACL* (2016).

Joint Entity and Relation Extraction

- Share RNN hidden layers



Miwa, Makoto, and Mohit Bansal. "End-to-end relation extraction using lstms on sequences and tree structures." *In proceedings of ACL* (2016).

Joint Entity and Relation Extraction

- Results on ACE

Settings	Macro-F1
No External Knowledge Resources	
Our Model (SPTree)	**0.844**
dos Santos et al. (2015)	0.841
Xu et al. (2015a)	0.840
+WordNet	
Our Model (SPTree + WordNet)	0.855
Xu et al. (2015a)	**0.856**
Xu et al. (2015b)	0.837

Miwa, Makoto, and Mohit Bansal. "End-to-end relation extraction using lstms on sequences and tree structures." *In proceedings of ACL* (2016).

Joint Entity and Relation Extraction

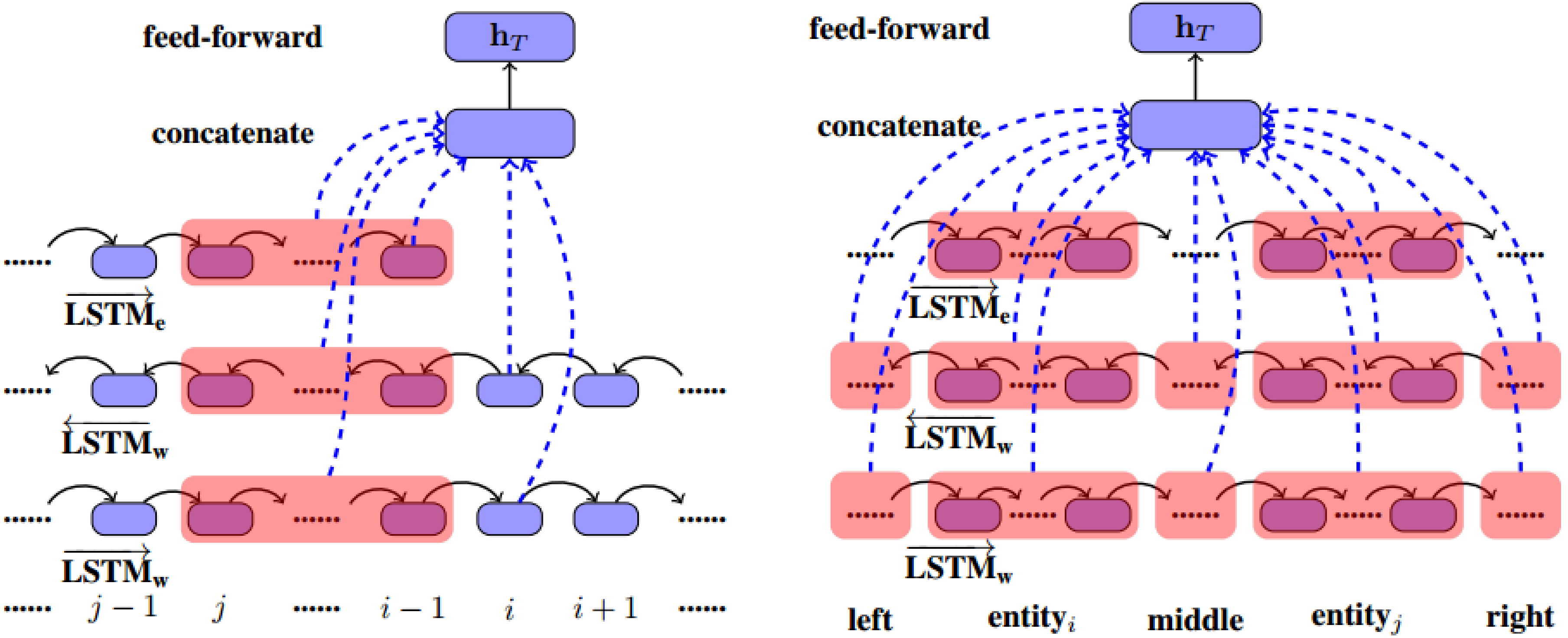
- Beam Search with Global Learning

- Novel Syntactic Features
  - Without any background on syntactic grammars

Zhang, Meishan, et al. "End-to-End Neural Relation Extraction with Global Optimization." *EMNLP*, 2017.

Joint Entity and Relation Extraction

- Share RNN Encoding Layers



6 features

12 features

Zhang, Meishan, et al. "End-to-End Neural Relation Extraction with Global Optimization." *EMNLP*, 2017.

Joint Entity and Relation Extraction

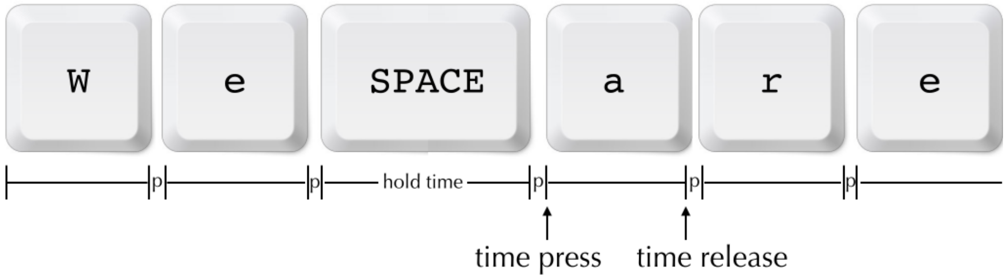
- Results on ACE05

Model	Beam	Relation F1
Local	1	50.9
Local(+SS)	1	51.2
Global	1	51.4
	3	51.8
	5	**52.6**

Zhang, Meishan, et al. "End-to-End Neural Relation Extraction with Global Optimization." *EMNLP*, 2017.

Keystroke and Shallow Syntactic Parsing

- Keystroke Logging



Token:	[Coefficient	of	determination]	[is	a]	[measure	used	in]	[statisitcal	model]	[analysis]
Pause (ms):	0	96	496	30769	96	2144	96	80	2975	240	680

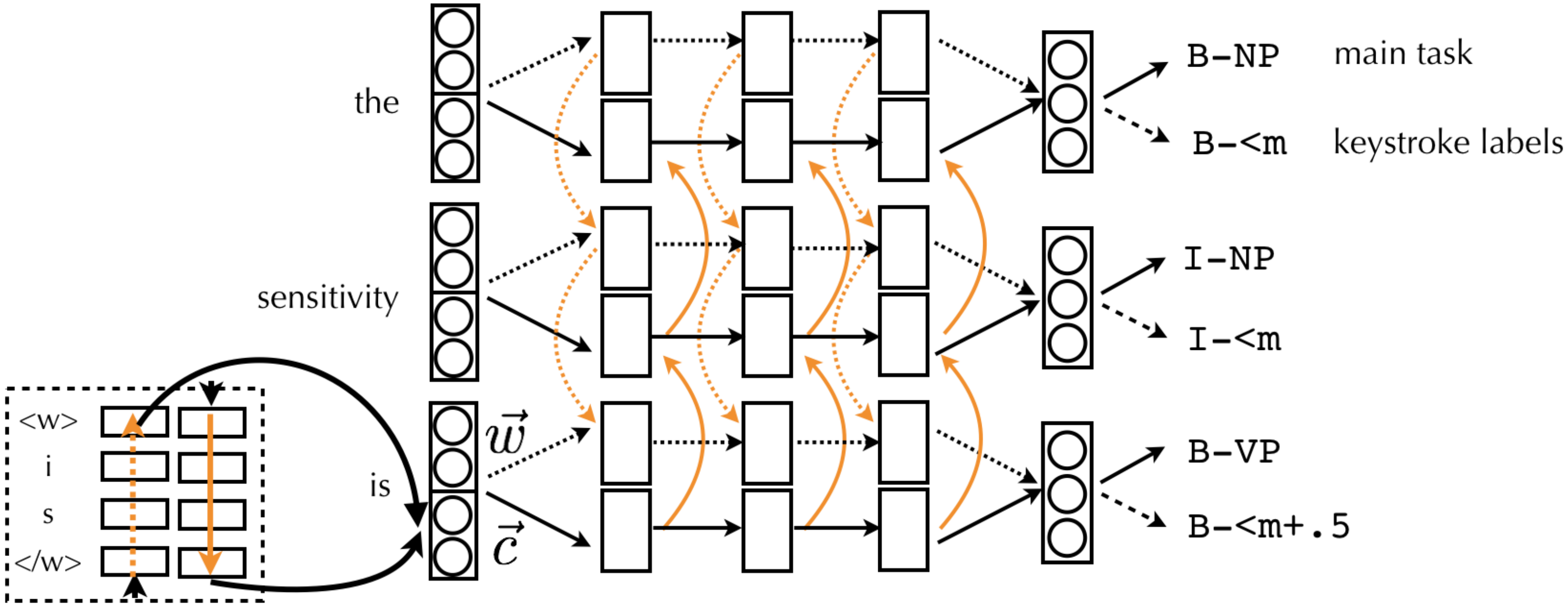
B-<m	B-<m+1	B-<m	I-<m	B-<m+.5	I-<m+.5	B->m+1
the	closer	the	number	is	to	1

Aggregate statistics

Barbara Plank. Keystroke dynamics as signal for shallow syntactic parsing. The 26th International Conference on Computational Linguistics (COLING). Osaka, Japan.

Keystroke and Shallow Syntactic Parsing

- Model



Barbara Plank. Keystroke dynamics as signal for shallow syntactic parsing. The 26th International Conference on Computational Linguistics (COLING). Osaka, Japan.

Keystroke and Shallow Syntactic Parsing

- Results

sentences	TRAIN	DEV	TEST
CONLL 2000	8936	–	2012
FOSTER	–	269	250
RITTER	–	–	2364
CCG	39604	1913	2407

Chunking and CCG data

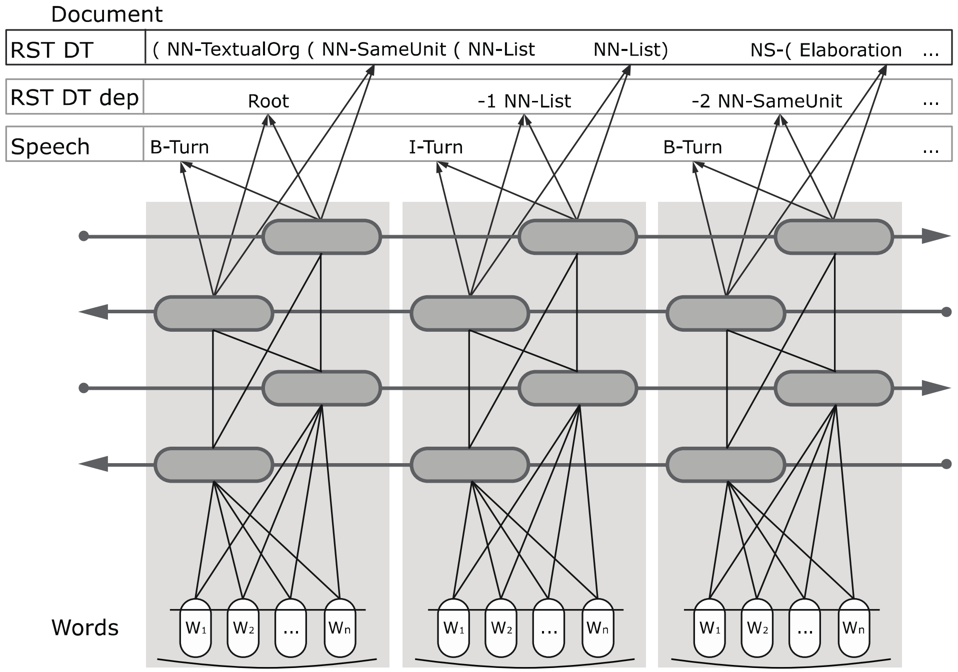
	FOSTER.DEV	FOSTER.TEST	RITTER	CCG
Baseline	73.93	73.61	66.65	92.41
+PAUSE	**74.63**†	**74.32**†	**66.91**†	**92.62**†
p-values	<0.01	<0.01	<0.01	<0.048

Barbara Plank. Keystroke dynamics as signal for shallow syntactic parsing. The 26th International Conference on Computational Linguistics (COLING). Osaka, Japan.

RST Discourse Parser

- Many tasks

	Task	# Doc	# Labels
Main task	Constituent	322	1955
Other views	Nuclearity	322	284
	Relation	322	1159
	Dependency	322	708
	Fine grained	322	2,700
Other tasks	Aspect	208	4
	Factuality	208	7
	Modality	208	10
	Polarity	208	3
	Tense	208	7
	Coreference	2,361	4
	PDTB	2,065	35
	Speech	446	2



Braud, Chloé, Barbara Plank, and Anders Søgaard. "Multi-view and multi-task training of RST discourse parsers." Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016.

RST Discourse Parser

- Results on RST Discourse Treebank

System	RSTFin	Fact	Speech	Asp	RSTDep	Nuc+lab	Mod	Pol	PDTB	Coref	Ten	Span	Nuclearity	Relation
Prior work														
DPLP concat	–	–	–	–	–	–	–	–	–	–	–	82.08	71.13	61.63
DPLP general	–	–	–	–	–	–	–	–	–	–	–	81.60	70.95	61.75
Our work														
Hier-LSTM	–	–	–	–	–	–	–	–	–	–	–	81.39	64.54	49.15
MTL-Hier-LSTM	✓	–	–	–	–	–	–	–	–	–	–	82.88	67.46	53.25
MTL-Hier-LSTM	–	✓	–	–	–	–	–	–	–	–	–	83.40	67.16	52.10
MTL-Hier-LSTM	–	–	✓	–	–	–	–	–	–	–	–	83.26	67.51	51.75
MTL-Hier-LSTM	–	–	–	✓	–	–	–	–	–	–	–	83.69	66.25	51.25
MTL-Hier-LSTM	–	–	–	–	✓	–	–	–	–	–	–	81.25	65.34	51.24
MTL-Hier-LSTM	–	–	–	–	–	✓	–	–	–	–	–	82.09	65.68	51.12
MTL-Hier-LSTM	–	–	–	–	–	–	✓	–	–	–	–	81.66	65.31	50.58
MTL-Hier-LSTM	–	–	–	–	–	–	–	✓	–	–	–	82.01	65.29	50.11
MTL-Hier-LSTM	–	–	–	–	–	–	–	–	✓	–	–	81.61	63.10	48.89
MTL-Hier-LSTM	–	–	–	–	–	–	–	–	–	✓	–	80.26	63.35	47.70
MTL-Hier-LSTM	–	–	–	–	–	–	–	–	–	–	✓	81.33	62.34	47.57
Best combination	–	–	–	–	✓	✓	✓	–	✓	–	–	83.62	69.77	55.11
Human annotation	–	–	–	–	–	–	–	–	–	–	–	88.70	77.72	65.75

Braud, Chloé, Barbara Plank, and Anders Søgaard. "Multi-view and multi-task training of RST discourse parsers." Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016.

Identifying beneficial task relations

- Not all tasks are mutually beneficial !

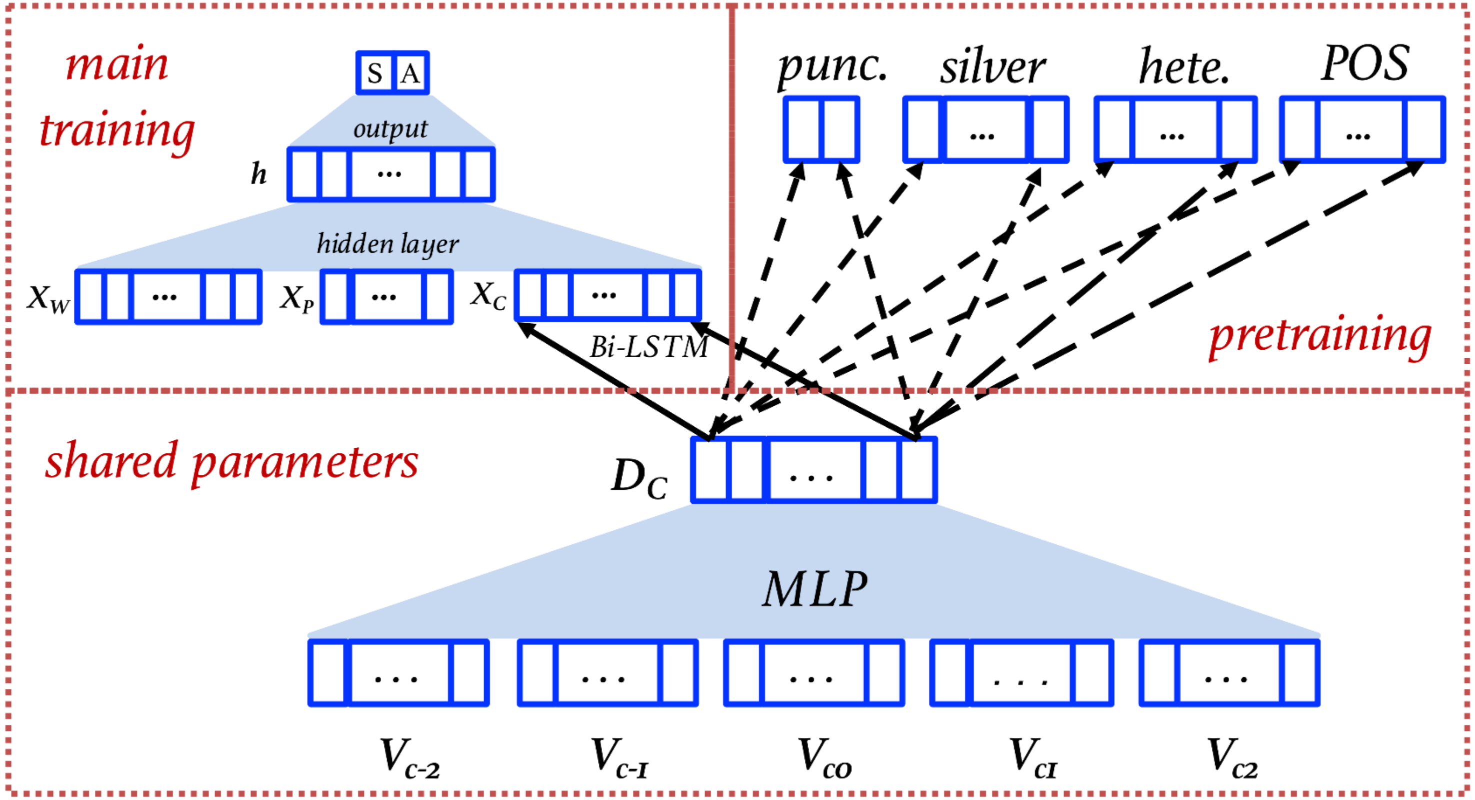
CCG Tagging
Chunking
Sentence Compression
Semantic frames
POS tagging
Hyperlink Prediction
Keyphrase Detection
MWE Detection
Super-sense Tagging

	CCG	CHU	COM	FNT	POS	HYP	KEY	MWE	SEM	STR
CCG		1.4	0.45	0.58	1.8	0.24	0.3	0.45	1.4	0.84
CHU	-0.052		-0.15	-0.12	-0.45	-0.5	-0.22	-0.27	-0.099	-0.32
COM	-5	1.3		1.3	-1.4	-2.4	-4.8	0.82	-3	-0.63
FNT	-5.8	-1	-6.1		-9.4	-5.7	-3.6	-9.4	-3	-0.68
POS	4.9	2.9	1.9	0.9		-0.85	-0.26	1.3	3.4	2.9
HYP	12	4	-11	9.2	22		1.5	-7.7	23	8.1
KEY	5.7	3.2	-1	-0.43	-1.3	-2.6		-4.7	0.59	0.69
MWE	18	20	7.4	5.5	1.6	-3.8	-5.8		16	8.6
SEM	-5	-0.76	-1.2	-0.81	-0.85	-1.3	-0.83	-1.1		-1.7
STR	-1.7	1.5	-0.26	-0.72	0.037	-1.5	-1.4	-1.6	1.7	

Bingel, Joachim, and Anders Søgaard. "Identifying beneficial task relations for multi-task learning in deep neural networks." arXiv preprint arXiv:1702.08303 (2017).
Hector Mart ´ ´ ınez Alonso and Barbara Plank. 2017. Multitask learning for semantic sequence prediction under varying data conditions. In EACL.
Mou, Lili, et al. "How transferable are neural networks in nlp applications?." arXiv preprint arXiv:1603.06111 (2016).

Pretraining: Word Segmentation

- Rich Multi-task pretraining of character window representations



Jie Yang, Yue Zhang, Fei Dong. Neural Word Segmentation with Rich Pretraining. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (2017)

Pretraining: Word Segmentation

- Results

Models	P	R	F
Baseline	95.3	95.5	95.4
Punc. pretrain	96.0	95.6	95.8
Auto-seg pretrain	95.8	95.6	95.7
Multitask pretrain	**96.4**	**96.0**	**96.2**
Sun and Xu (2011) baseline	95.2	94.9	95.1
Sun and Xu (2011) multi-source semi	95.9	95.6	95.7
Zhang et al. (2016b) neural	95.3	94.7	95.0
Zhang et al. (2016b)* hybrid	96.1	95.8	96.0
Chen et al. (2015a) window	95.7	95.8	95.8
Chen et al. (2015b) char LSTM	96.2	95.8	96.0
Zhang et al. (2014) POS and syntax	–	–	95.7
Wang et al. (2011) statistical semi	95.8	95.8	95.8
Zhang and Clark (2011) statistical	95.5	94.8	95.1

Jie Yang, Yue Zhang, Fei Dong. Neural Word Segmentation with Rich Pretraining. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (2017)

Pretraining: Word Segmentation

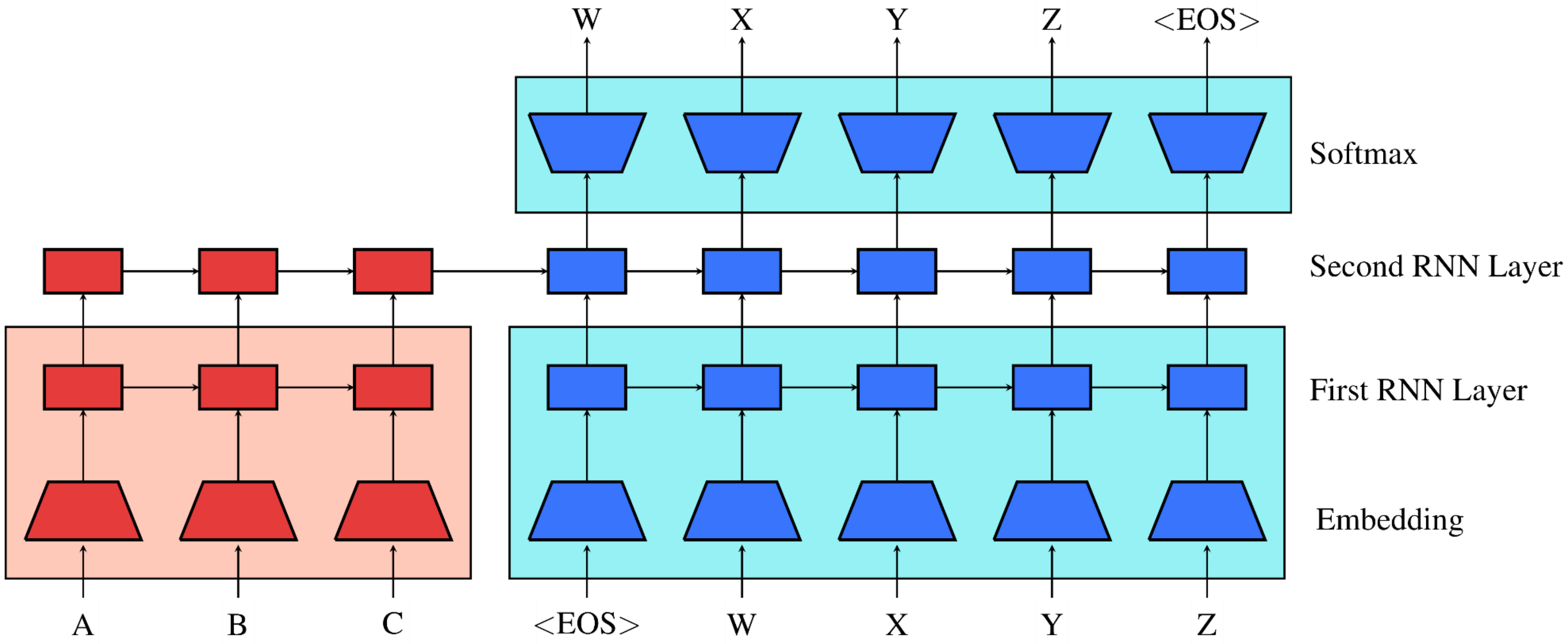
- Results

F1 measure	PKU	MSR	AS	CityU	Weibo
Multitask pretrain	**96.3**	97.5	**95.7**	**96.9**	**95.5**
Cai and Zhao (2016)	95.5	96.5	–	–	–
Zhang et al. (2016b)	95.1	97.0	–	–	–
Zhang et al. (2016b)*	95.7	**97.7**	–	–	–
Pei et al. (2014)	95.2	97.2	–	–	–
Sun et al. (2012)	95.4	97.4	–	–	–
Zhang and Clark (2007)	94.5	97.2	94.6	95.1	–
Zhang et al. (2006)	95.1	97.1	95.1	95.1	–
Sun et al. (2009)	95.2	97.3	–	94.6	–
Sun (2010)	95.2	96.9	95.2	95.6	–
Wang et al. (2014)	95.3	97.4	95.4	94.7	–
Xia et al. (2016)	–	–	–	–	95.4

Jie Yang, Yue Zhang, Fei Dong. Neural Word Segmentation with Rich Pretraining. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (2017)

Pretraining: Language Translation and Language Modelling

- Language Model Pretrain for both the source and target



Ramachandran, Prajit, Peter J. Liu, and Quoc V. Le. “Unsupervised pretraining for sequence to sequence learning.” In Proceeddings of EMNLP, (2016).

Pretraining: Language Translation and Language Modelling

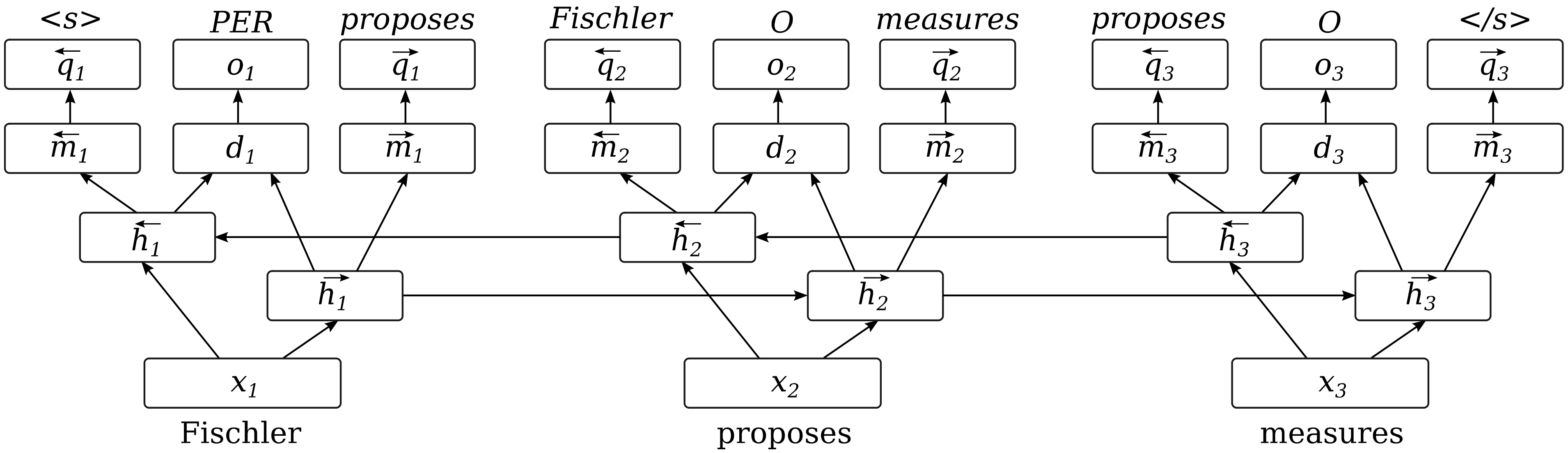
- Results on WMT

System	*ensemble?*	*BLEU* *newstest2014*	*newstest2015*
Phrase Based MT (Williams et al., 2016)	-	21.9	23.7
Supervised NMT (Jean et al., 2015)	single	-	22.4
Edit Distance Transducer NMT (Stahlberg et al., 2016)	single	21.7	24.1
Edit Distance Transducer NMT (Stahlberg et al., 2016)	ensemble 8	22.9	25.7
Backtranslation (Sennrich et al., 2015a)	single	22.7	25.7
Backtranslation (Sennrich et al., 2015a)	ensemble 4	23.8	26.5
Backtranslation (Sennrich et al., 2015a)	ensemble 12	**24.7**	27.6
No pretraining	single	21.3	24.3
Pretrained seq2seq	single	**24.0**	**27.0**
Pretrained seq2seq	ensemble 5	**24.7**	**28.1**

Ramachandran, Prajit, Peter J. Liu, and Quoc V. Le. “Unsupervised pretraining for sequence to sequence learning.” In Proceeddings of EMNLP, (2016).

Language Model Pretraining

- Embeddings from Language Models (ELMo)



Peters, Matthew E., et al. "Deep contextualized word representations." arXiv preprint arXiv:1802.05365 (2018).

Language Model Pretraining

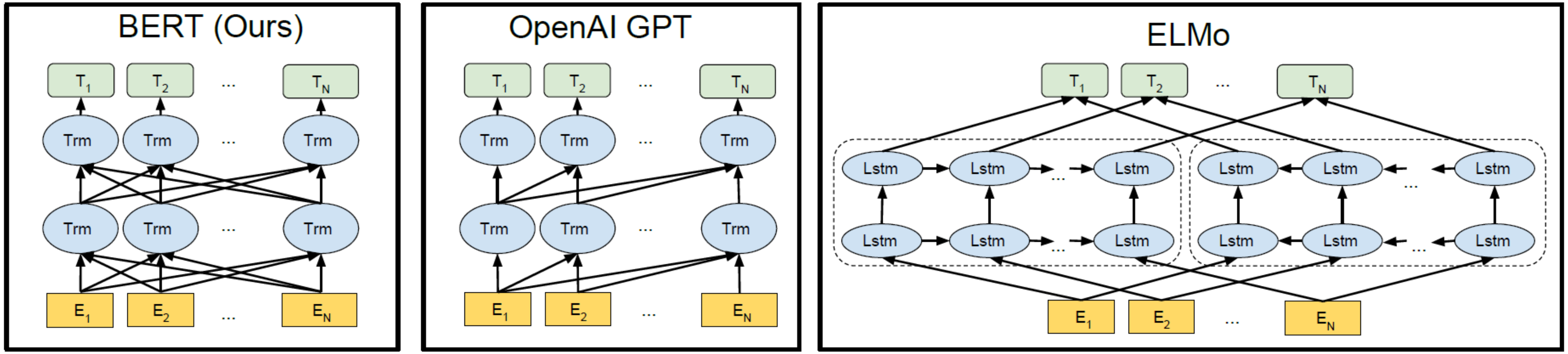
- Results

Task	Previous SOTA		Our baseline	ELMo + baseline	Increase (absolute/relative)
SQuAD	Liu et al. (2017)	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Chen et al. (2017)	88.6	88.0	88.7 ± 0.17	0.7 / 5.8%
SRL	He et al. (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al. (2017)	67.2	67.2	70.4	3.2 / 9.8%
NER	Peters et al. (2017)	91.93 ± 0.19	90.15	92.22 ± 0.10	2.06 / 21%
SST-5	McCann et al. (2017)	53.7	51.4	54.7 ± 0.5	3.3 / 6.8%

Peters, Matthew E., et al. "Deep contextualized word representations." arXiv preprint arXiv:1802.05365 (2018).

Language Model Pretraining

- BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding



Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805 (2018).

Language Model Pretraining

- Results on GLUE

System	MNLI-(m/mm)	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE	**Average**
	392k	363k	108k	67k	8.5k	5.7k	3.5k	2.5k	-
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.9	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
$\text{BERT}_{\text{BASE}}$	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
$\text{BERT}_{\text{LARGE}}$	**86.7/85.9**	**72.1**	**91.1**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**81.9**

Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805 (2018).

Language Model Pretraining

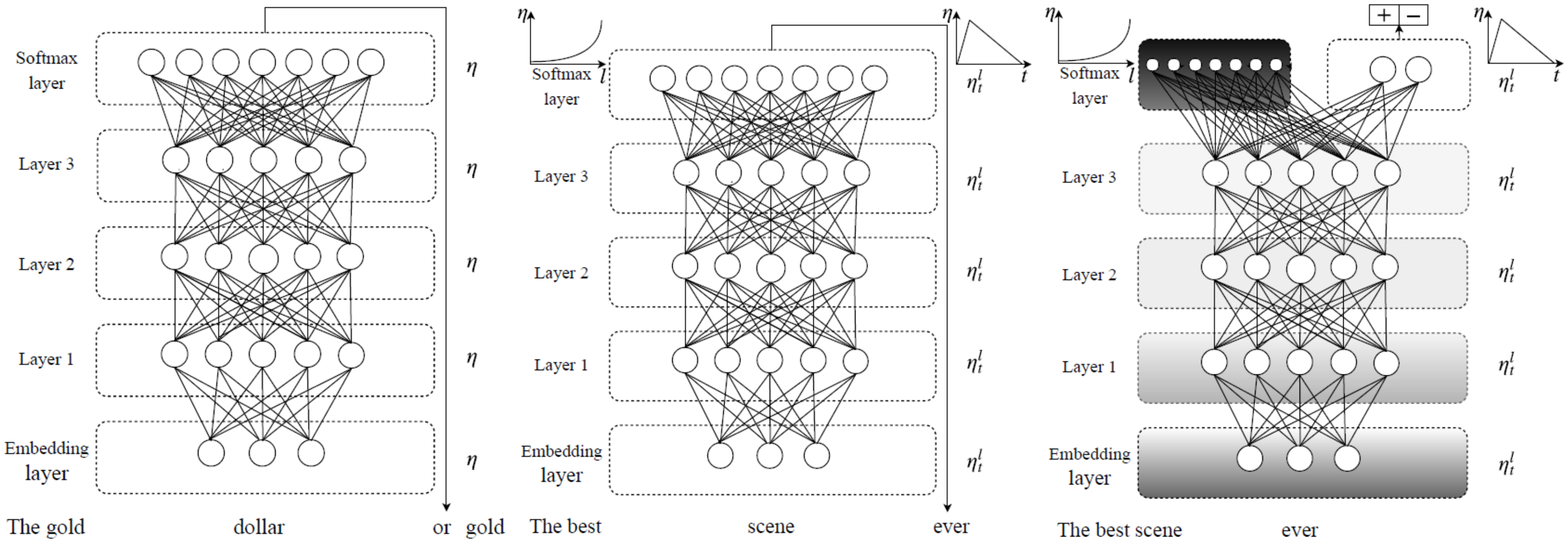
- Results on SQuAD

System	Dev EM	Dev F1	Test EM	Test F1
Leaderboard (Oct 8th, 2018)				
Human	-	-	82.3	91.2
#1 Ensemble - nlnet	-	-	86.0	91.7
#2 Ensemble - QANet	-	-	84.5	90.5
#1 Single - nlnet	-	-	83.5	90.1
#2 Single - QANet	-	-	82.5	89.3
Published				
BiDAF+ELMo (Single)	-	85.8	-	-
R.M. Reader (Single)	78.9	86.3	79.5	86.6
R.M. Reader (Ensemble)	81.2	87.9	82.3	88.5
Ours				
$\text{BERT}_{\text{BASE}}$ (Single)	80.8	88.5	-	-
$\text{BERT}_{\text{LARGE}}$ (Single)	84.1	90.9	-	-
$\text{BERT}_{\text{LARGE}}$ (Ensemble)	85.8	91.8	-	-
$\text{BERT}_{\text{LARGE}}$ (Sgl.+TriviaQA)	**84.2**	**91.1**	**85.1**	**91.8**
$\text{BERT}_{\text{LARGE}}$ (Ens.+TriviaQA)	**86.2**	**92.2**	**87.4**	**93.2**

Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805 (2018).

Language Model Pretraining

- Universal Language Model Fine-tuning for Text Classification



(a) LM pre-training (b) LM fine-tuning (c) Classifier fine-tuning

Howard, Jeremy and Sebastian Ruder. "Universal Language Model Fine-tuning for Text Classification." *ACL*, 2018.

Language Model Pretraining

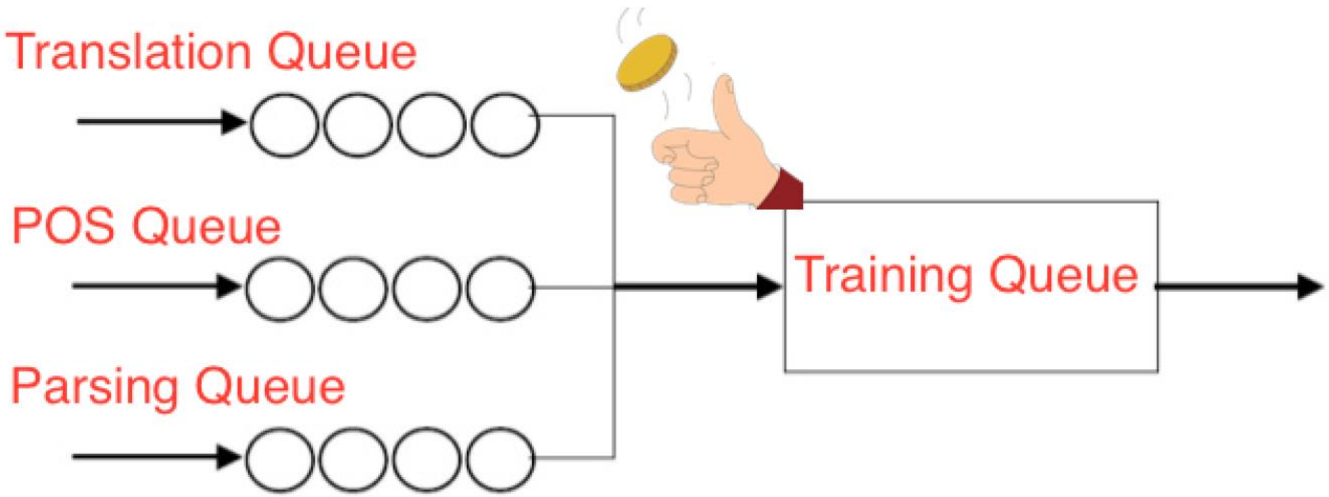
- Results on classification tasks

	AG	DBpedia	Yelp-bi	Yelp-full
Char-level CNN (Zhang et al., 2015)	9.51	1.55	4.88	37.95
CNN (Johnson and Zhang, 2016)	6.57	0.84	2.90	32.39
DPCNN (Johnson and Zhang, 2017)	6.87	0.88	2.64	30.58
ULMFiT (ours)	**5.01**	**0.80**	**2.16**	**29.98**

Howard, Jeremy and Sebastian Ruder. "Universal Language Model Fine-tuning for Text Classification." *ACL*, 2018.

Correlation between multi-task learning and pretraining

- Tasks



Kiperwasser, Eliyahu and Miguel Ballesteros. "Scheduled Multi-Task Learning: From Syntax to Translation." *TACL,* 2018.

Correlation between multi-task learning and pretraining

- Results on IWSLT English German

Scheduler	Tasks	BLEU	POS	UAS	LAS
No MTL	NMT	27.70	–	–	–
	POS	–	95.41	–	–
	Parsing (Unlabeled)	–	–	80.28	–
Constant Scheduler	NMT + POS	**30.4**	93.51	–	–
	NMT + Parsing	28.73	–	79.78	74.25
	NMT + POS + Parsing	29.08	94.80	79.38	74.13
Exponent Scheduler	NMT + POS	30.15	89.05	–	–
	NMT + Parsing	29.37	–	67.60	60.71
	NMT + POS + Parsing	29.55	91.48	72.85	66.44
Sigmoid Scheduler	NMT + POS	30.2	90.74	–	–
	NMT + Parsing	28.78	–	69.26	62.43
	NMT + POS + Parsing	28.93	89.11	65.92	58.46

Kiperwasser, Eliyahu and Miguel Ballesteros. "Scheduled Multi-Task Learning: From Syntax to Translation." *TACL,* 2018.

Joint Entity and Sentiment Extraction

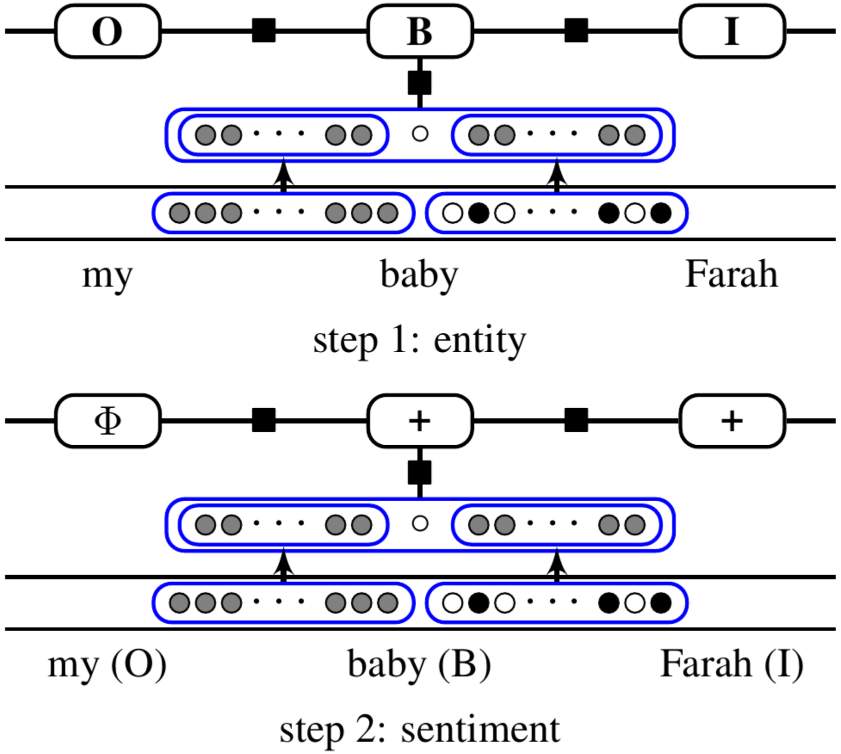
- Multi-task with CRF

So excited to meet my [**baby Farah**]$_+$!!!
[**Baseball Warehouse**]$_+$: easy to understand information.
The [**#Afghan #Parlaiment Speaker**]$_-$ should Resign .
Saw [**Erykah Badu**]$_-$ last night , vile venue unfortunately .
[**AW service**]$_0$ will be back at work .

Zhang, Meishan, Yue Zhang, and Duy-Tin Vo. "Neural Networks for Open Domain Targeted Sentiment." *EMNLP*. 2015.

Joint Entity and Sentiment Extraction

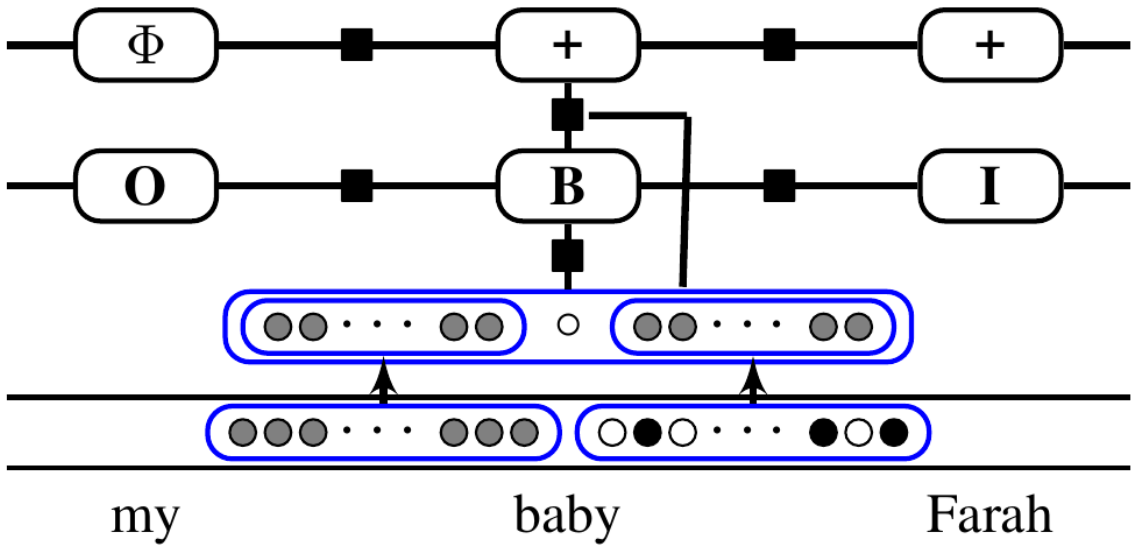
- Pipeline



Zhang, Meishan, Yue Zhang, and Duy-Tin Vo. "Neural Networks for Open Domain Targeted Sentiment." *EMNLP*. 2015.

Joint Entity and Sentiment Extraction

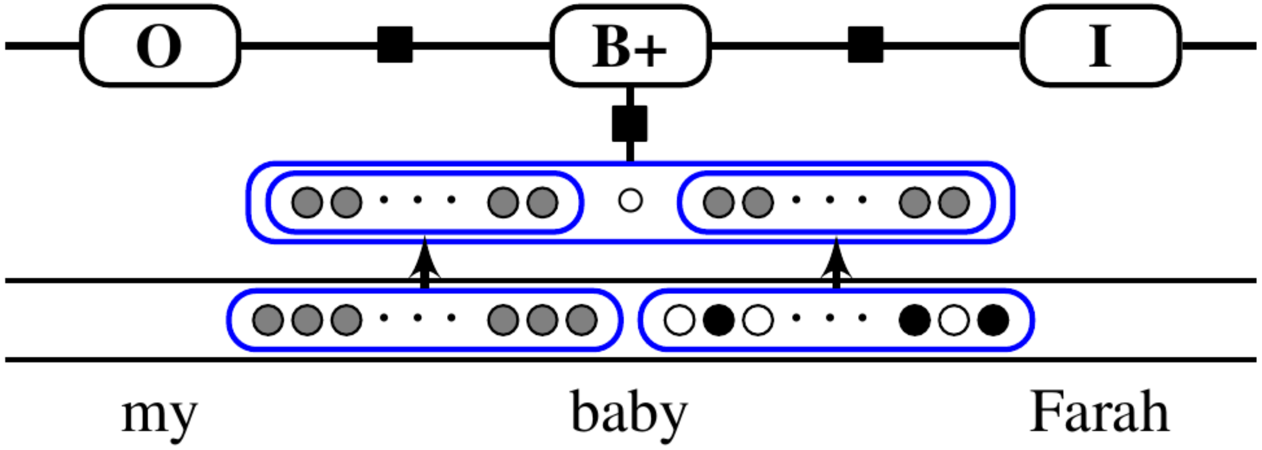
- Joint



Zhang, Meishan, Yue Zhang, and Duy-Tin Vo. "Neural Networks for Open Domain Targeted Sentiment." *EMNLP*. 2015.

Joint Entity and Sentiment Extraction

- Collapsed



Zhang, Meishan, Yue Zhang, and Duy-Tin Vo. "Neural Networks for Open Domain Targeted Sentiment." *EMNLP*. 2015.

Joint Entity and Sentiment Extraction

- Results

Model	English						Spanish					
	Entity			SA			Entity			SA		
	P	R	F	P	R	F	P	R	F	P	R	F
Pipeline												
discrete	59.37	34.83	43.84	42.97	25.21	31.73	**70.77**	47.75	57.00	**46.55**	31.38	37.47
neural	53.64	44.87	48.67	37.53	31.38	34.04	65.59	47.82	55.27	41.50	30.27	34.98
integrated	**60.69**	**51.63**	**55.67**	**43.71**	**37.12**	**40.06**	70.23	**62.00**	**65.76**	45.99	**40.57**	**43.04**
Joint												
discrete	59.55	34.06	43.30	43.09	24.67	31.35	71.08	47.56	56.96	46.36	31.02	37.15
neural	54.45	42.12	47.17	37.55	28.95	32.45	65.05	47.79	55.07	40.28	29.58	34.09
integrated	**61.47**	**49.28**	**54.59**	**44.62**	**35.84**	**39.67**	**71.32**	**61.11**	**65.74**	**46.67**	**39.99**	**43.02**
Collapsed												
discrete	**64.16**	26.03	36.95	**48.35**	19.64	27.86	73.18	35.11	47.42	**49.85**	23.91	32.30
neural	58.53	37.25	45.30	43.12	27.44	33.36	67.43	43.2	52.64	42.61	27.27	33.25
integrated	63.55	**44.98**	**52.58**	46.32	**32.84**	**38.36**	**73.51**	**53.3**	**61.71**	47.69	**34.53**	**40.00**

Zhang, Meishan, Yue Zhang, and Duy-Tin Vo. "Neural Networks for Open Domain Targeted Sentiment." *EMNLP*. 2015.

Neural Graph-based Models (Multi-task Learning)

- Cross Task
- Cross Lingual
- Cross Domain
- Cross Standard

Neural Graph-based Models (Multi-task Learning)

- Cross Task
- Cross Lingual
  - Standard
  - Regularization
  - Stacking
  - Pretraing
  - Adversarial Training
- Cross Domain
- Cross Standard

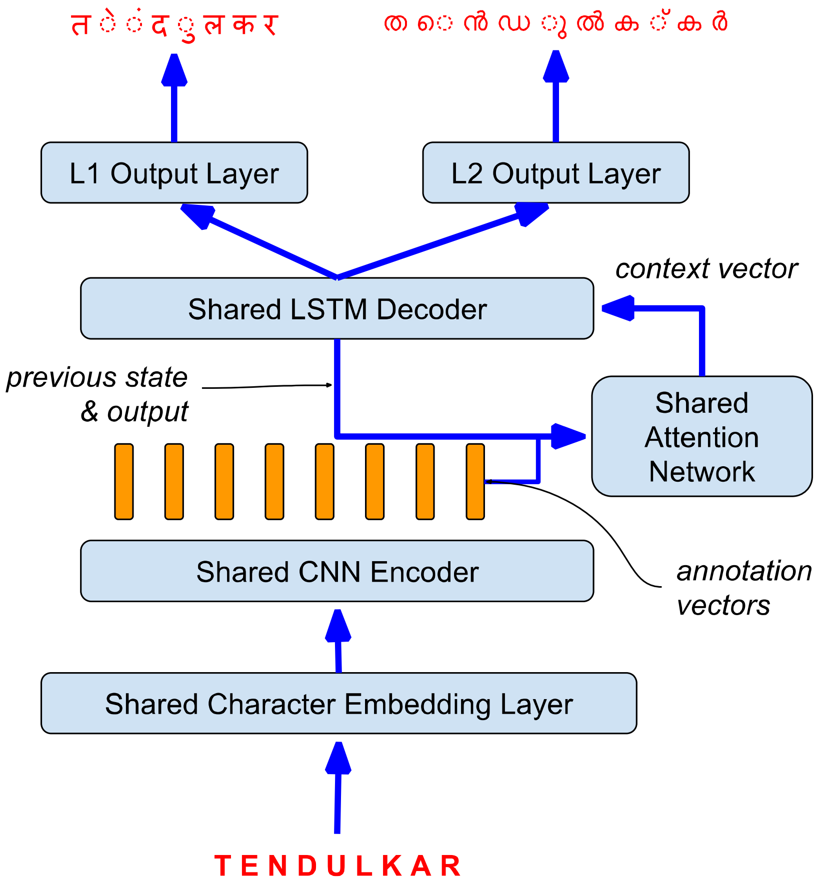
Standard: Multi-lingual Neural Transliteration

- Orthographically similar languages
  - (i) highly overlapping phoneme sets.
  - (ii) mutually compatible orthographic systems.
  - (iii) similar grapheme to phoneme mappings.

Kunchukuttan, Anoop, et al. "Leveraging Orthographic Similarity for Multilingual Neural Transliteration." Transactions of the Association for Computational Linguistics 6 (2018): 303-316.

Standard: Multi-lingual Neural Transliteration

- Standard multi-task



Kunchukuttan, Anoop, et al. "Leveraging Orthographic Similarity for Multilingual Neural Transliteration." Transactions of the Association for Computational Linguistics 6 (2018): 303-316.

Standard: Multi-lingual Neural Transliteration

- Results on NEWS 2015

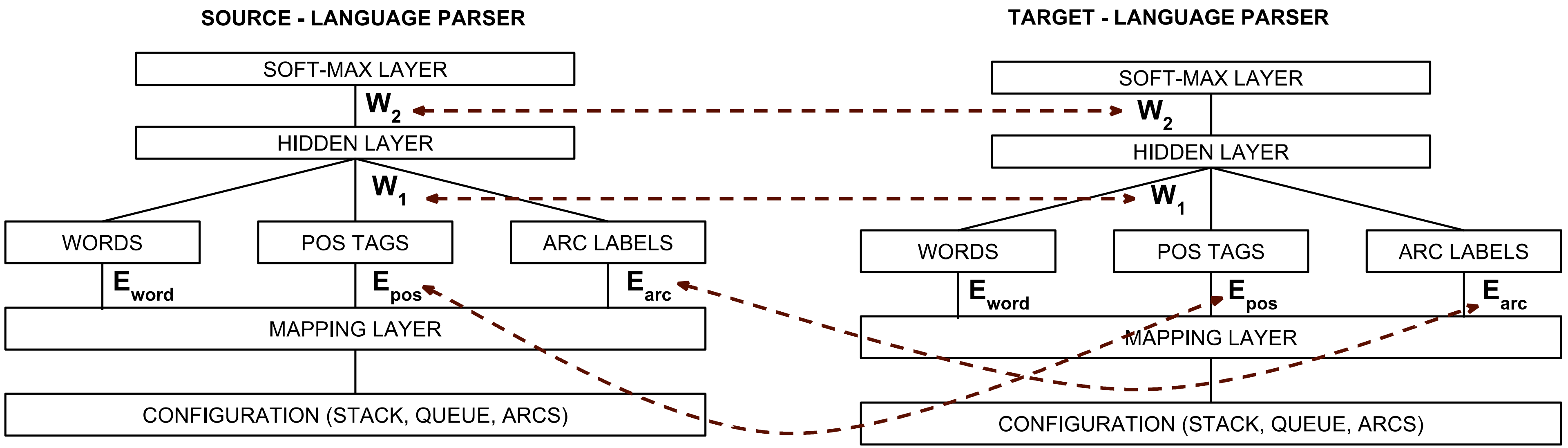
Pair	P	B	M	Pair	P	B	M
Similar Source and Target Languages							
Indic-Indic (45.5%)							
bn-hi	**29.74**	19.08	27.69	kn-bn	28.59	24.04	**37.47**
bn-kn	17.62	18.14	**27.74**	kn-ta	34.89	30.85	**38.30**
hi-bn	29.92	25.46	**39.15**	ta-hi	**29.07**	19.24	28.97
hi-ta	25.15	28.62	**38.70**	ta-kn	26.99	19.86	**29.06**
Similar Source Languages							
Slavic-Arabic (55.8%)				*Indic-English (24.2%)*			
cs-ar	38.91	37.10	**59.17**	bn-en	**55.23**	48.93	54.01
pl-ar	34.70	34.80	**44.83**	hi-en	49.19	38.26	**51.11**
sk-ar	43.26	37.49	**62.21**	kn-en	42.79	33.77	**47.70**
sl-ar	41.90	36.74	**62.04**	ta-en	**33.93**	23.22	25.93
Similar Target Languages							
Arabic-Slavic (176.8%)				*English-Indic (1.1%)*			
ar-cs	15.41	12.08	**36.76**	en-bn	42.90	41.70	**46.10**
ar-pl	13.68	12.26	**24.21**	en-hi	60.50	**64.10**	60.70
ar-sk	15.24	13.82	**38.72**	en-kn	48.70	52.00	**53.90**
ar-sl	18.31	13.63	**44.35**	en-ta	52.90	**57.80**	55.30

Comparison of bilingual (B) and multilingual (M) neural models as well as bilingual PBSMT (P) models (top-1 accuracy %). Figure in brackets for each dataset shows average increase in transliteration accuracy for multilingual neural model over bilingual neural model. Best accuracies for each language pair in **bold**.

Kunchukuttan, Anoop, et al. "Leveraging Orthographic Similarity for Multilingual Neural Transliteration." Transactions of the Association for Computational Linguistics 6 (2018): 303-316.

Regularization: Low resource dependency parsing

- English to Low Resource
  - Transferred Parameters $E_{word}^{en}, E_{pos}^{en}, E_{arc}^{en}, W_1^{en}, W_2^{en}$



Duong, Long, et al. "Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Vol. 2. 2015.

Regularization: Low resource dependency parsing

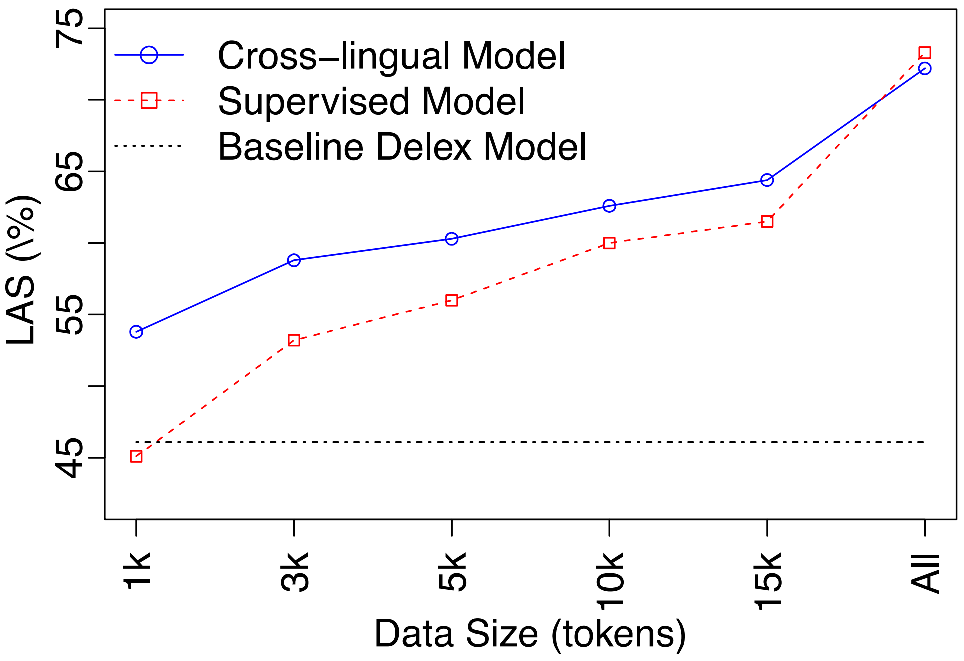
“ To allow parameter sharing between languages we could jointly train the parser on the source and target language simultaneously. However, we leave this for **future work**. First we train a lexicalized neural network parser on the source resource-rich language (English), as described in Section 2. "

$$\mathcal{L} = \sum_{i=1}^{N} \log P(y^{(i)}|x^{(i)}) - \frac{\lambda_1}{2}\Big[\|W_1^{pos} - W_1^{en:pos}\|_F^2 + \|W_1^{arc} - W_1^{en:arc}\|_F^2 + \|W_2 - W_2^{en}\|_F^2\Big] - \frac{\lambda_2}{2}\Big[\|E_{pos} - E_{pos}^{en}\|_F^2 + \|E_{arc} - E_{arc}^{en}\|_F^2\Big]$$

Duong, Long, et al. "Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Vol. 2. 2015.

Regularization: Low resource dependency parsing

- Results



Save human label effort

Duong, Long, et al. "Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Vol. 2. 2015.

Language Embedding: Multi-lingual parser

- ‘**Future Work**’ mentioned in the previous work.
- Seven languages jointly trained
- Words: Cross-lingual embeddings and cross-lingual word cluster
- Languages: Language embeddings!

Ammar, Waleed, et al. "Many languages, one parser." arXiv preprint arXiv:1602.01595 (2016).

Language Embedding: Multi-lingual parser

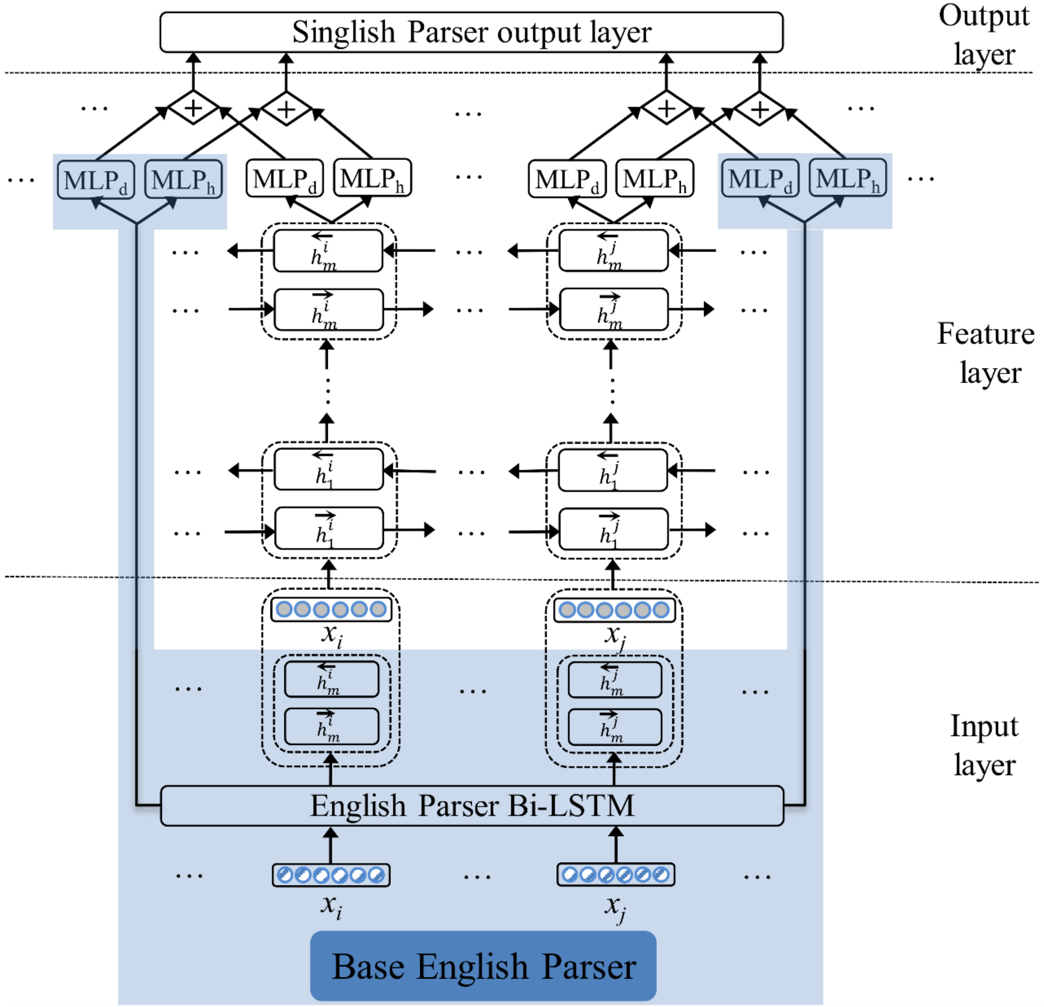
- Results on UD Treebank

LAS	target language							average
	de	en	es	fr	it	pt	sv	
monolingual	**79.3**	**85.9**	83.7	81.7	88.7	85.7	83.5	84.0
MALOPA	70.4	69.3	72.4	71.1	78.0	74.1	65.4	71.5
+lexical	76.7	82.0	82.7	81.2	87.6	82.1	81.2	81.9
+language ID	78.6	84.2	83.4	**82.4**	**89.1**	84.2	82.6	83.5
+fine-grained POS	78.9	85.4	**84.3**	**82.4**	89.0	**86.2**	**84.5**	**84.3**

Ammar, Waleed, et al. "Many languages, one parser." arXiv preprint arXiv:1602.01595 (2016).

Stacking: Singlish Parsing

- Variation on Parameter Sharing



Hongmin Wang, Yue Zhang, GuangYong Leonard Chan, Jie Yang, Hai Leong Chieu. Universal Dependencies Parsing for Colloquial Singaporean English. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Vancouver, Canada, July.

Stacking: Singlish Parsing

- Results

System	Accuracy
ENG-on-SIN	81.39%
Base-ICE-SIN	78.35%
Stack-ICE-SIN	89.50%

POS tagging

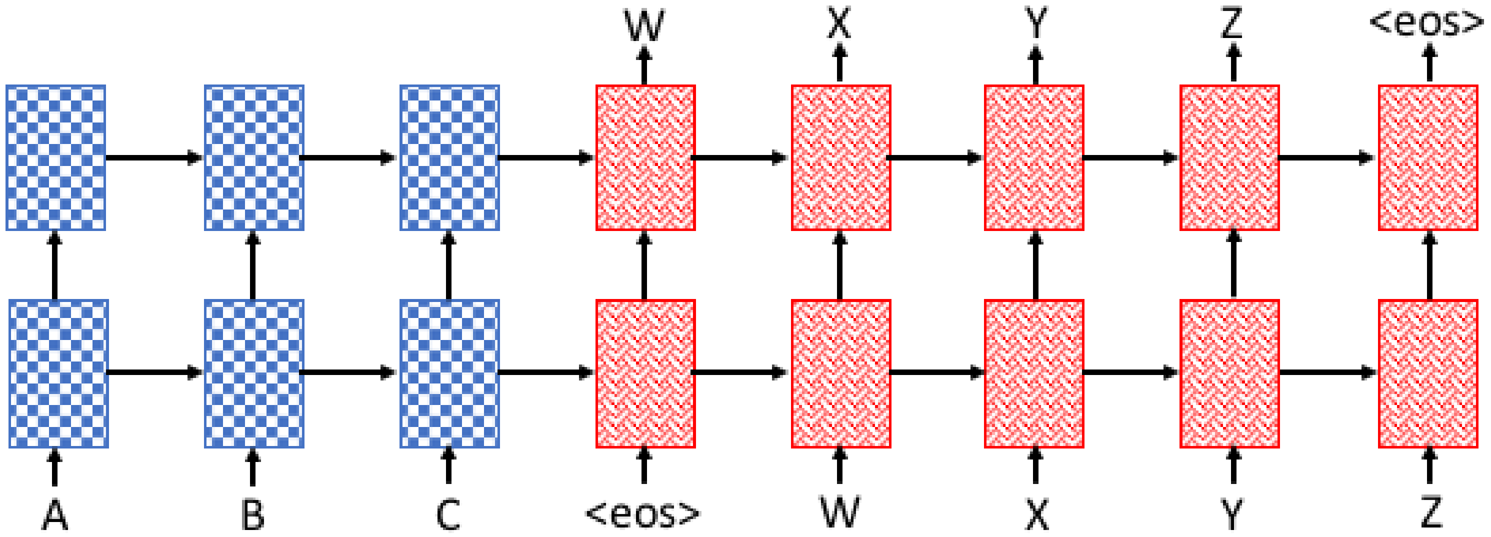
Trained on	System	UAS	LAS
English	ENG-on-SIN	75.89	65.62
Singlish	Baseline	75.98	66.55
	Base-Giga100M	77.67	67.23
	Base-GloVe6B	78.18	68.51
	Base-ICE-SIN	79.29	69.27
Both	ENG-plus-SIN	82.43	75.64
	Stack-ICE-SIN	**84.47**	**77.76**

Dependency Parsing

Hongmin Wang, Yue Zhang, GuangYong Leonard Chan, Jie Yang, Hai Leong Chieu. Universal Dependencies Parsing for Colloquial Singaporean English. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Vancouver, Canada, July.

Pretraining: Low resource neural machine translation

- Variation on model structure: Rich Resource(EN→ FR) pretraining, low resource (EN → UZ) fine-tuning



Zoph, Barret, et al. "Transfer learning for low-resource neural machine translation." In Proceeddings of EMNLP, (2016).

Pretraining: Low resource neural machine translation

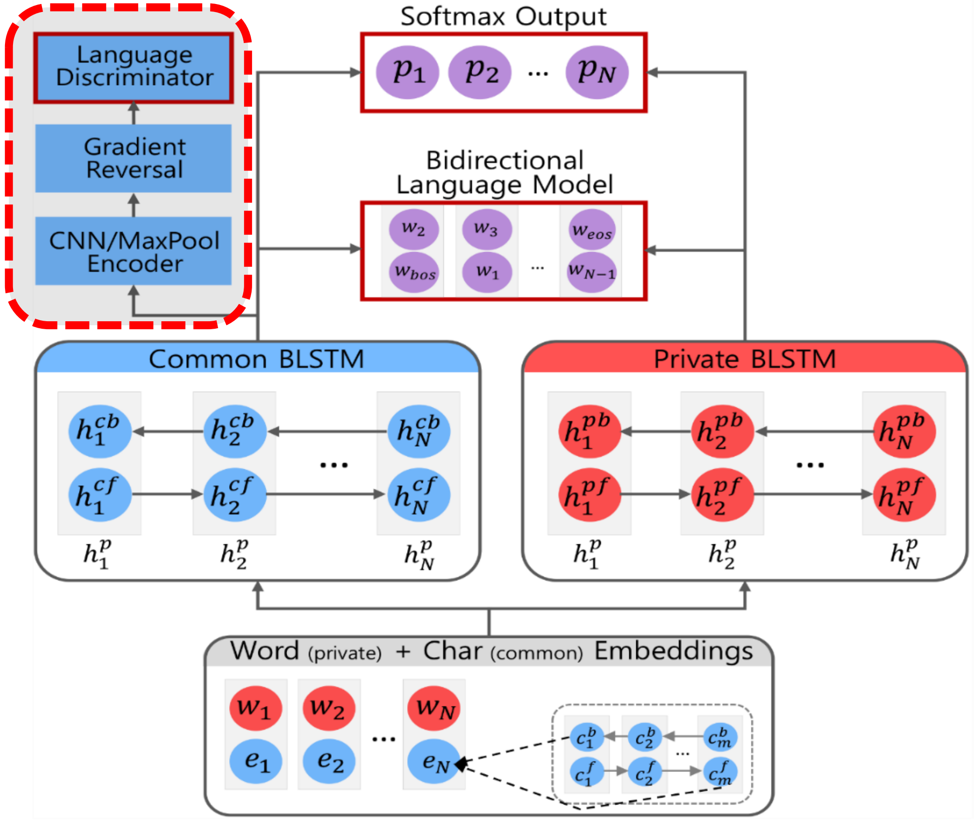
- Results

Language Pair	Parent	Train Size	BLEU ↑	PPL ↓
Uzbek–English	None	1.8m	10.7	22.4
	French–English	1.8m	**15.0 (+4.3)**	**13.9**
French′–English	None	1.8m	13.3	28.2
	French–English	1.8m	**20.0 (+6.7)**	**10.9**

Zoph, Barret, et al. "Transfer learning for low-resource neural machine translation." In Proceeddings of EMNLP, (2016).

Adversarial Training: Cross-lingual Sequence Labelling

- Adversarial training
- Language model auxiliary task



Kim, Joo-Kyung, et al. "Cross-Lingual Transfer Learning for POS Tagging without Cross-Lingual Resources." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Adversarial Training: Cross-lingual Sequence Labelling

- Results on UD treebank

		Target only		Source (English) → Target				
Language Family	Language	p	p,l	p,l	c,l	p,c,l	c,l+a	p,c,l+a
Germanic	Swedish	87.43	90.49	**91.02**	90.45	90.48	90.72	90.70
	Danish	86.42	90.00	90.74	90.69	90.02	90.16	**90.79**
	Dutch	76.76	82.24	**82.61**	82.46	82.10	82.58	82.15
	German	86.25	88.95	89.10	88.69	88.93	88.08	**89.68**
	Avg	84.22	87.92	**88.37**	88.07	87.88	87.88	88.33
Slavic	Slovenian	87.02	89.97	90.29	90.00	90.32	89.58	**90.59**
	Polish	82.10	84.13	85.21	85.41	85.30	85.46	**85.50**
	Slovak	76.22	81.03	82.95	**83.40**	82.68	82.70	83.17
	Bulgarian	87.32	**92.81**	92.68	92.07	92.30	92.20	92.39
	Avg	83.16	86.98	87.78	87.72	87.65	87.48	**87.91**
Romance	Romanian	88.67	**91.44**	**91.44**	90.87	91.22	90.85	91.37
	Portuguese	90.66	93.73	93.55	93.90	93.81	93.58	**94.20**
	Italian	89.78	93.99	93.82	93.27	93.46	93.51	**94.00**
	Spanish	85.91	**91.07**	90.59	90.59	**91.07**	90.17	90.88
	Avg	88.76	92.56	92.35	92.16	92.39	92.03	**92.61**
Indo-Iranian	Persian	90.64	**92.40**	91.98	91.97	92.12	92.18	91.83
Uralic	Hungarian	89.14	90.65	91.45	91.48	90.91	**91.52**	90.72
	Total Avg	86.02	89.49	89.82	89.66	89.62	89.52	**89.86**

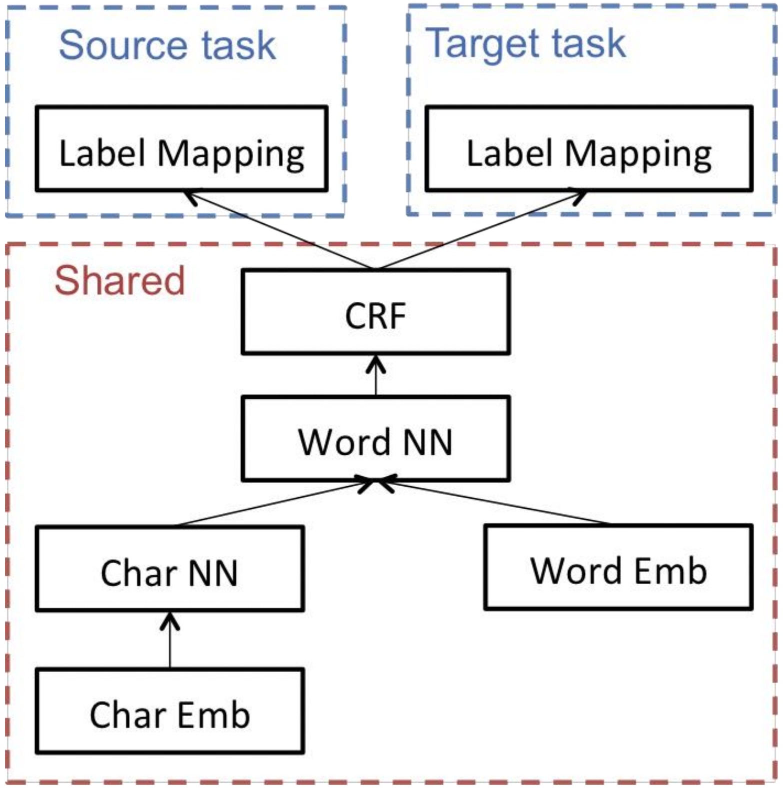
Kim, Joo-Kyung, et al. "Cross-Lingual Transfer Learning for POS Tagging without Cross-Lingual Resources." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Neural Graph-based Models (Multi-task Learning)

- Cross Task
- Cross Lingual
- Cross Domain
- Cross Standard

Sequence Tagging

- Standard Multi-task



Yang, Zhilin, Ruslan Salakhutdinov, and William W. Cohen. "Transfer learning for sequence tagging with hierarchical recurrent networks." ICLR. 2017.

Sequence Tagging

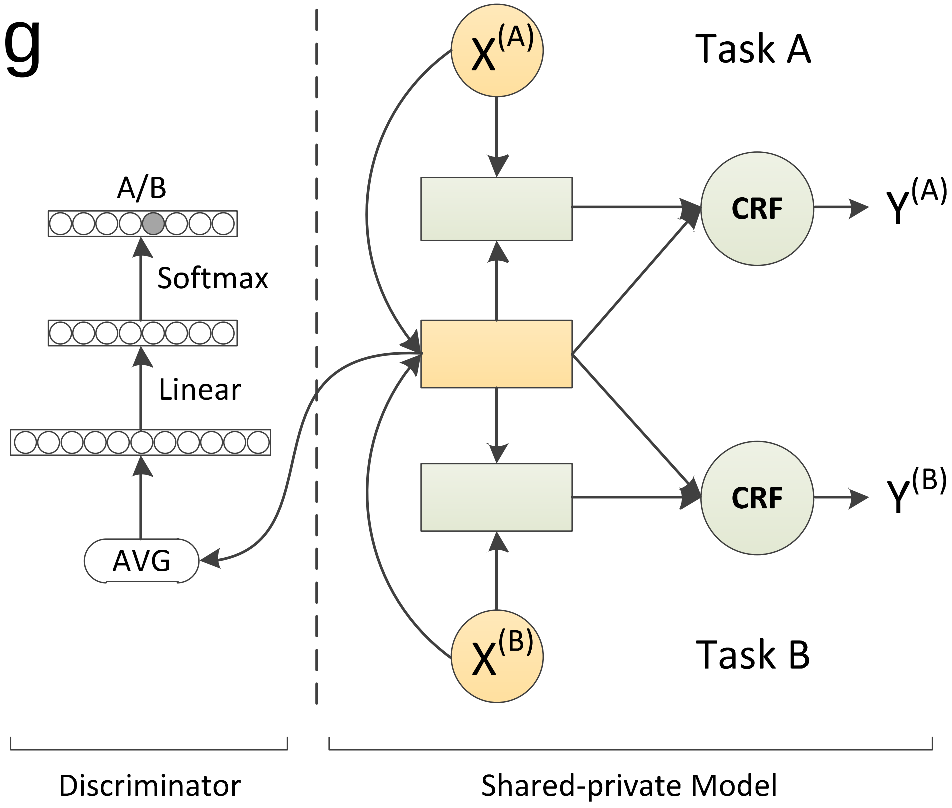
- Results

Source	Target	Model	Setting	Transfer	No Transfer	Delta
PTB	Twitter/0.1	T-A	dom	83.65	74.80	8.85
CoNLL03	Twitter/0.1	T-A	dom	43.24	34.65	8.59
PTB	CoNLL03/0.01	T-B	app	74.92	68.64	6.28
PTB	CoNLL00/0.01	T-B	app	86.73	83.49	3.24
CoNLL03	PTB/0.001	T-B	app	87.47	84.16	3.31
Spanish	CoNLL03/0.01	T-C	ling	72.61	68.64	3.97
CoNLL03	Spanish/0.01	T-C	ling	60.43	59.84	0.59
PTB	Genia/0.001	T-A	dom	92.62	83.26	9.36
CoNLL03	Genia/0.001	T-B	dom&app	87.47	83.26	4.21
Spanish	Genia/0.001	T-C	dom&app&ling	84.39	83.26	1.13
PTB	Genia/0.001	T-B	dom	89.77	83.26	6.51
PTB	Genia/0.001	T-C	dom	84.65	83.26	1.39

Yang, Zhilin, Ruslan Salakhutdinov, and William W. Cohen. "Transfer learning for sequence tagging with hierarchical recurrent networks." ICLR. 2017.

Chinese Word Segmentation

- Adversarial training



Xinchi Chen, Zhan Shi, Xipeng Qiu, Xuanjing Huang. Adversarial Multi-Criteria Learning for Chinese Word Segmentation, ACL, 2017.

Chinese Word Segmentation

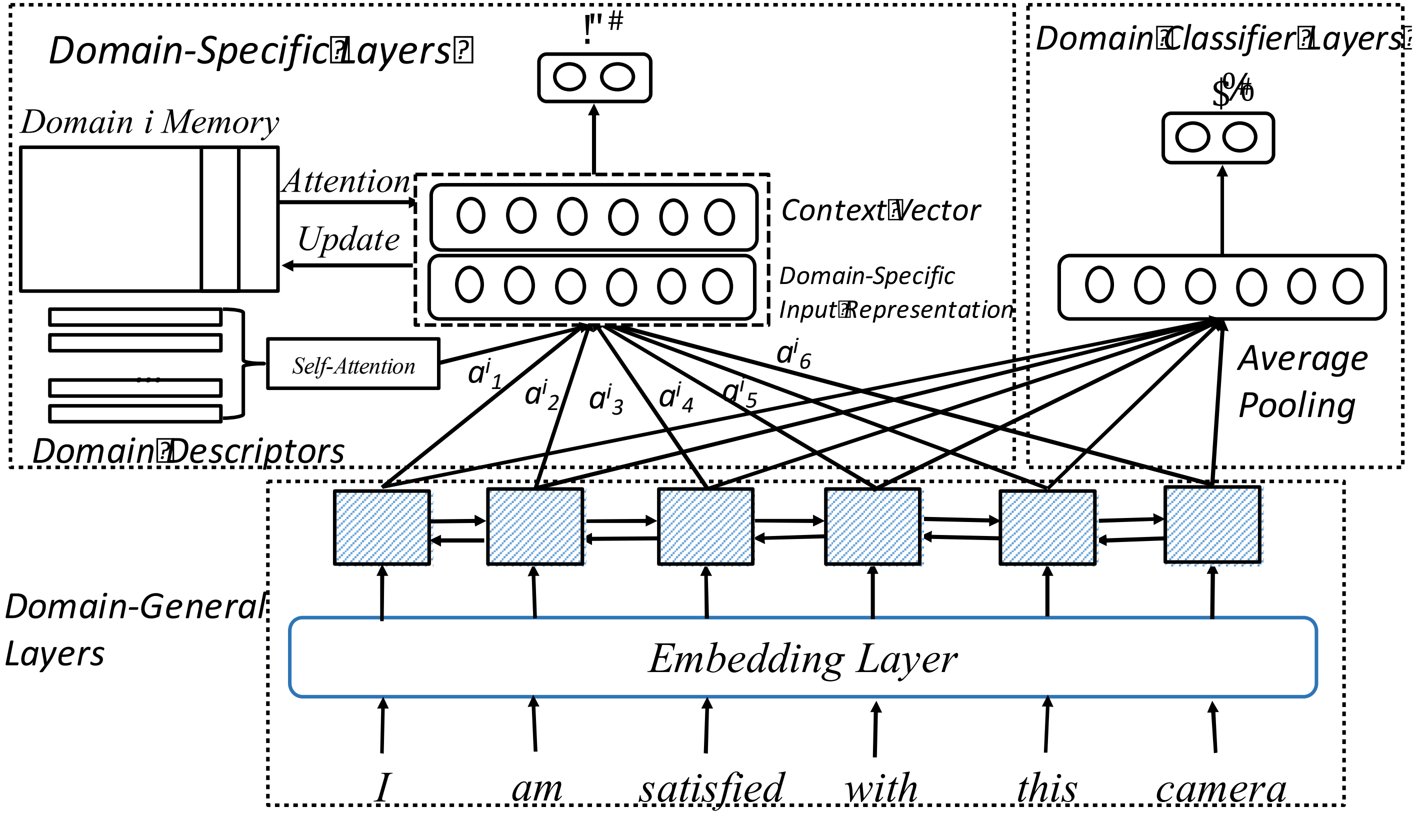
- Results

Adversarial Multi-Criteria Learning										
Model-I+ADV	P	95.95	94.17	94.86	96.02	93.82	95.39	92.46	96.07	94.84
	R	96.14	95.11	93.78	96.33	94.70	95.70	93.19	96.01	95.12
	F	**96.04**	94.64	**94.32**	**96.18**	**94.26**	**95.55**	**92.83**	96.04	**94.98**
	OOV	71.60	73.50	72.67	82.48	77.59	81.40	63.31	77.10	74.96
Model-II+ADV	P	96.02	94.52	94.65	96.09	93.80	95.37	92.42	95.85	94.84
	R	95.86	94.98	93.61	95.90	94.69	95.63	93.20	96.07	94.99
	F	95.94	**94.75**	94.13	96.00	94.24	95.50	92.81	95.96	94.92
	OOV	72.76	75.37	73.13	82.19	77.71	81.05	62.16	76.88	75.16
Model-III+ADV	P	95.92	94.25	94.68	95.86	93.67	95.24	92.47	96.24	94.79
	R	95.83	95.11	93.82	96.10	94.48	95.60	92.73	96.04	94.96
	F	95.87	94.68	94.25	95.98	94.07	95.42	92.60	**96.14**	94.88
	OOV	70.86	72.89	72.20	81.65	76.13	80.71	63.22	77.88	74.44

Xinchi Chen, Zhan Shi, Xipeng Qiu, Xuanjing Huang. Adversarial Multi-Criteria Learning for Chinese Word Segmentation, ACL, 2017.

Multi-domain Sentiment Classification

- Adversarial training
- Domain Embeddings
- Memory Network



Qi Liu, Yue Zhang, Jiangming Liu, 2018. Learning Domain Representation for Multi-domain Sentiment Classification. In Proceedings of 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), New Orleans, Louisiana, June.

Multi-domain Sentiment Classification

- Results

	In domain							Cross domain									
Dataset	MTRL	Mix	Multi	DSR	DSR-sa	DSR-ctx	DSR-at	MTRL	Mix	MDA	Multi	FEMA	NDA	DSR	DSR-sa	DSR-ctx	DSR-at
Apparel	0.883	0.912	0.921	0.927	0.928	0.92	**0.938***	0.828	0.843	0.863	0.854	0.865	0.873	0.882	0.899	0.896	**0.909***
Electronics	0.853	0.881	**0.899**	0.884	0.879	0.883	0.891	0.804	0.826	0.836	0.849	0.845	0.834	0.857	0.859	0.861	**0.875***
Office	0.863	0.88	0.89	0.903	0.914	0.925	**0.933***	0.824	0.825	0.818	0.824	0.843	0.839	0.854	0.876	0.883	**0.894***
Automotive	0.842	0.864	0.873	0.886	0.891	0.902	**0.917***	0.791	0.786	0.791	0.797	0.816	0.826	0.835	0.847	0.857	**0.867***
Gourmet	0.814	0.838	0.84	0.852	0.856	0.858	**0.863***	0.777	0.775	0.764	0.784	0.796	0.803	0.814	0.826	**0.832**	0.828
Outdoor	0.853	0.889	0.899	0.903	0.907	0.915	**0.927***	0.785	0.796	0.805	0.815	0.836	0.829	0.856	0.861	0.867	**0.887***
Baby	0.816	0.853	0.86	0.875	0.877	0.892	**0.91***	0.803	0.816	0.814	0.821	0.834	0.84	0.845	0.878	0.873	**0.895***
Grocery	0.862	0.886	0.898	0.907	0.911	0.917	**0.933***	0.806	0.817	0.826	0.846	0.846	0.862	0.88	0.873	0.865	**0.886***
Software	0.851	0.876	0.88	0.893	0.898	0.904	**0.92***	0.795	0.811	0.816	0.836	0.845	0.836	0.85	0.862	0.884	**0.897***
Beauty	0.816	0.843	0.8567	0.862	0.867	0.864	**0.889***	0.756	0.768	0.775	0.785	0.795	0.804	0.812	0.812	0.838	**0.851***
Health	0.871	0.901	0.904	0.896	0.897	0.896	**0.907**	0.785	0.807	0.819	0.832	0.845	0.848	0.843	0.834	0.857	**0.871***
Sports	0.851	0.883	0.899	0.889	0.882	0.895	**0.9**	0.759	0.768	0.775	0.784	0.816	0.819	0.821	0.836	0.848	**0.864***
Book	0.743	0.803	0.79	0.804	0.809	0.815	**0.822***	0.694	0.705	0.716	0.723	0.745	0.743	0.751	0.758	0.779	**0.798***
Jewelry	0.816	0.891	0.881	0.893	0.891	0.894	**0.909***	0.762	0.769	0.774	0.785	0.795	0.808	0.815	0.835	0.857	**0.874***
Camera	0.912	0.937	0.968	0.966	0.959	0.968	**0.989***	0.869	0.878	0.886	0.896	0.894	0.908	0.917	0.925	0.942	**0.963***
Kitchen	0.815	0.858	0.863	0.875	0.887	0.894	**0.913***	0.759	0.768	0.775	0.776	0.794	0.818	0.826	0.856	0.865	**0.884** *
Toy	0.823	0.863	0.875	0.881	0.884	0.88	**0.892***	0.814	0.824	0.815	0.803	0.813	0.832	0.826	0.843	0.845	**0.857***
Phone	0.879	0.936	0.94	0.943	**0.949***	0.941	0.933	0.805	0.813	0.808	0.818	0.821	0.833	0.836	0.856	0.874	**0.894***
Magazine	0.835	0.874	0.872	0.883	0.895	0.917	**0.937***	0.805	0.819	0.817	0.816	0.83	0.841	0.845	0.857	0.871	**0.896***
Video	0.851	0.873	0.882	0.891	0.896	0.912	**0.925***	0.754	0.774	0.794	0.795	0.815	0.822	0.834	0.845	0.855	**0.875***
Games	0.867	0.886	0.89	0.883	0.886	0.887	**0.9***	0.681	0.684	0.708	0.718	0.723	0.734	0.746	0.765	**0.781**	0.778
Music	0.752	0.782	0.8	0.798	0.8	0.798	**0.81***	0.775	0.769	0.779	0.784	0.795	0.824	0.815	0.823	0.842	**0.858***
Dvd	0.795	0.826	0.834	0.847	0.854	0.867	**0.889***	0.801	0.794	0.804	0.794	0.814	0.827	0.835	0.845	0.851	**0.875***
Instrument	0.873	0.943	**0.957***	0.896	0.906	0.898	0.9	0.814	0.805	0.813	0.815	0.825	0.836	0.833	0.835	0.845	**0.865***
Tools	0.887	0.915	0.931	0.928	0.93	0.932	**0.94***	0.805	0.814	0.828	0.835	0.846	0.857	0.864	0.866	0.873	**0.897***
Average	0.841	0.875	0.884	0.887	0.89	0.895	**0.907***	0.786	0.794	0.801	0.807	0.82	0.827	0.835	0.847	0.858	**0.873***

Qi Liu, Yue Zhang, Jiangming Liu, 2018. Learning Domain Representation for Multi-domain Sentiment Classification. In Proceedings of 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), New Orleans, Louisiana, June.

Neural Graph-based Models (Multi-task Learning)

- Cross Task
- Cross Lingual
- Cross Domain
- Cross Standard

POS tagging

- Same language, different standard

CTB:

中国	最大	氨纶丝	生产	基地	在	连云港	建成	。
NR	JJ	NN	NN	NN	P	NR	VV	PU

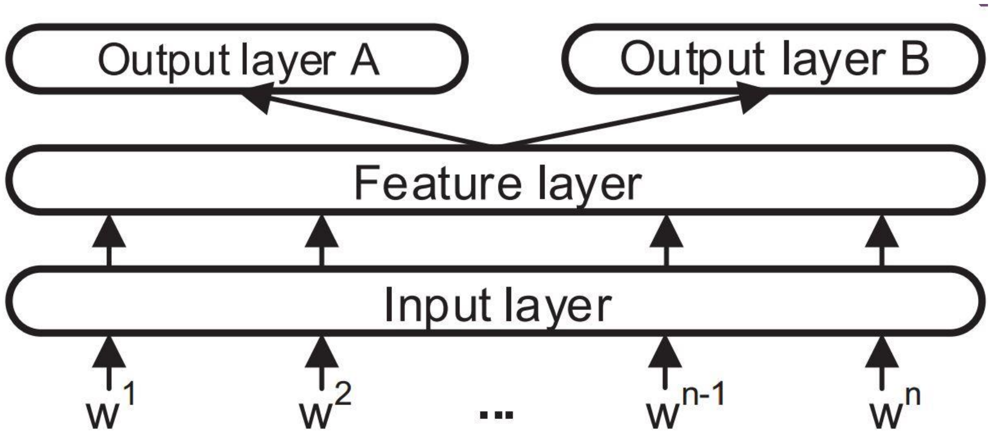
PD:

第四	条	防震	减灾	工作	，	应当	纳入	国民经济	和	社会	发展	计划	。
m	q	vn	vn	vn	w	v	v	n	c	n	vn	n	w

Chen, Hongshen, Yue Zhang, and Qun Liu. "Neural Network for Heterogeneous Annotations." *EMNLP*. 2016.

POS tagging

- Standard neural multi-view model



Chen, Hongshen, Yue Zhang, and Qun Liu. "Neural Network for Heterogeneous Annotations." *EMNLP*. 2016.

POS tagging

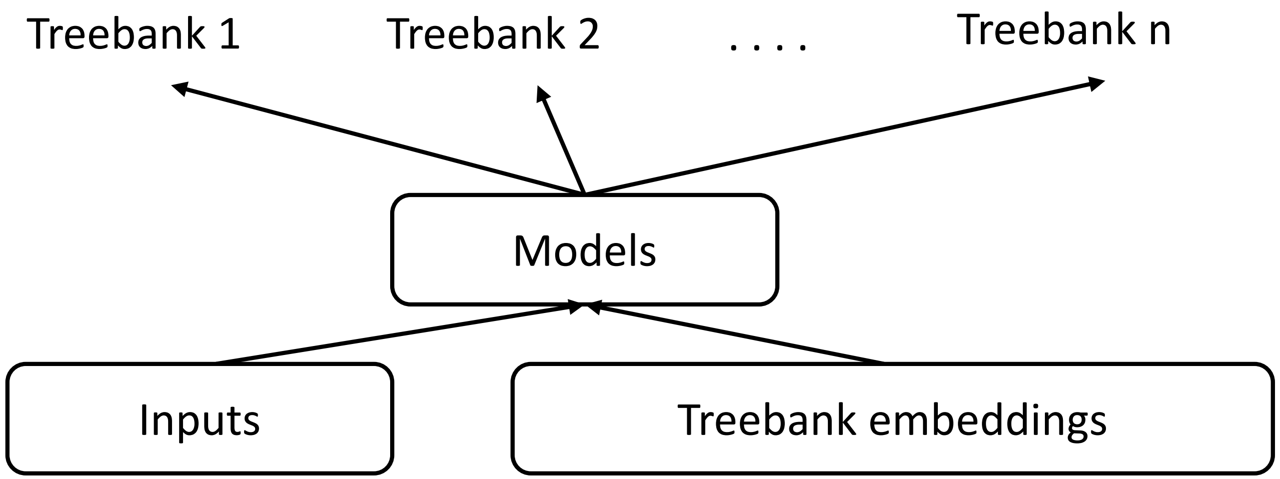
- Results

System	Accuracy
CRF Baseline (Li et al., 2015)	94.10
CRF Stacking (Li et al., 2015)	94.81
CRF Multi-view (Li et al., 2015)	95.00
NN Baseline	94.24
NN Stacking	94.74
NN Feature Stacking	95.01
NN Feature Stacking & Fine-tuning	95.32
NN Multi-view	95.40
Integrated NN Multi-view & Stacking	**95.53**

Chen, Hongshen, Yue Zhang, and Qun Liu. "Neural Network for Heterogeneous Annotations." *EMNLP*. 2016.

Dependency parsing

- Same language with multiple treebanks
- Treebank embeddings



Stymne, Sara, et al. "Parser Training with Heterogeneous Treebanks." arXiv preprint arXiv:1805.05089 (2018).

Dependency parsing

- Results

			Same treebank test set				PUD test set			
Language	**Treebank**	**Size**	SINGLE	CONCAT	C+FT	TB-EMB	SINGLE	CONCAT	C+FT	TB-EMB
Czech	PDT	68495	86.7	87.5^{+}	**88.3***	87.2^{+}	**81.7**	**81.7**	81.6	81.2
	CAC	23478	86.0	87.8^{+}	88.1^{+}	**88.5**$^{+}$	75.0		81.3	81.1
	FicTree	10160	84.3	89.3^{+}	**89.5**$^{+}$	89.2^{+}	66.1		79.8	80.3
	CLTT	860	72.5	86.2^{+}	**86.9**$^{+}$	86.0^{+}	42.1		80.8	80.9
English	EWT	12543	82.2	82.1	82.5	**83.0**	80.7	80.0	81.7*	**81.9***
	LinES	2738	72.1	76.7^{+}	**77.3**$^{+}$	**77.3**$^{+}$	62.6		75.9	74.5
	ParTUT	1781	80.5	83.5^{+}	85.4^{+}	**85.7**$^{+}$	68.0		78.1	76.9
Finnish	FTB	14981	76.4$^{\times}$	74.4	80.1*	**80.6***	46.7	73.0	54.6	53.1
	TDT	12217	78.1$^{\times}$	70.6	**80.6***	80.3*	78.6$^{\times}$		**81.3***	80.9*
French	FTB	14759	83.2	83.2	83.9*	**84.1***	72.0	79.4	76.7	74.1
	GSD	14554	84.5	84.1	85.3	**85.6**$^{\times}$	79.1		80.2*	**80.3***
	Sequoia	2231	84.0	86.0^{+}	**89.8***	89.1*	69.5		78.1	77.6
	ParTUT	803	79.8	80.5	89.1*	**90.3***	63.4		78.8	77.5
Italian	ISDT	12838	87.7	**87.9**	87.7	87.6	85.4	86.0	85.7	86.0
	PoSTWITA	2808	71.4	76.7^{+}	76.8^{+}	**77.0**$^{+}$	68.5		85.7	85.3
	ParTUT	1781	83.4	89.2^{+}	**89.3**$^{+}$	88.8^{+}	77.4		85.8^{+}	**86.1**$^{+}$
Portuguese	GSD	9664	88.3	87.3	89.0*	**89.1***	74.0	76.8^{+}	75.2	74.9
	Bosque	8331	84.7	84.2	86.2$^{\times}$	**86.3***	75.2		77.5^{+}	**77.6**$^{+}$
Russian	SynTagRus	48814	90.2$^{\times}$	89.4	**90.4**$^{\times}$	**90.4**$^{\times}$	66.0	68.7	66.3	66.4
	GSD	3850	74.7$^{\times}$	73.4	79.8*	**80.8***	70.1$^{\times}$		77.6*	**78.0***
Spanish	AnCora	14305	87.2$^{\times}$	86.2	87.5$^{\times}$	**87.6**$^{\times}$	75.2	79.9	77.7	76.4
	GSD	14187	84.7	83.0	85.8$^{\times}$	**86.2***	79.8		80.8^{+}	**80.9***
Swedish	Talbanken	4303	79.6	79.1	80.2	**80.6**$^{\times}$	70.3	72.0^{+}	73.2*	**73.6***
	LinES	2738	74.3	76.8	**77.3**$^{+}$	77.1^{+}	64.0		70.0	69.0
Average			81.4	82.7^{+}	**84.9***	**84.9***	77.9	77.5	80.0*	**80.1***

Stymne, Sara, et al. "Parser Training with Heterogeneous Treebanks." arXiv preprint arXiv:1805.05089 (2018).

Thanks!