Joint Models for NLP

Yue Zhang

Outline

- Motivation
- Statistical Models
- Deep Learning Models

Outline

- Motivation
- Statistical Models
- Deep Learning Models

Motivation

- Subtasks in NLP
  - Segmentation ⟹ POS tagging

布朗访问上海

⇩

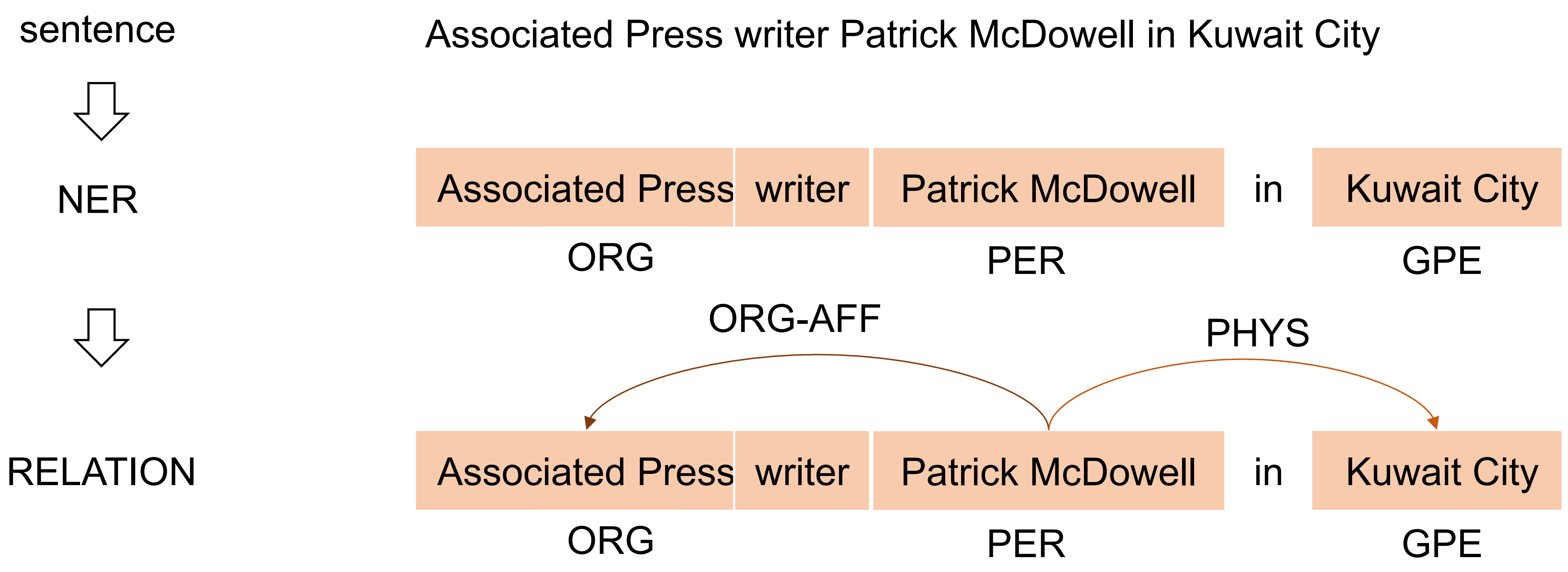
布朗/ 访问/ 上海/

⇩

布朗/NR 访问/VV 上海/NR

Motivation

- Subtasks in NLP
  - NER and Relation



Motivation

- Subtasks 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 mixing
- Challenge
  - Joint learning
  - Search

Outline

- Motivation
- Statistical Models
- Deep Learning Models

Statistical Models

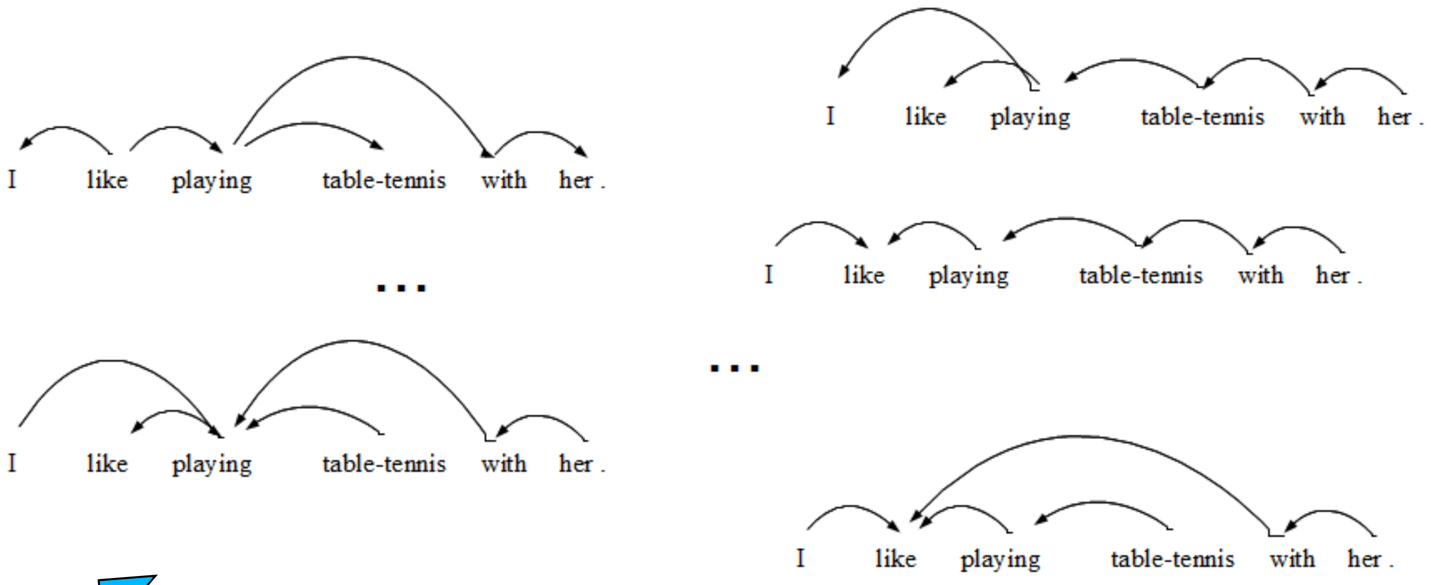
- Graph-Based Methods
- Transition-Based Methods

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



✓ **Over 100 possible trees for this seven-word sentence.**
✓ **Over one million trees for a 20-word sentence.**

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)

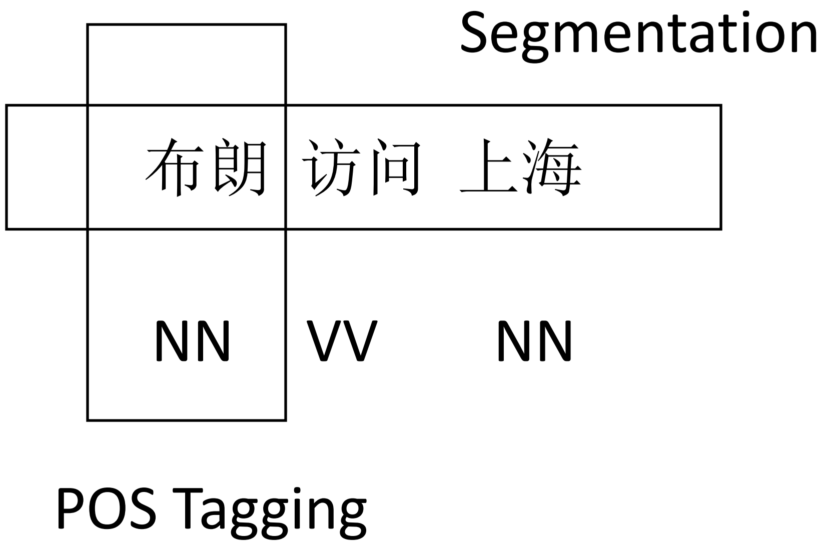
Joint Label Structure

- Two questions to building a Chinese POS tagger:
  - Should we perform Chinese POS tagging strictly after word segmentation in two separate phases (one at-a-time approach), or perform both word segmentation and POS tagging in a combined, single step simultaneously (all-at-once approach)?
  - Should we assign POS tags on a word-by-word basis (like in English), making use of word features in the surrounding context (word-based), or on a character-by-character basis with character features (character-based)?

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 Label Structure

- Collapsing labels



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 Label Structure

- 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 Label Structure

- One-at-a-Time, Word-Based POS Tagger : Feature

(a) $W_n(n = -2,-1,0,1,2)$

(b) $W_n W_{n+1}(n = -2,-1,0,1)$

(c) $W_{-1} W_1$

(d) $Pu(W_0)$

(e) $T(W_{-2})T(W_{-1})T(W_0)T(W_1)T(W_2)$

(f) $POS(W_{-1})$

(g) $POS(W_{-2})POS(W_{-1})$

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 Label Structure

- 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 Label Structure

- One-at-a-Time, Character-Based POS Tagger : 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) $POS(C_{-1W_0})$

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

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 Label Structure

- 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 Label Structure

- Results on the various methods(local maximum entropy)

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.

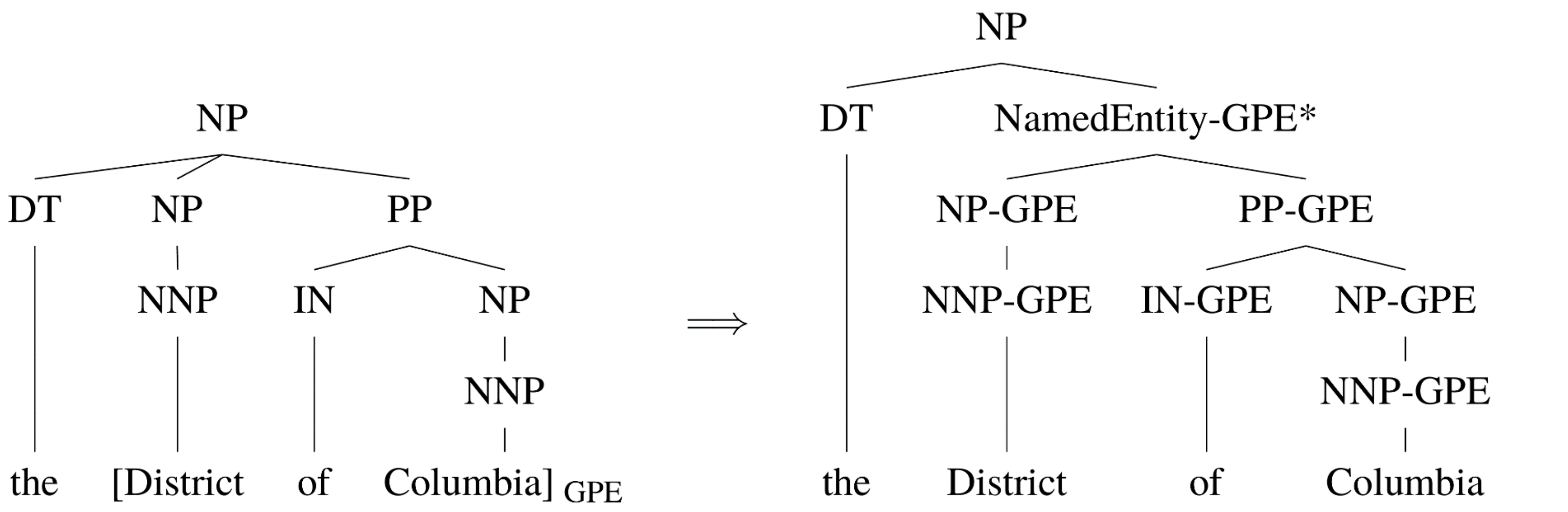
Joint Label Structure

- Results Discussions
  - Character-based approach is better than word-based approach. Unlike in English where each English letter by itself does not possess any meaning, many Chinese characters have well defined meanings. In addition, since the OOV rate for Chinese words is much higher than the OOV rate for Chinese characters, in the presence of an unknown word, using the component characters in the word to help predict the correct POS is a good heuristic.
  - The all-at-once approach, which considers all aspects of available information in an integrated, unified framework, can make better informed decisions but incurs a higher computational cost.

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

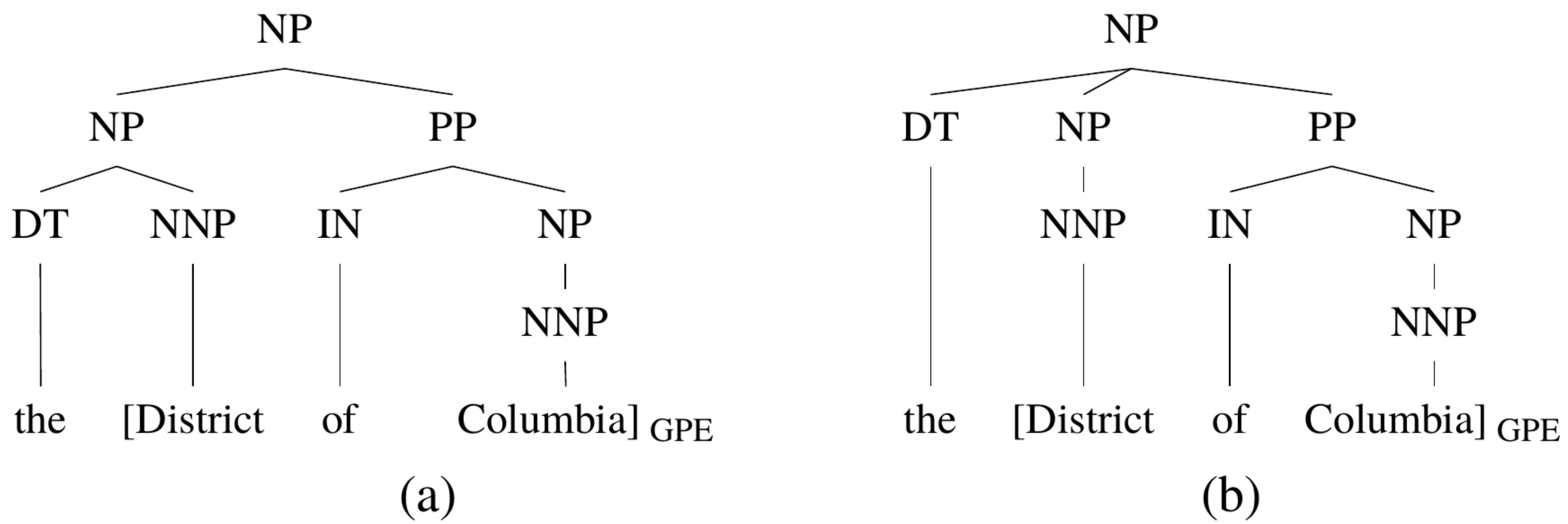
- A joint model of both parsing and named entity recognition.



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

- Data : LDC2008T04 OntoNotes Release 2.0 corpus (Hovy et al., 2006).

	Training		Testing	
	Range	# Sent.	Range	# Sent.
ABC	0–55	1195	56–69	199
CNN	0–375	5092	376–437	1521
MNB	0–17	509	18–25	245
NBC	0–29	552	30–39	149
PRI	0–89	1707	90–112	394
VOA	0–198	1512	199–264	383

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:

		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)

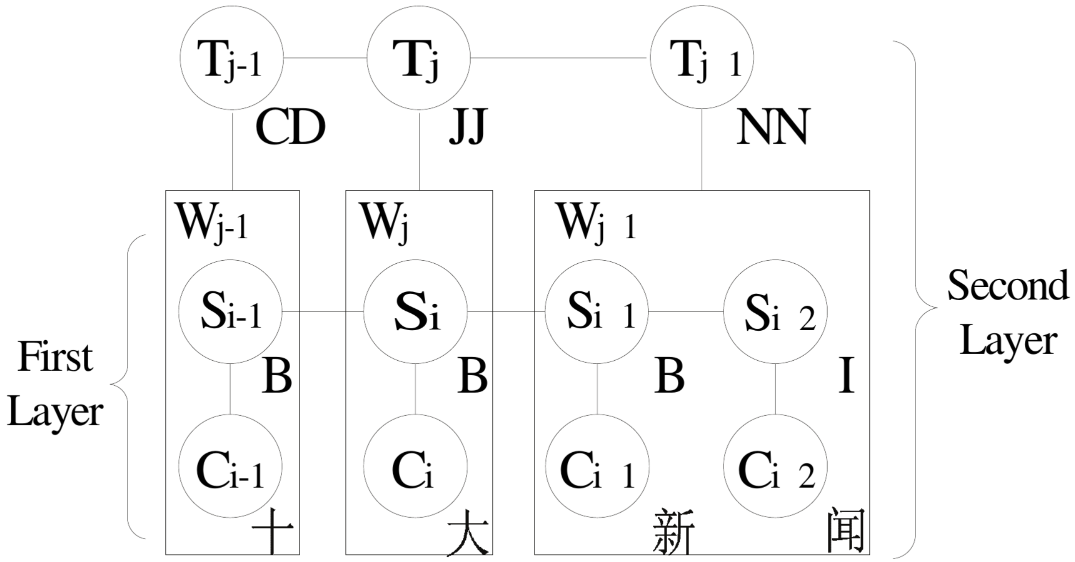
Joint Word Segmentation and POS Tagging

- This method performs joint decoding of separately trained Conditional Random Field(CRF) models, while guarding against violations of hard-constraints.
- 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 Word 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 Word 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 Word 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.

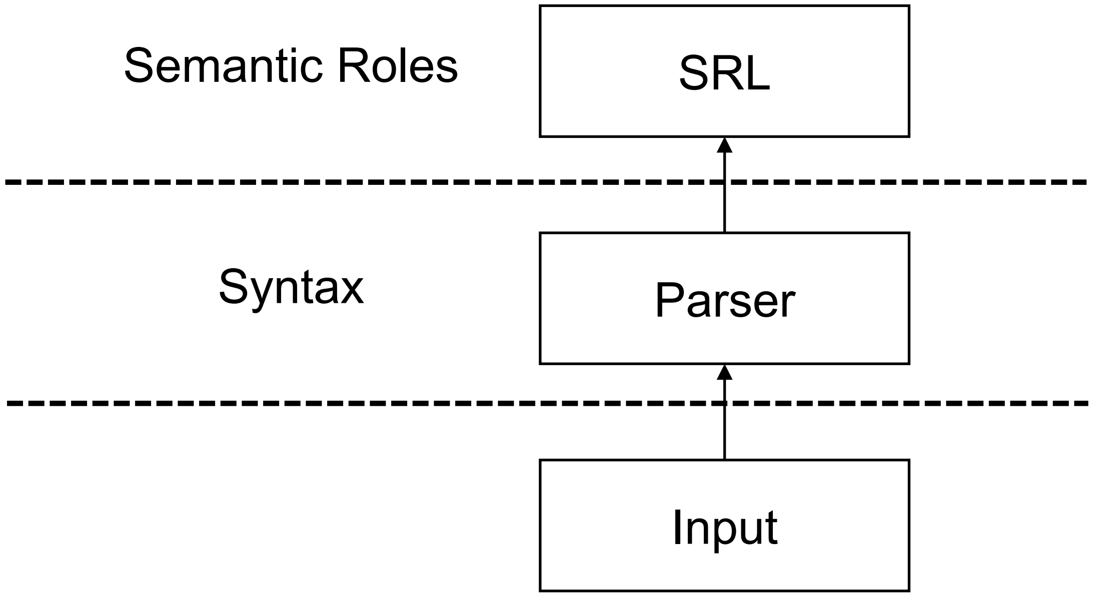
Joint Parsing and SRL

- The goal of this investigation is to narrow the gap between SRL results from gold parses and from automatic parses. The paper aims to achieve this by jointly performing parsing and semantic role labeling in a single probabilistic model. In both parsing and SRL, state-of-the-art systems are probabilistic; therefore, their predictions can be combined in a principled way by multiplying probabilities. This paper rerank the 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

- 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

- 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

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

- Detailed results on the WSJ test

Test WSJ	Precision	Recall	$F_{\beta=1}$
Overall	68.57%	64.99%	66.73
A0	69.47%	74.35%	71.83
A1	66.90%	64.91%	65.89
A2	64.42%	61.17%	62.75
A3	62.14%	50.29%	55.59
A4	72.73%	70.59%	71.64
A5	50.00%	20.00%	28.57
AM-ADV	55.90%	49.60%	52.57
AM-CAU	76.60%	49.32%	60.00
AM-DIR	57.89%	38.82%	46.48
AM-DIS	79.73%	73.75%	76.62
AM-EXT	66.67%	43.75%	52.83
AM-LOC	50.26%	53.17%	51.67
AM-MNR	54.32%	51.16%	52.69
AM-MOD	98.50%	95.46%	96.96
AM-NEG	98.20%	94.78%	96.46
AM-PNC	46.08%	40.87%	43.32
AM-PRD	0.00%	0.00%	0.00
AM-REC	0.00%	0.00%	0.00
AM-TMP	72.15%	67.43%	69.71
R-A0	0.00%	0.00%	0.00
R-A1	0.00%	0.00%	0.00
R-A2	0.00%	0.00%	0.00
R-A3	0.00%	0.00%	0.00
R-A4	0.00%	0.00%	0.00
R-AM-ADV	0.00%	0.00%	0.00
R-AM-CAU	0.00%	0.00%	0.00
R-AM-EXT	0.00%	0.00%	0.00
R-AM-LOC	0.00%	0.00%	0.00
R-AM-MNR	0.00%	0.00%	0.00
R-AM-TMP	0.00%	0.00%	0.00
V	99.21%	86.24%	92.27

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)

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

- A model that jointly identifies opinion-related entities, including opinion expressions, opinion targets and opinion holders as well as the associated opinion linking relations, IS-ABOUT and IS-FROM.

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

Joint Entity and Sentiment

- Example:
  - 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.

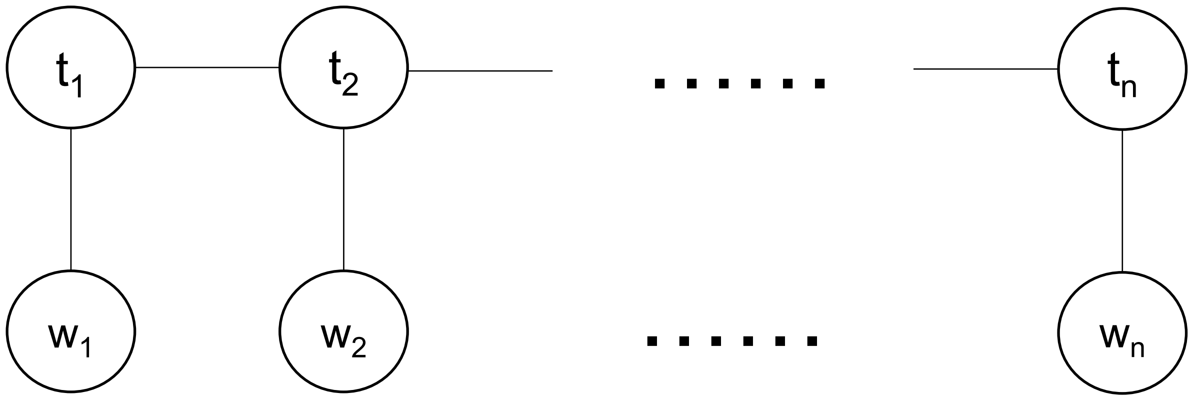
Joint Entity and Sentiment

- 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; Then, it treat the relation extraction problem as a combination of two binary classification problems and use L1-regularized logistic regression to train the classifiers; finally 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 model for opinion target relation
- A model for opinion holder relation

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

Joint Entity and Sentiment

- Joint training objective by linearposition

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 Opinion Entity Extraction

	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 Opinion Relation Extraction

	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

- This method is a single model with both supertagging and parsing features, rather than separating them into distinct models chained together in a pipeline.

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.

Lexicalized Grammar

- **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.

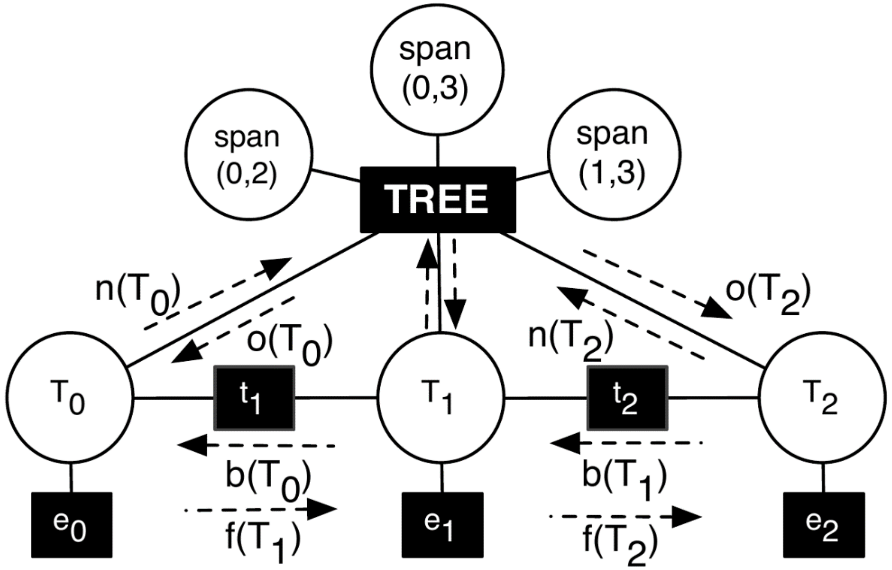
Lexicalized Grammar

- 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.

Lexicalized Grammar

- Loopy belief propagation and dual decomposition
- 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.

Lexicalized Grammar

$$\arg\max_{y \in Y, z \in Z} f(y) + g(z) \qquad (9)$$

$$\text{such that } y(i,t) = z(i,t) \text{ for all } (i,t) \in I \qquad (10)$$

$$L(u) = \max_{y \in Y}(f(y) - \sum_{i,t} u(i,t)y(i,t)) \qquad (11)$$
$$+ \max_{z \in Z}(f(z) + \sum_{i,t} u(i,t)z(i,t))$$

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.

Lexicalized Grammar

- 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	-	-	-
$\text{BP}_{k=1}$	87.70	93.28	94.44	**88.35**	93.69	**94.73**	**88.20**	**93.28**	**94.60**	88.78	93.66	94.81
$\text{BP}_{k=25}$	87.70	93.31	94.44	88.33	**93.72**	94.71	88.19	93.27	94.59	88.80	93.68	94.81
$\text{DD}_{k=1}$	87.40	93.09	94.23	87.38	93.15	94.03	87.74	93.10	94.33	87.67	93.07	94.02
$\text{DD}_{k=25}$	**87.71**	**93.32**	**94.44**	88.29	93.71	94.67	88.14	93.24	94.59	**88.80**	**93.68**	**94.82**

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)

Joint Modeling (Single task)

- A Single Model

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

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

Joint Word Segmentation and POS Tagging

- This paper propose a joint segmentation and POS tagging model that does not impose any hard constraints on the interaction between word and POS information. Fast decoding is achieved by using a novel multiple-beam search algorithm. The system uses a discriminative statistical model, trained using the generalized perceptron algorithm.

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 Word Segmentation and POS Tagging

- The averaged perceptron algorithm is adopted with the union of feature templates from the baseline segmentor and POS tagger as the feature templates

Inputs: training examples (x_i, y_i)
Initialization: set $\vec{w} = 0$
Algorithm:
  for $t = 1..T$, $i = 1..N$
    calculate $z_i = \arg\max_{y \in \mathbf{GEN}(x_i)} \Phi(y) \cdot \vec{w}$
    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 Word Segmentation and POS Tagging

- Feature templates for the baseline segmentor

1	word w	9	word w immediately before character c
2	word bigram $w_1 w_2$	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 $c_1 c_2$ 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 Word 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 wc		
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 Word Segmentation and POS Tagging

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

Input: raw sentence $sent$ – a list of characters
Variables: candidate sentence $item$ – a list of (word, tag) pairs;
maximum word-length record $maxlen$ for each tag;
the agenda list $agendas$;
the tag dictionary $tagdict$;
$start\_index$ for current word;
end_index for current word
Initialization: $agendas[0] = [``"]$,
$agendas[i] = [] \; (i! = 0)$

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: $agendas[sent$.length].best_item

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

Joint Word Segmentation and POS Tagging

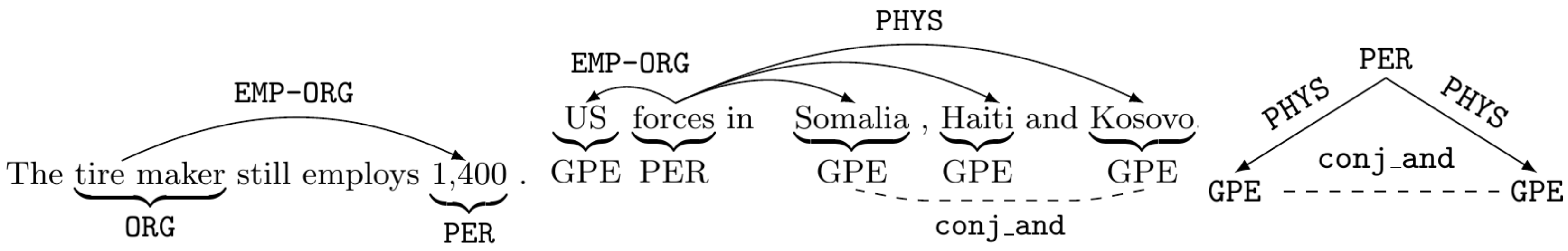
- The comparison of overall accuracies by 10-fold cross validation using CTB

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

- An incremental joint framework to simultaneously extract entity mentions and relations using structured perceptron with efficient beam-search. A segment-based decoder based on the idea of semi-Markov chain is adopted to the new framework as opposed to traditional token-based tagging.



(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

- Similar idea to (Zhang and Clark 2008)
  - A Single Model

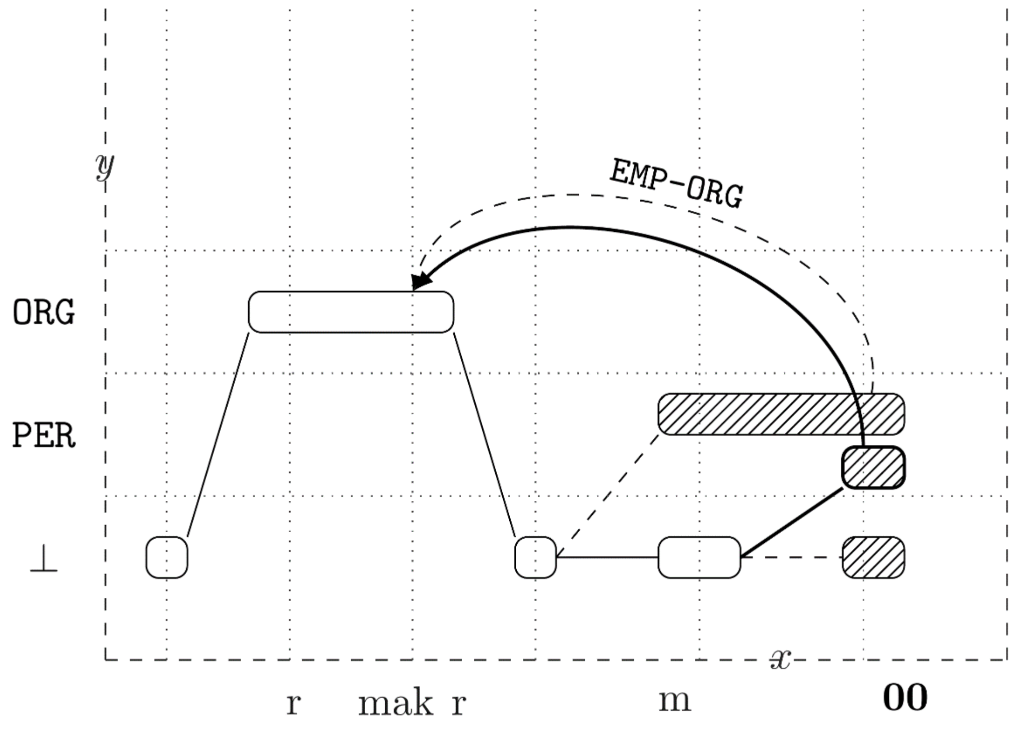
$$\hat{y} = \underset{y' \in \mathcal{Y}(x)}{\operatorname{argmax}} \, \mathbf{f}(x, y') \cdot \mathbf{w}$$

  - Beam Search

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

Joint Entity Relation Extraction

- Example of decoding steps



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

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

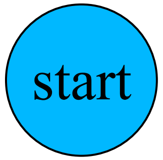
A Transition System

- 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

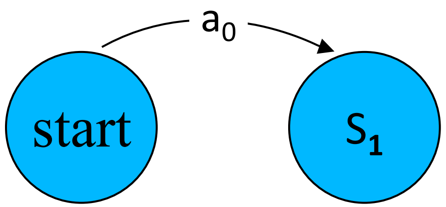
- Automata



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

A Transition System

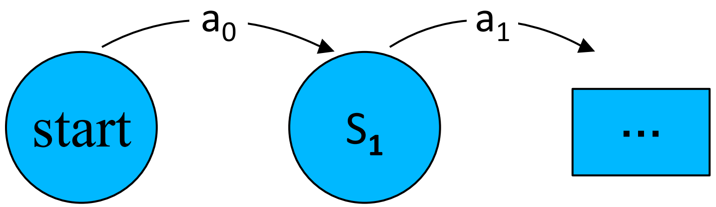
- Automata



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

A Transition System

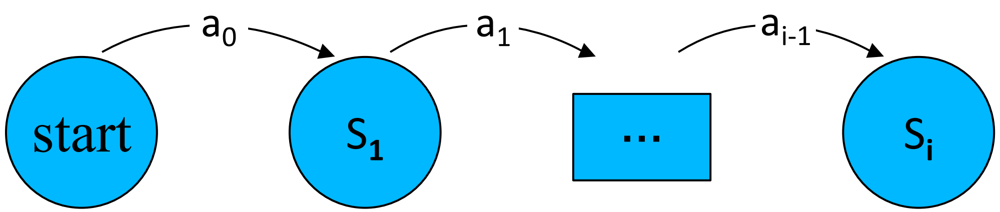
- Automata



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

A Transition System

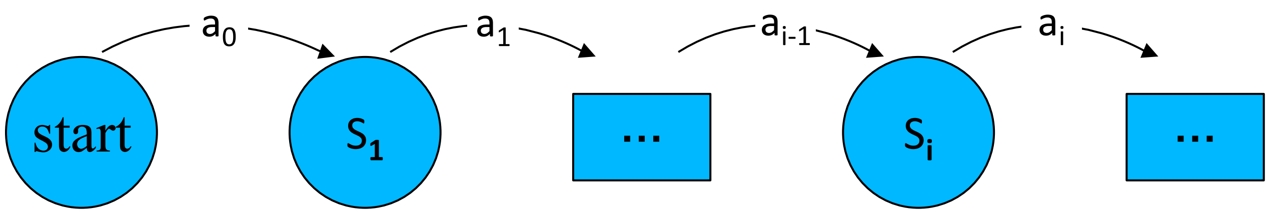
- Automata



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

A Transition System

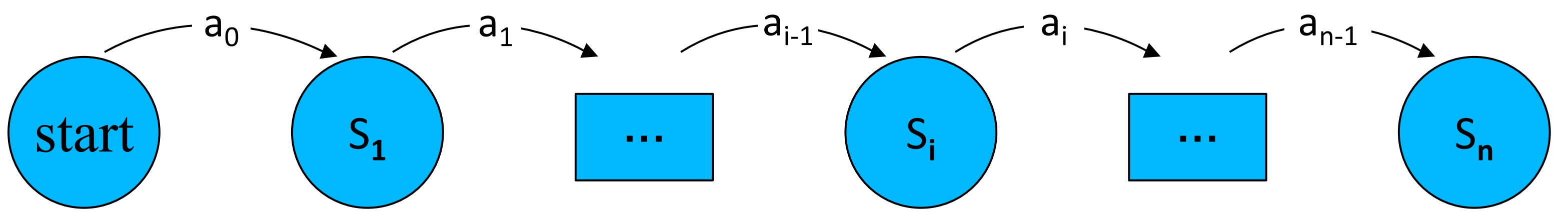
- Automata



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

A Transition System

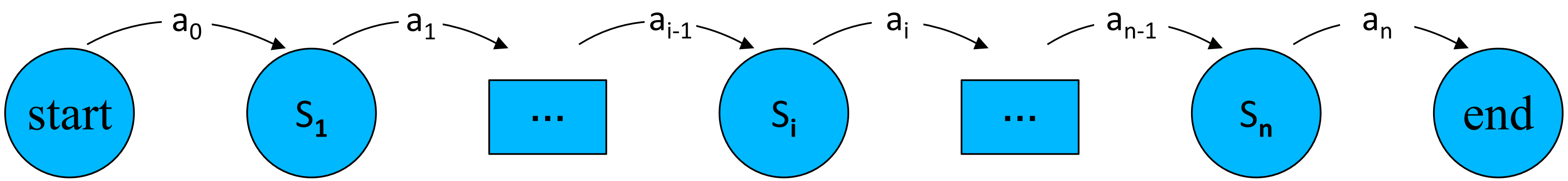
- Automata



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

A Transition System

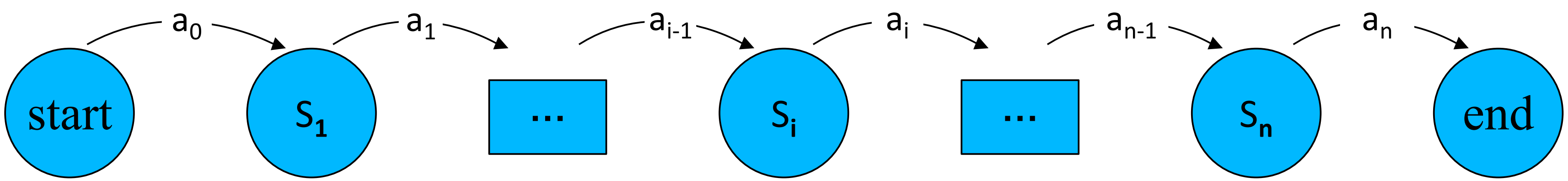
- Automata



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

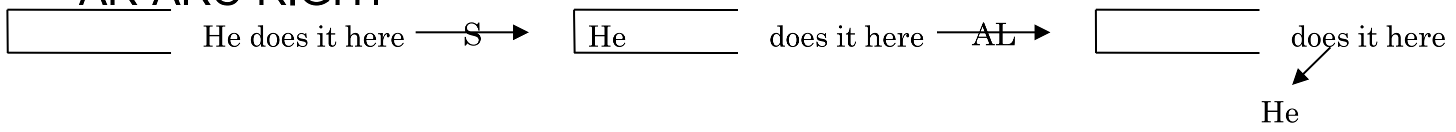
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT

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

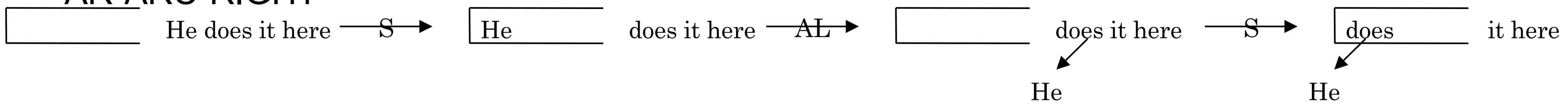
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



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

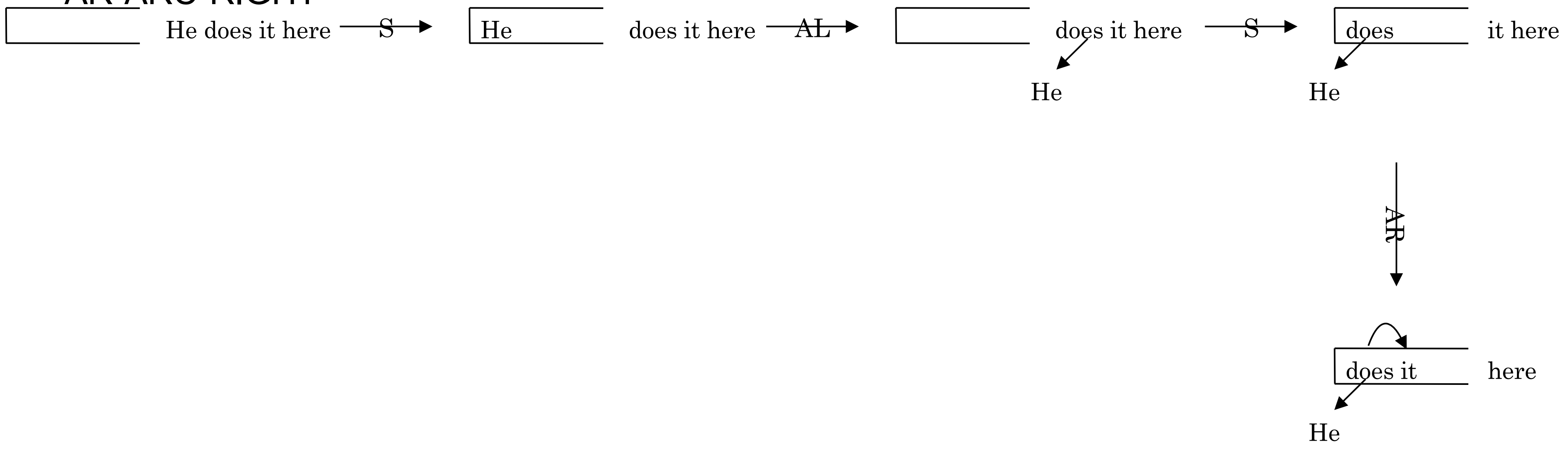
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



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

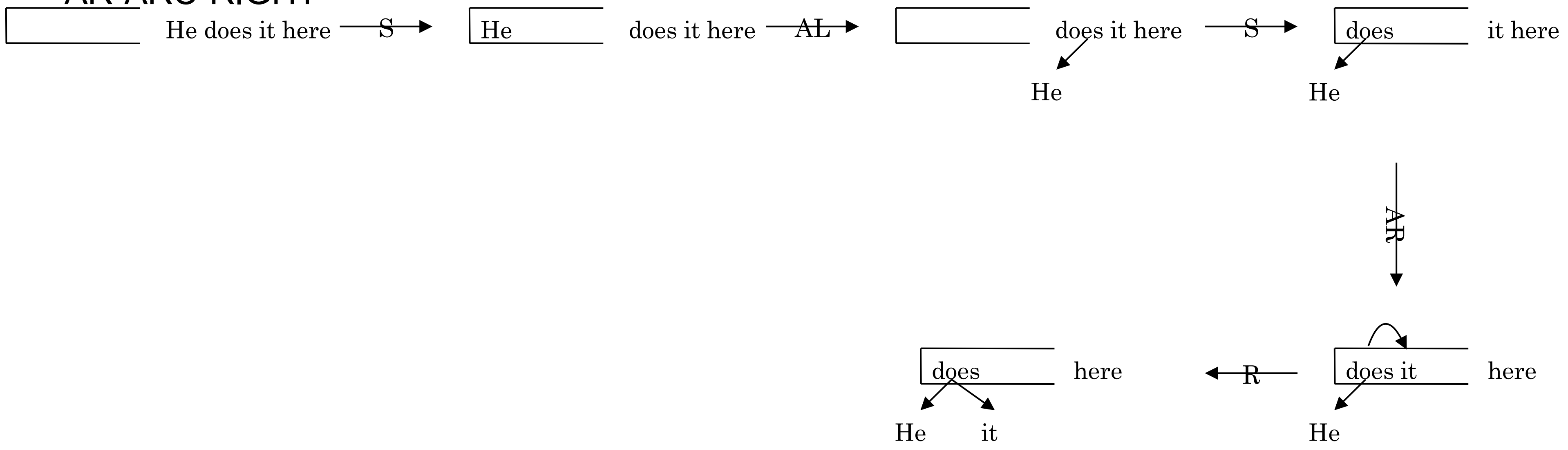
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



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

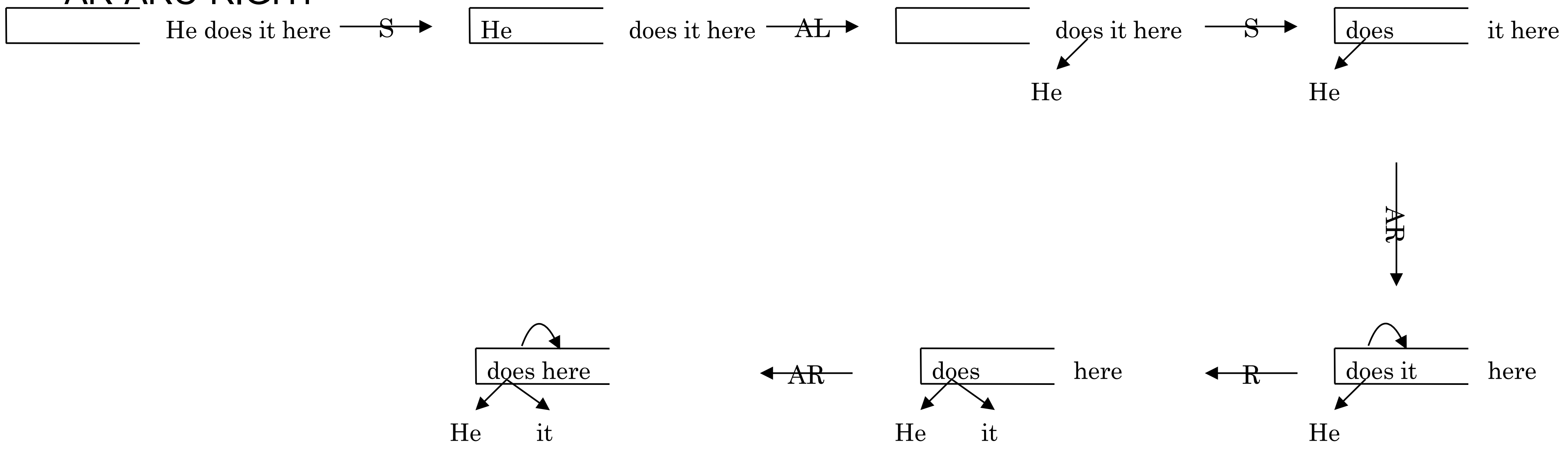
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



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

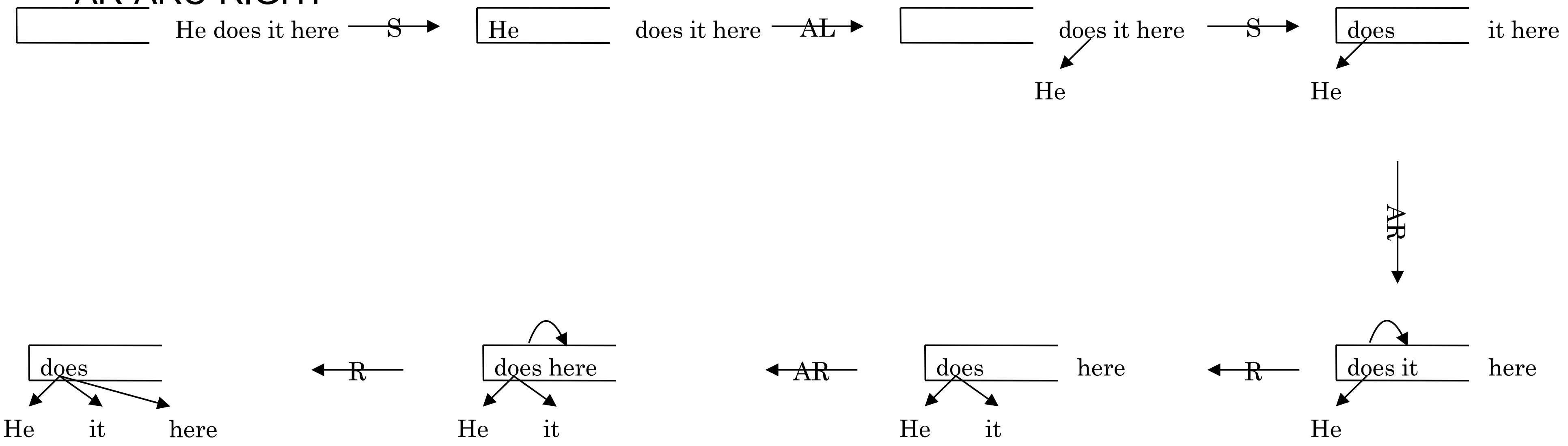
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



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

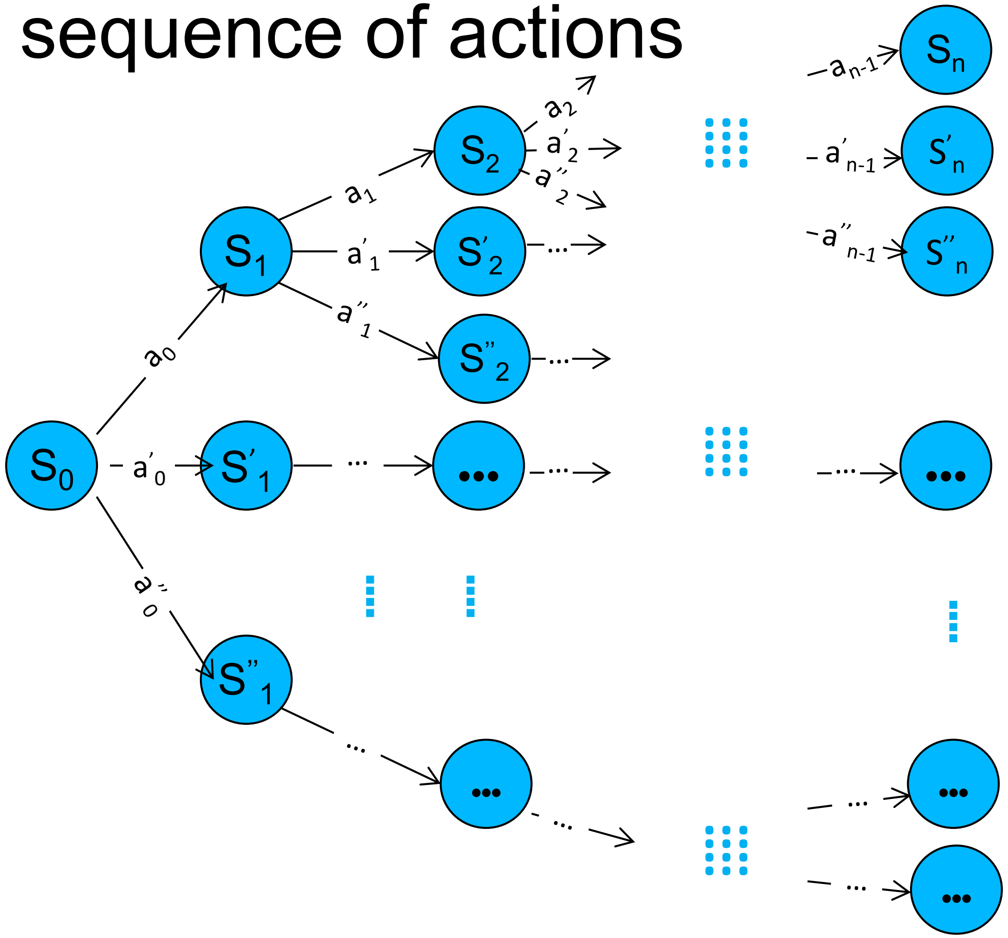
- An Example
  - S-SHIFT
  - R-REDUCE
  - AL-ARC-LEFT
  - AR-ARC-RIGHT



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

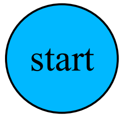
Search Space

- Find the best sequence of actions
- Exponential



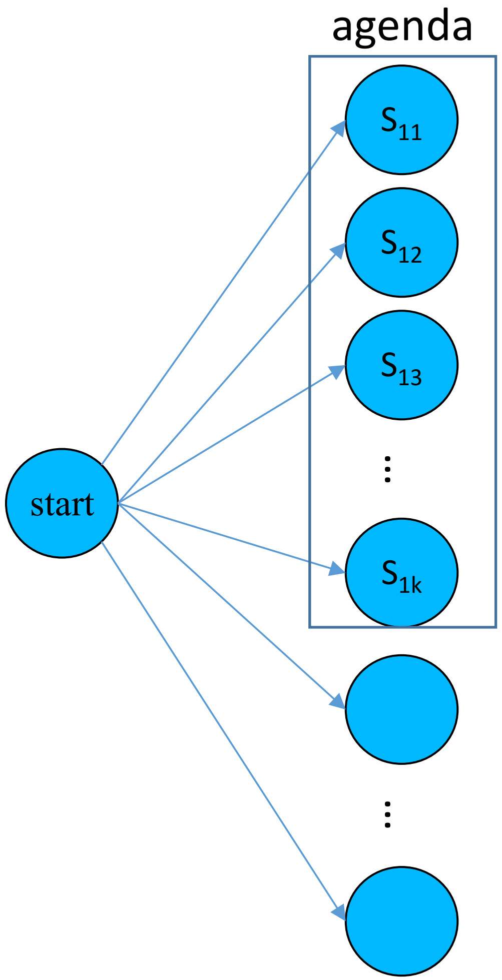
Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Beam-search decoding



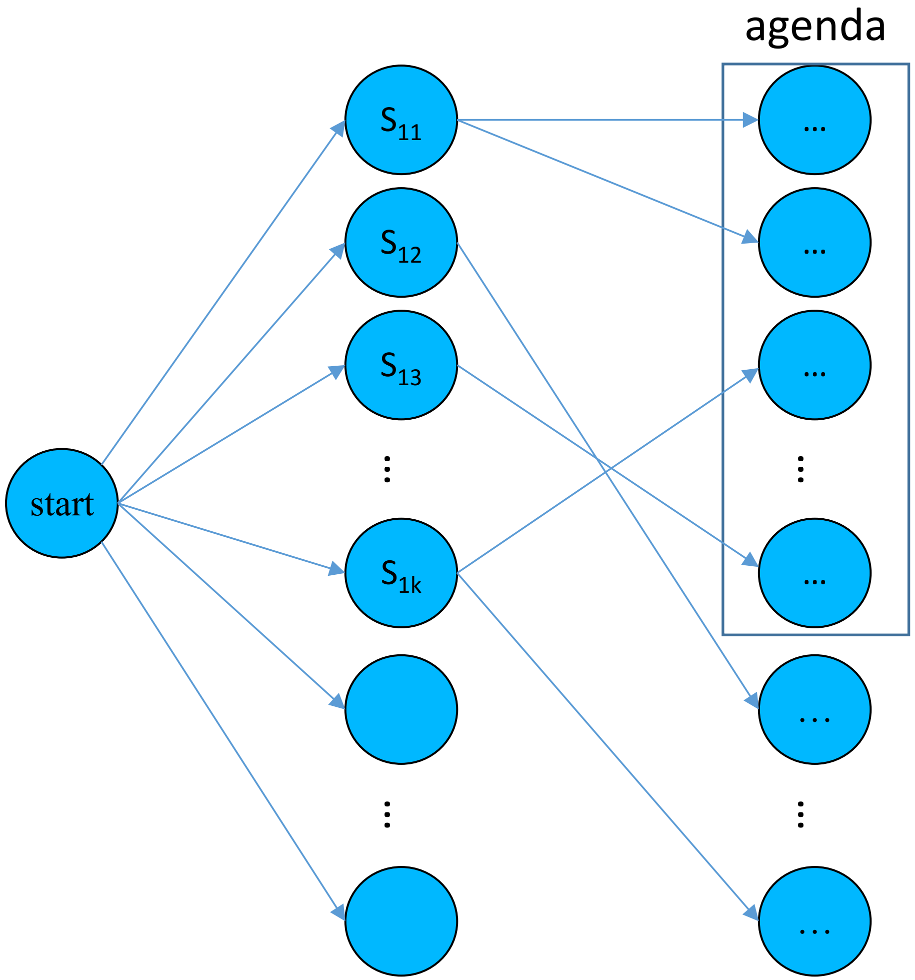
Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Beam-search decoding



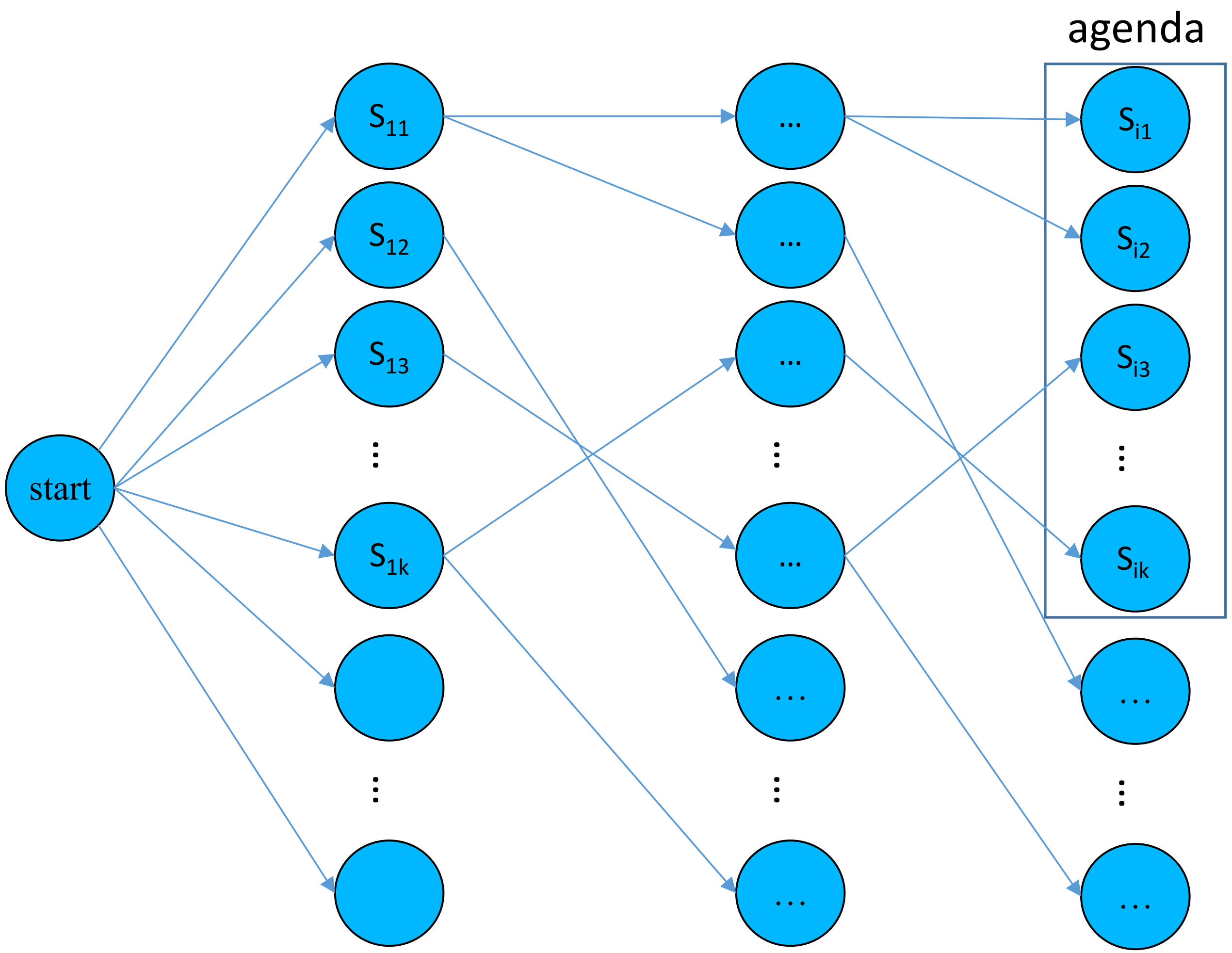
Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Beam-search decoding



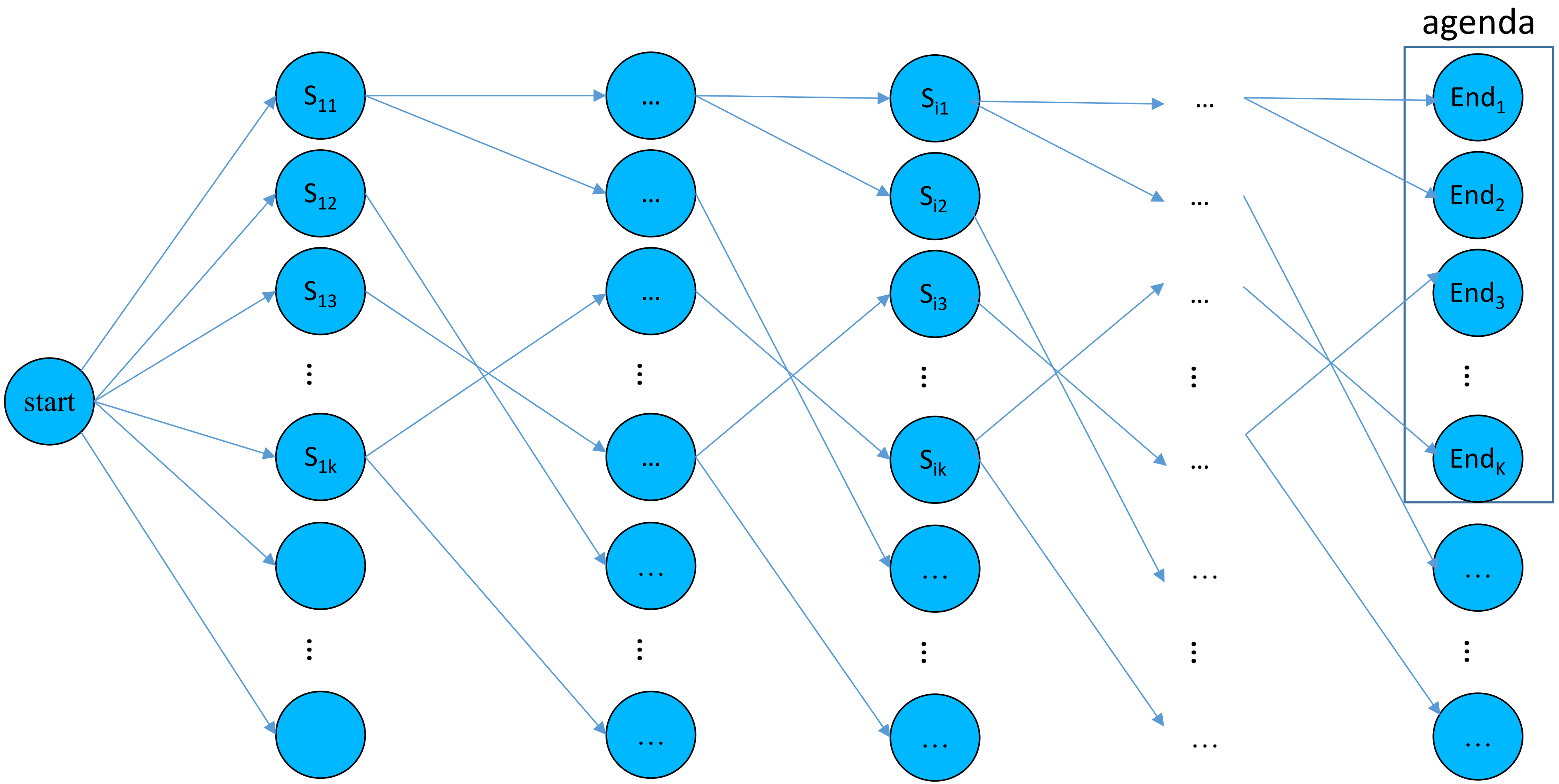
Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Beam-search decoding



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

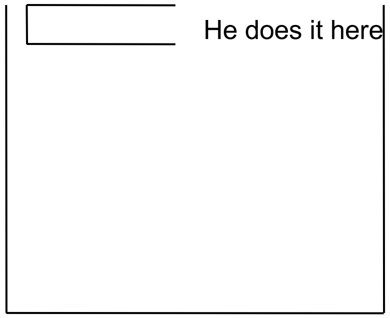
Beam-search decoding



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

Beam-search decoding

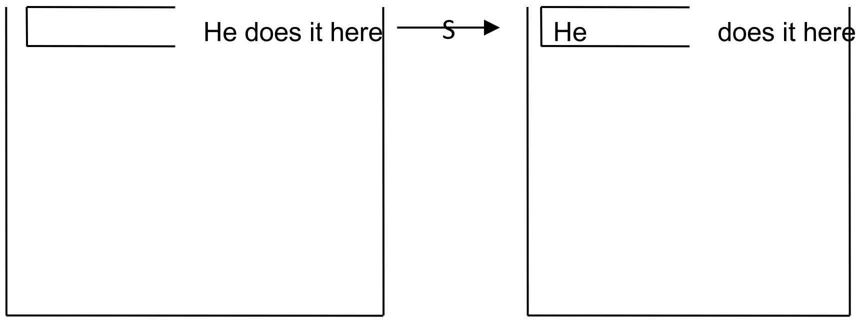
- Dependency Parsing Example
  - Decoding



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

Beam-search decoding

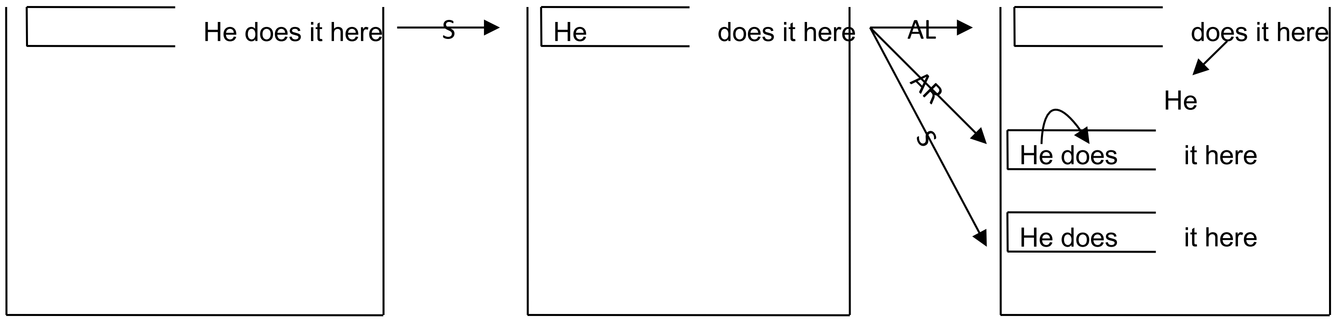
- Dependency Parsing Example
  - Decoding



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

Beam-search decoding

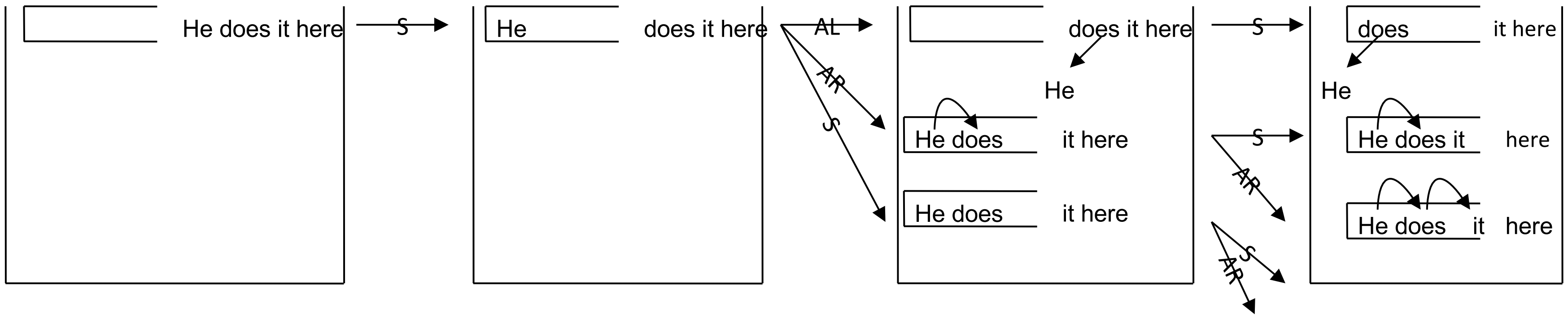
- Dependency Parsing Example
  - Decoding



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

Beam-search decoding

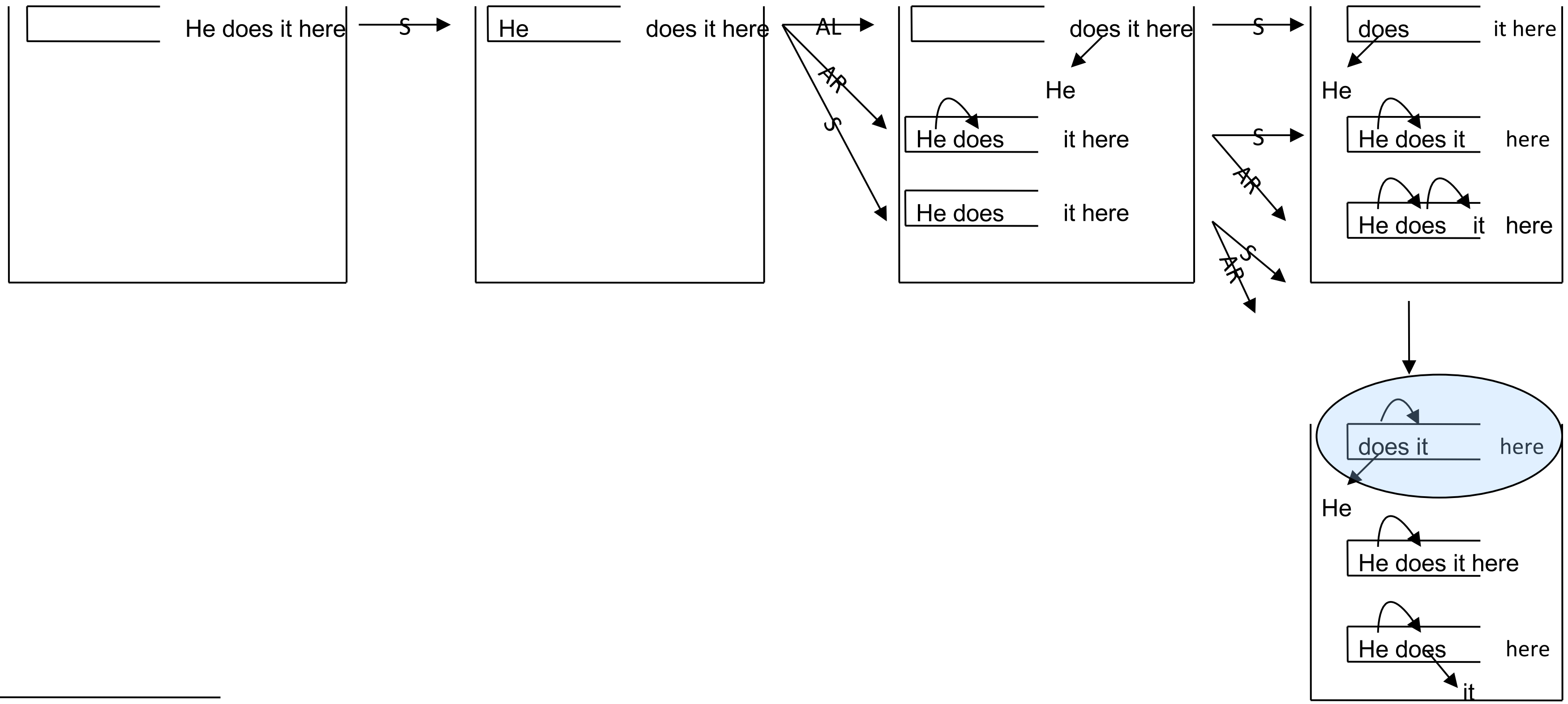
- Dependency Parsing Example
  - Decoding



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

Beam-search decoding

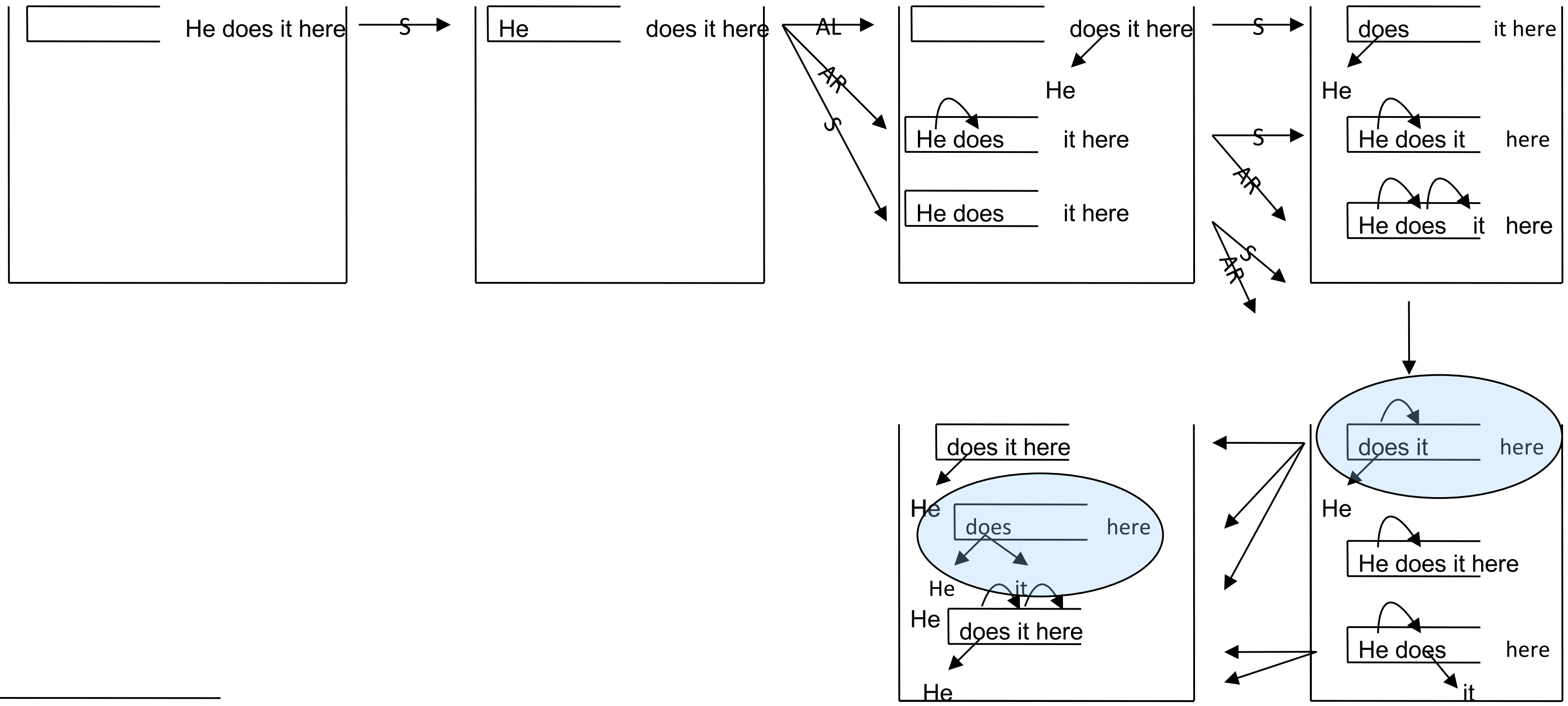
- Dependency Parsing Example
  - Decoding



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

Beam-search decoding

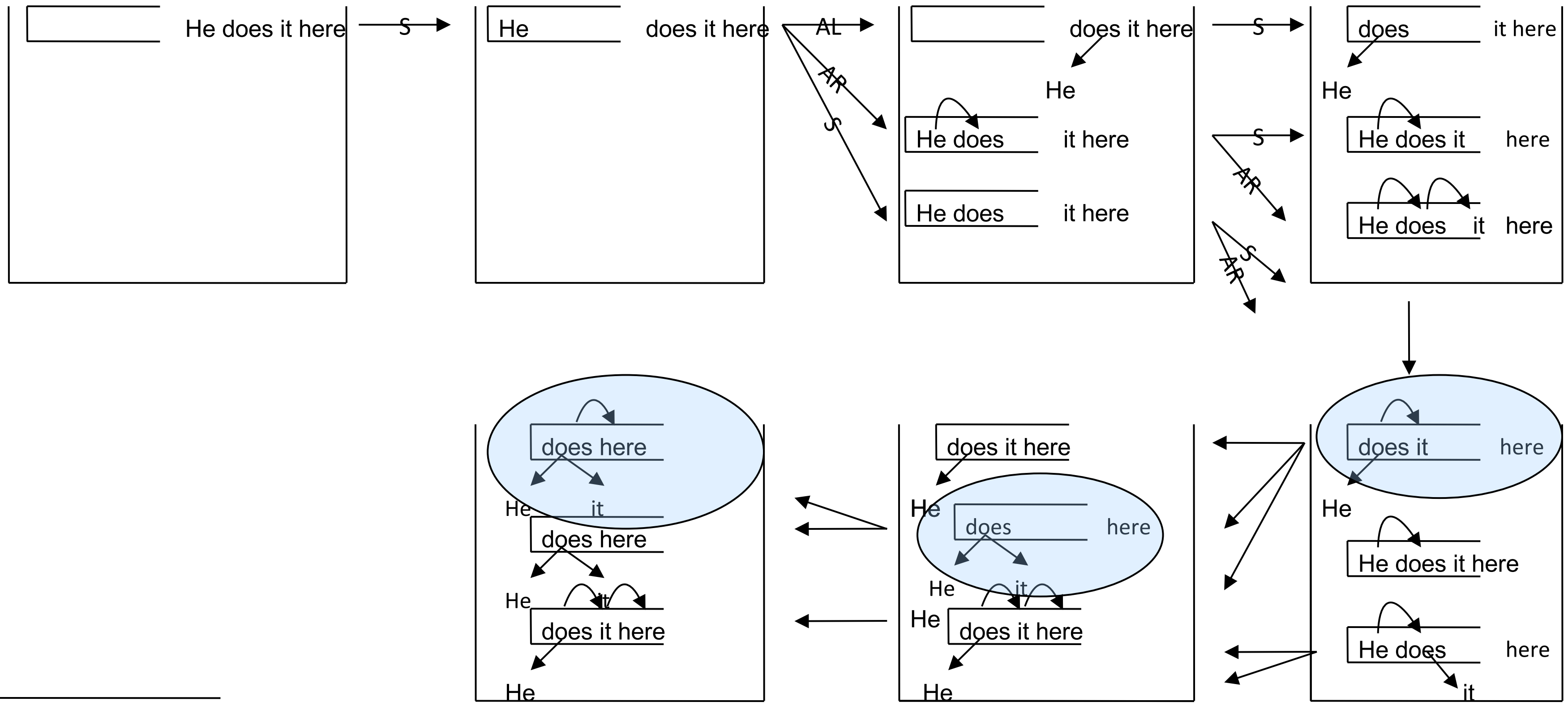
- Dependency Parsing Example
  - Decoding



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

Beam-search decoding

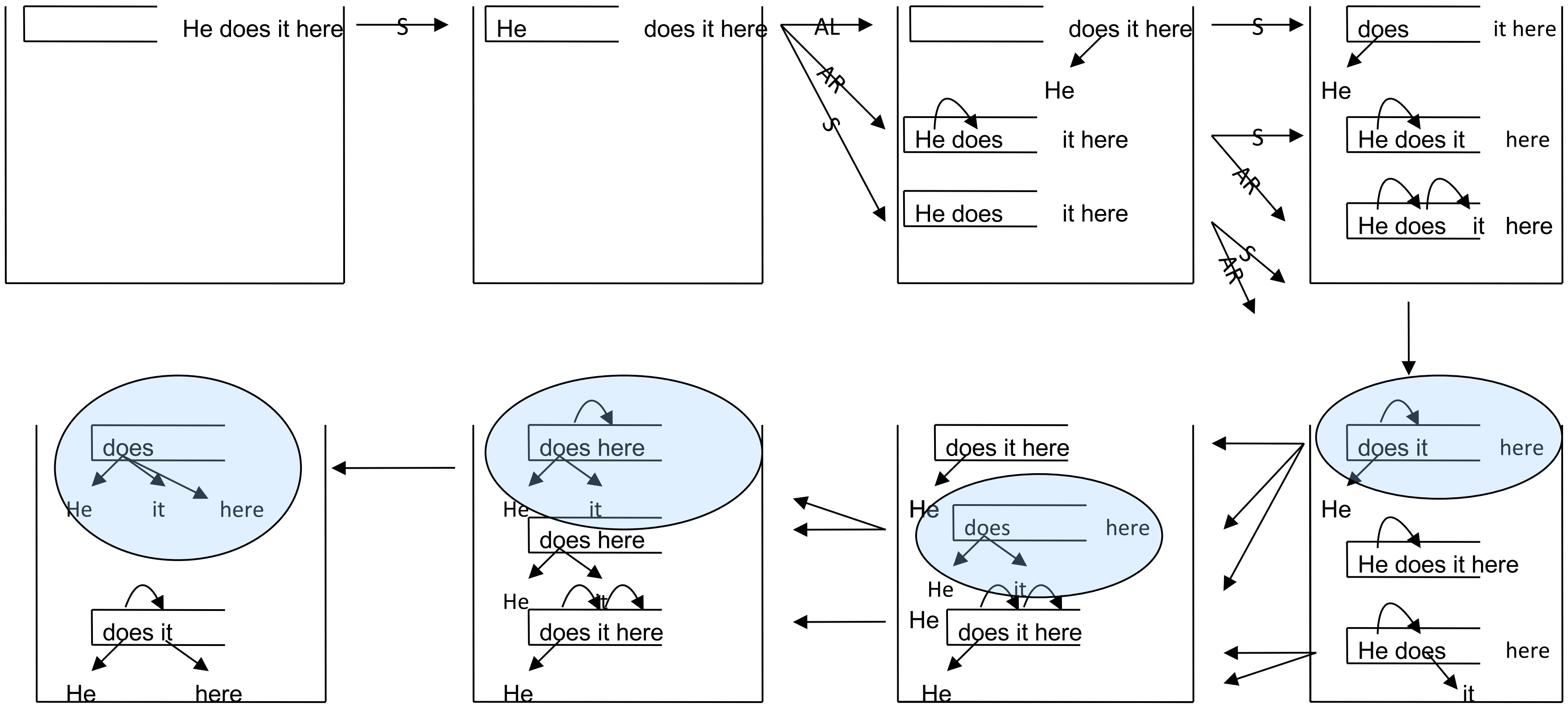
- Dependency Parsing Example
  - Decoding



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

Beam-search decoding

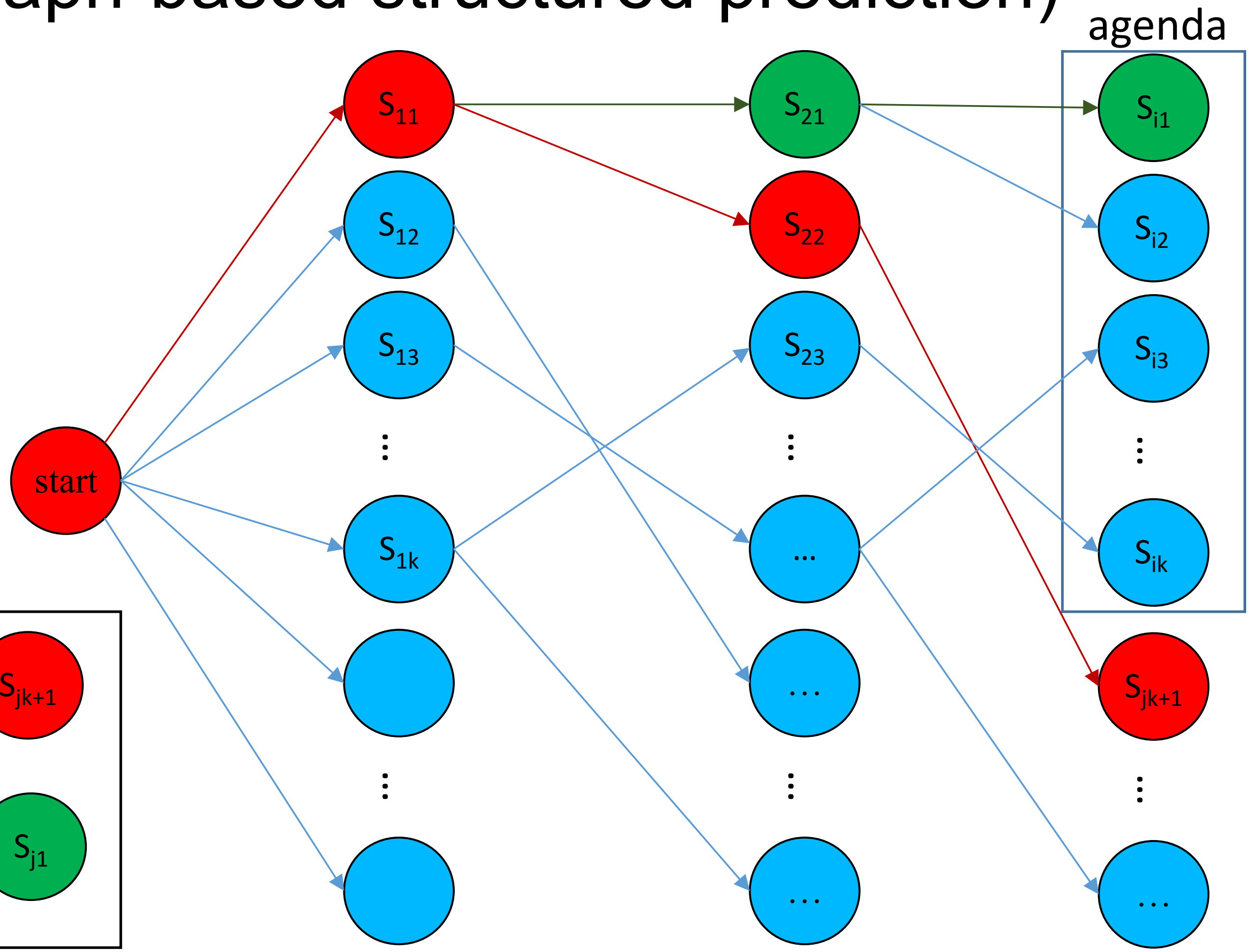
- Dependency Parsing Example
    - Decoding

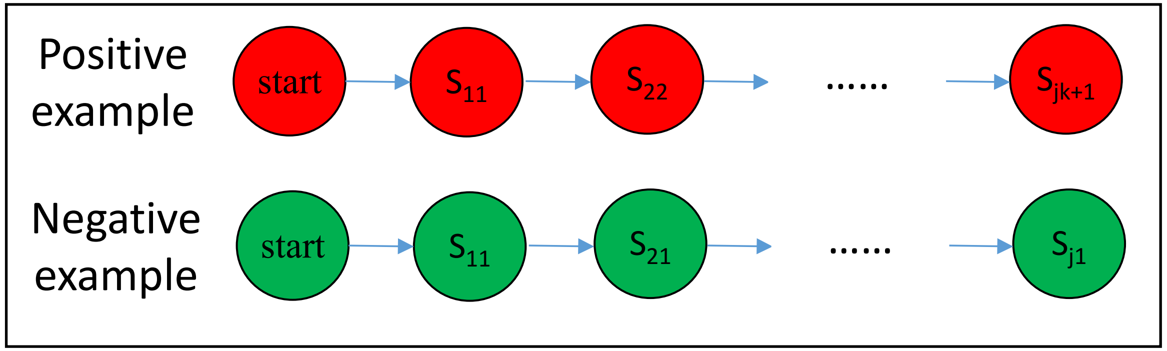


Yue Zhang and Stephen Clark. 2011. *Syntactic Processing Using the Generalized Perceptron and Beam Search*. In *Computational Linguistics*, 37(1), March.

Learning guided search

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





Advantages

- Low computation complexity
- Arbitrary linear features
  - Enabled by learning-guided-search

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

Advantages

- State-of-the-art **accuracies** and **speeds**
  - For a wide range of tasks
- 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.

Joint Segmentation and 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 Tagging

- The transition system
  - Initial state

我喜欢读书

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 Tagging

- The transition system
  - Separate(PN)

我/PN	喜欢读书

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 Tagging

- The transition system
  - Separate (V)

我/PN 喜/V

欢读书

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 Tagging

- The transition system
  - Append

我/PN 喜欢/V

读书

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 Tagging

- The transition system
  - Separate (V)

我/PN　喜欢/V　读/V	书

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 Tagging

- The transition system
  - Separate (N)

我/PN 喜欢/V 读/V 书/N

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 Tagging

- The transition system
  - End state

我/PN 喜欢/V 读/V 书/N

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 Tagging

- 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.

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 Tagging

- 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.

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 Tagging

- Experiments
  - Penn Chinese Treebank 5 (CTB-5)

	CTB files	# sent.	# words
Training	1-270	18089	493,939
	400-1151		
Develop	301-325	350	6,821
Test	271-300	348	8,008

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 Tagging

- Experiments

Accuracy comparisons between various joint segmentors and POS-taggers on CTB5

	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

- The observations lead to the solution of joint segmentation, POS-tagging and 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]

- The chunking knowledge can potentially improve segmentation, this paper explore a joint model that performs segmentation, POS-tagging and chunking simultaneously.
- To address the sparsity of full chunk features, a semi-supervised method is proposed to derive chunk cluster features from large-scale automatically-chunked data.

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

- Word-based chunking example
  - Action: Initial state

stack

queue

他/NR 到达/VV 北京/NR 机场/NN 。/PU

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

- Word-based chunking example
  - Action: SEP(NP)

stack	queue
[NP 他/NR	到达/VV 北京/NR 机场/NN 。/PU

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

- Word-based chunking example
  - Action: SEP(VP)

stack	queue
[NP 他/NR][VP 到达/VV	北京/NR 机场/NN 。/PU

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

- Word-based chunking example
  - Action: SEP(NP)

stack

[NP 他/NR][VP 到达/VV]
[NP 北京/NR]

queue

机场/NN 。/PU

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

- Word-based chunking example
  - Action: APP(NP)

stack

[NP 他/NR][VP 到达/VV]
[NP 北京/NR 机场/NN

queue

。/PU

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

- Word-based chunking example
  - Action: SEP(O)

stack

[NP 他/NR][VP 到达/VV]
[NP 北京/NR 机场/NN]
[O 。/PU]

queue

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

- Word-based chunking feature template

ID	**Feature Templates**
1	N_0w
2	N_0t
3	N_1w
4	N_1t
5	N_2w
6	N_2t
7	$N_0w \cdot N_0t$
8	$N_1w \cdot N_1t$
9	$N_2w \cdot N_2t$
10	$N_0w \cdot N_1w$
11	$N_0w \cdot N_1t$
12	$N_0t \cdot N_1w$
13	$N_0w \cdot N_1w \cdot N_0t$
14	$N_0w \cdot N_1w \cdot N_1t$
15	$N_1w \cdot N_2w$
16	$N_1w \cdot N_2t$
17	$N_1t \cdot N_2w$
18	$N_1t \cdot N_2t$
19	$w_1 \cdot N_0 \cdot T_0$, where $len(C_0) = 1$
20	$start\_word(C_0)T_0$
21	$start\_POS(C_0)T_0$
22	$end\_word(C_0)T_0$
23	$end\_POS(C_0)T_0$
24	$w \cdot end\_word(C_0) \cdot T_0$ where $w \in C_0$ and $w \neq end\_word(C_0)$
25	$t \cdot end\_POS(C_0) \cdot T_0$ where $t \in POSset(C_0)$ and $p \neq end\_POS(C_0)$
26	$w \cdot label(w) \cdot T_0$ for all w in C_0
27	$bigram(w) \cdot label(w) \cdot T_0$ for all w in C_0
28	$biPOS(w) \cdot label(w) \cdot T_0$ for all w in C_0
29	$POSset(C_0) \cdot T_0$
30	$T_0 \cdot T_{-1}$
31	$end\_word(C_{-1}) \cdot T_{-1} \cdot start\_word(C_0) \cdot T_0$
32	$end\_word(C_{-1}) \cdot T_{-1} \cdot end\_word(C_0) \cdot T_0$
33	$start\_word(C_{-1}) \cdot T_{-1} \cdot start\_word(C_0) \cdot T_0$
34	$end\_POS(C_{-1}) \cdot T_{-1} \cdot start\_POS(C_0) \cdot T_0$
35	$end\_POS(C_{-1}) \cdot T_{-1} \cdot end\_POS(C_0) \cdot T_0$
36	$start\_POS(C_{-1}) \cdot T_{-1} \cdot start\_POS(C_0) \cdot T_0$
37	$end\_word(C_{-1}) \cdot T_0$; $end\_POS(C_{-1}) \cdot T_0$
38	$T_{-1} \cdot T_0 \cdot start\_word(C_0)$
39	$T_{-1} \cdot T_0 \cdot start\_POS(C_0)$
40	$POSset(C_{-1}) \cdot T_{-1} \cdot POSset(C_0) \cdot T_0$

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

stack	deque	queue
		他 到 达 北 京 机 场 。

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)

stack	deque	queue
	[他/NR	到 达 北 京 机 场 。

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

stack	deque	queue
	[他/NR]	到 达 北 京 机 场 。

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)

stack	deque	queue
[NP 他/NR]		到 达 北 京 机 场 。

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)

stack	deque	queue
[NP 他/NR]	[到/VV	达 北 京 机 场 。

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

stack	deque	queue
[NP 他/NR]	[到达/VV	北 京 机 场 。

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

stack	deque	queue
[NP 他/NR]	[到达/VV]	北 京 机 场 。

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)

stack	deque	queue
[NP 他/NR] [VP 到达/VV]		北 京 机 场 。

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)

stack	deque	queue
[NP 他/NR] [VP 到达/VV]	[北/NR	京 机 场 。

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

stack	deque	queue
[NP 他/NR] [VP 到达/VV]	[北京/NR	机 场 。

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

stack	deque	queue
[NP 他/NR] [VP 到达/VV]	[北京/NR]	机 场 。

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)

stack	deque	queue
[NP 他/NR] [VP 到达/VV] [NP 北京/NR]		机 场 。

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)

stack	deque	queue
[NP 他/NR] [VP 到达/VV] [NP 北京/NR]	[机/NN	场 。

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

stack	deque	queue
[NP 他/NR] [VP 到达/VV] [NP 北京/NR]	[机场/NN	。

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

stack

[NP 他/NR]
[VP 到达/VV]
[NP 北京/NR]

deque

[机场/NN]

queue

。

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

stack	deque	queue
[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN]		。

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)

stack	deque	queue
[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN]	[。/PU	

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

stack	deque	queue
[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN]	[。/PU]	

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)

stack	deque	queue
[NP 他/NR] [VP 到达/VV] [NP 北京/NR 机场/NN] [O 。/PU]		

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

- Statistics of the CTB4 corpus

	Sections	Sentences	Words
Training	1-300 326-899	9,528	232,085
Dev	301-325	350	6,821
Test	900-1078	5,290	165,862

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 of word-based chunking

Method	**CHUNK**
CRFs	90.74
SVMs	91.46
Chen, Zhang and Isahara (2006)	91.68
Zhou, Qu and Zhang (2012)	92.11
Our Baseline	91.43
Pipeline	69.02

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 of semi-supervised models

	SEG	**POS**	**CHUNK**
Supervised	89.85	81.94	70.96
Semi-ALL	**91.00**	**82.71**	**72.29**
Semi-C	90.67	82.45	72.09
Semi-C_0	90.71	82.59	71.98
Semi-W	90.72	82.53	71.62

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

- Comparison between the pipeline and joint models

	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”.
- This paper proposed a transition-based model for joint word segmentation, POS tagging and text normalization.

工作鸭梨大啊！ $\Rightarrow$ 工作/NN 压力/NN 大/VA 啊/SP !/PU

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

Sentence: 工作鸭梨大啊! (How great work pressure is!)

State	Action	Stack	Queue	Dictionary
S_i	......	Org: 工作 鸭梨 work pear Nor: 工作 work	大啊! big ah!	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 ....
S_{i+1}	APP("大")	Org: 工作 鸭梨大 work pear big Nor: 工作 work	啊! (ah!)	
	SEP("大")	Org: 工作 鸭梨 大 work pear big Nor: 工作 work		
	SEPS("大", "压力")	Org: 工作 鸭梨 大 work pear big Nor: 工作 压力 work pressure		

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

stack

queue

工 作 鸭 梨 大 啊 ！

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)

stack

工/NN

queue

作 鸭 梨 大 啊 !

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)

stack

工作/NN

queue

鸭 梨 大 啊 ！

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)

stack

工作/NN 鸭/NN

queue

梨 大 啊 ！

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)

stack	queue
工作/NN 鸭梨/NN	大 啊 ！

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)

stack

工作/NN 压力/NN 大/VA

queue

啊 ！

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)

stack	queue
工作/NN 压力/NN 大/VA 啊/SP	！

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)

stack

工作/NN 压力/NN 大/VA 啊/SP ！/PU

queue

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

- Normalization dictionary
- Using CTB data.

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

	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.

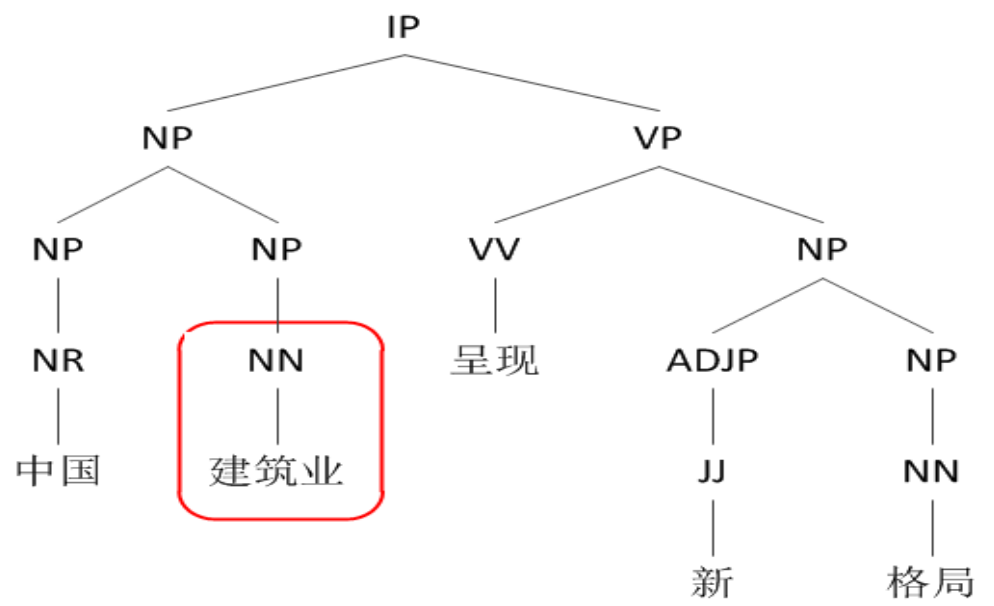
Joint Segmentation, POS-tagging and Constituent Parsing

- This paper investigate Chinese parsing from the character-level, extending the notion of phrase-structure trees by annotating internal structures of words.

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

- Traditional: word-based Chinese parsing

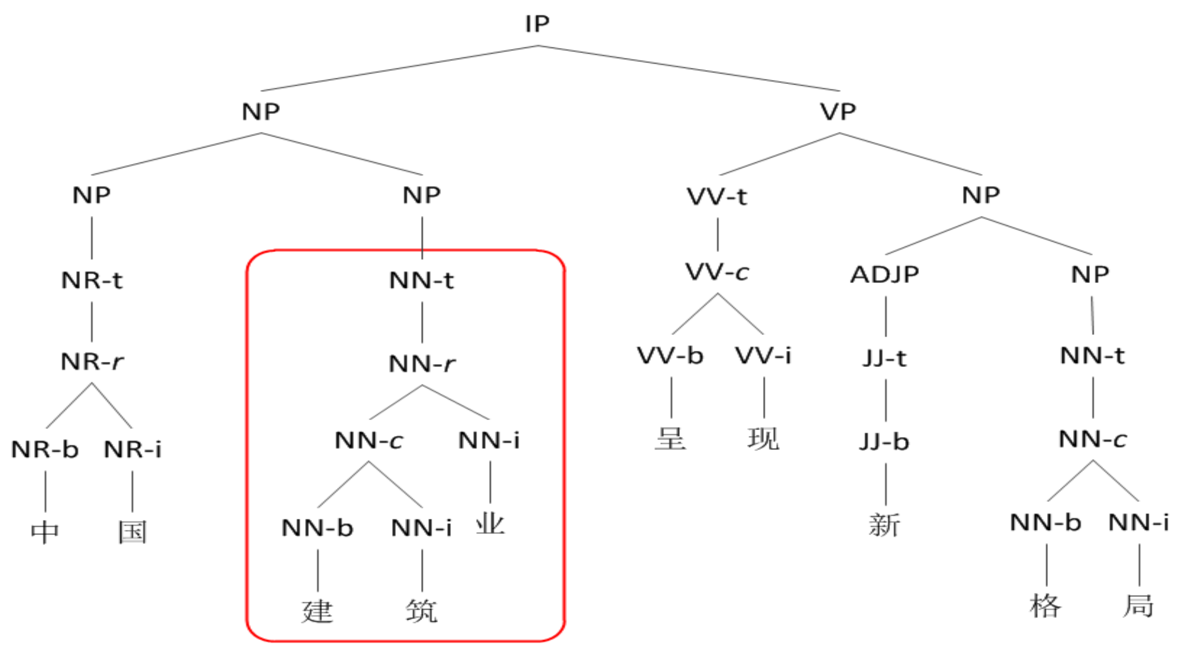


CTB-style word-based syntax tree for “中国 (China) 建筑业 (architecture industry) 呈现 (show) 新 (new) 格局 (pattern)”.

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

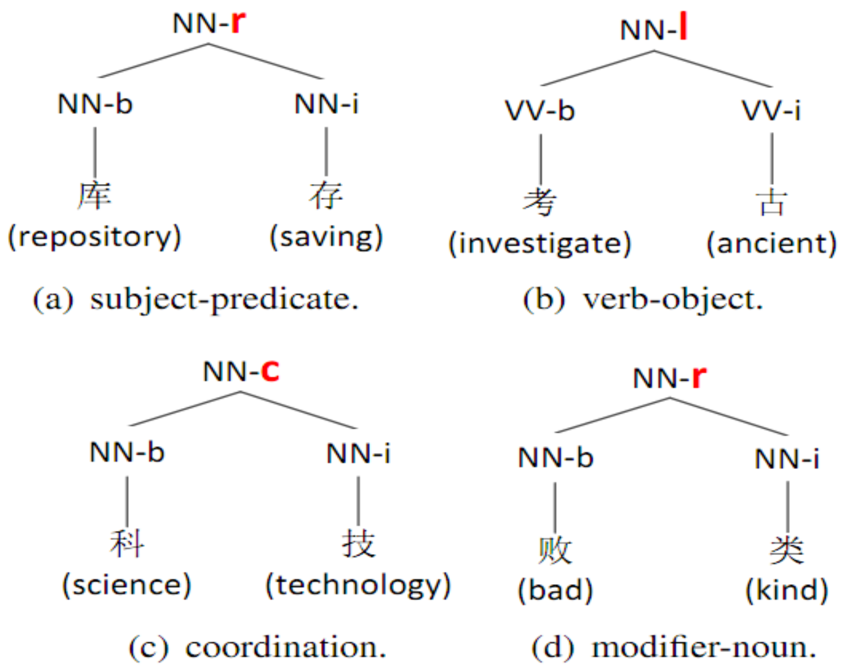


Character-level syntax tree with hierarchal word structures for “中 (middle) 国 (nation) 建 (construction) 筑 (building) 业 (industry) 呈 (present) 现 (show) 新 (new) 格 (style) 局 (situation)”.

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

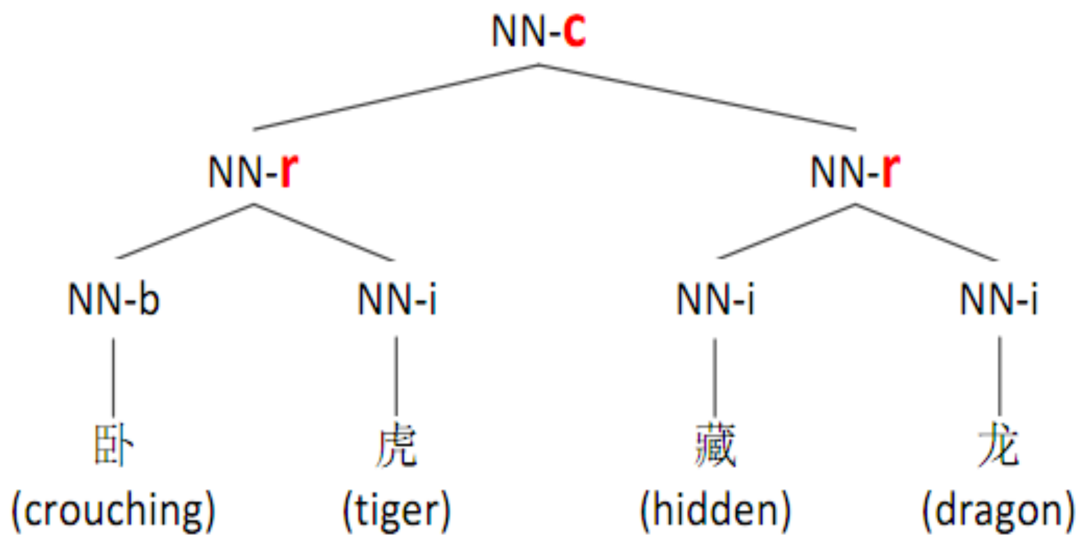
- Why character-based?
  - Chinese words have syntactic structures.



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

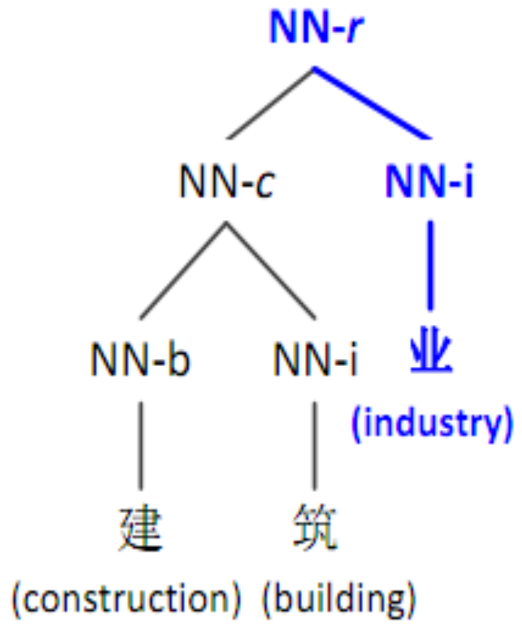
- Why character-based?
  - Chinese words have syntactic structures.



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

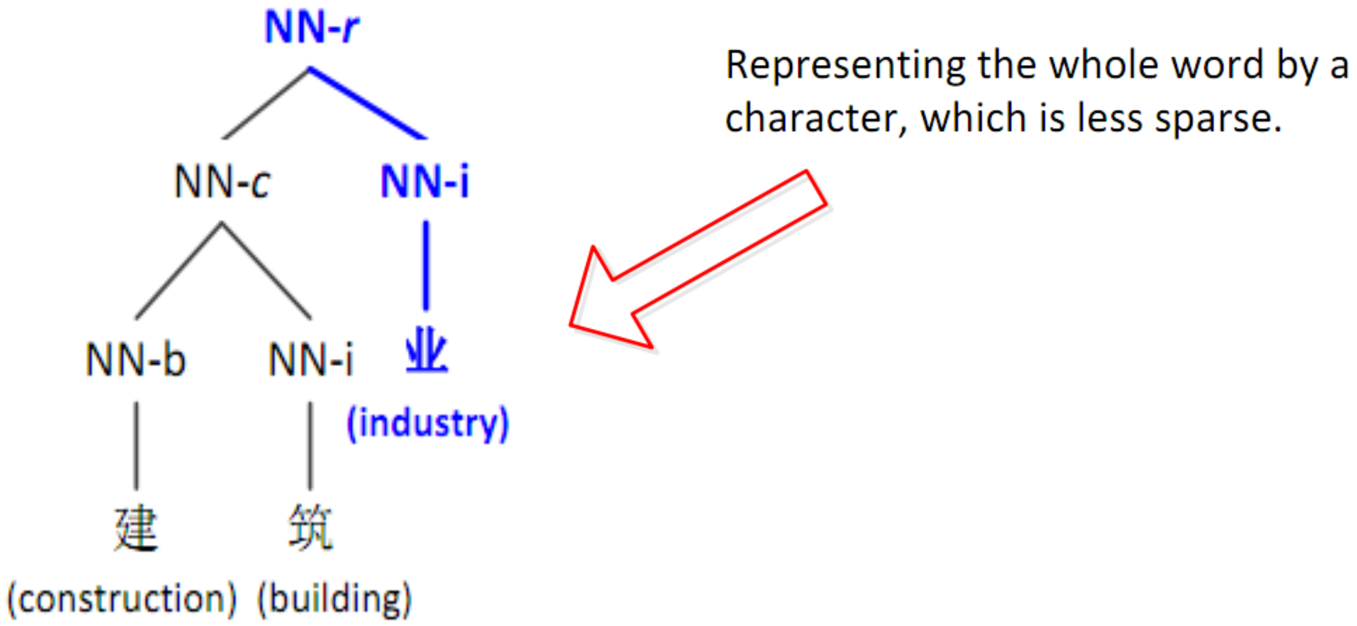
- Why character-based?
  - Deep character information of word structures.



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.

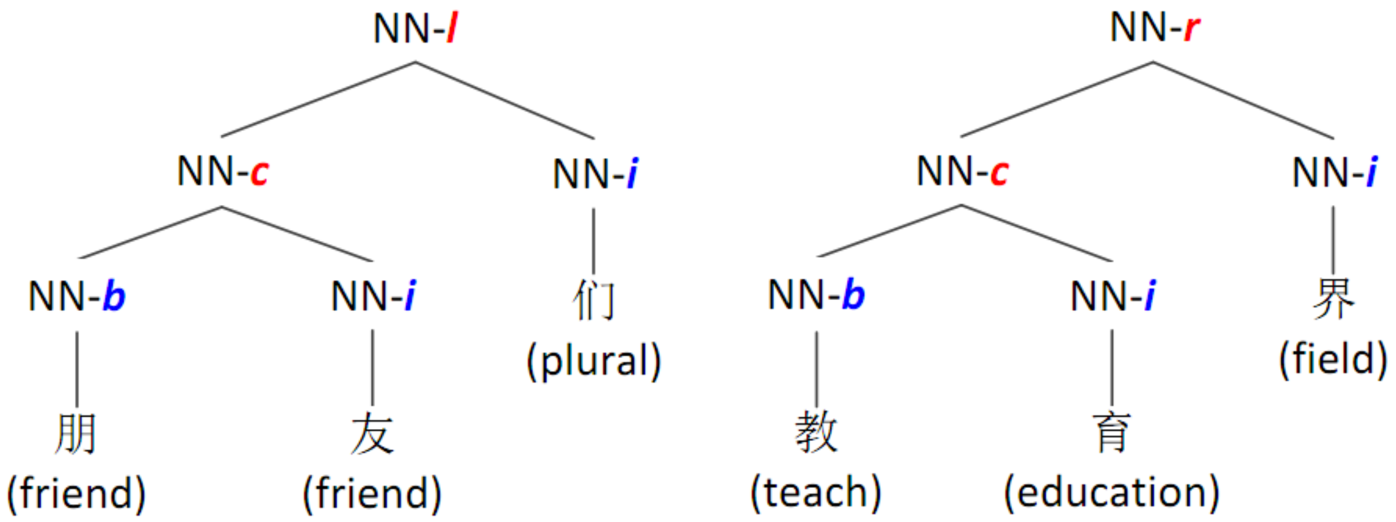
Joint Segmentation, POS-tagging and Constituent Parsing

- Why character-based?
  - Build syntax tree from character sequences.
    - Not require segmentation or POS-tagging as input.
    - Benefit from joint framework, avoid error propagation.

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

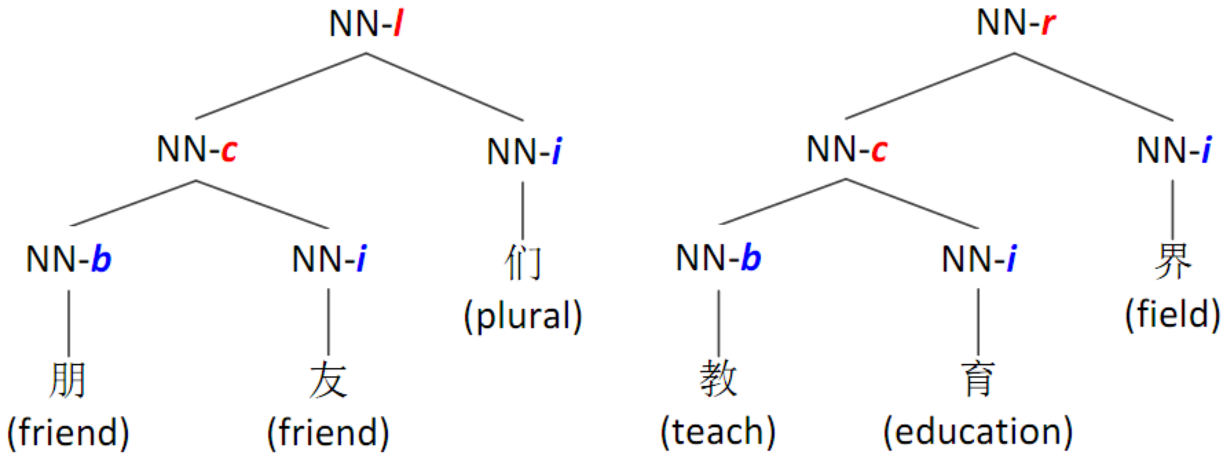
- Word structure annotation
  - Binarized tree structure for each word.



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

- Word structure annotation
  - Binarized tree structure for each word.

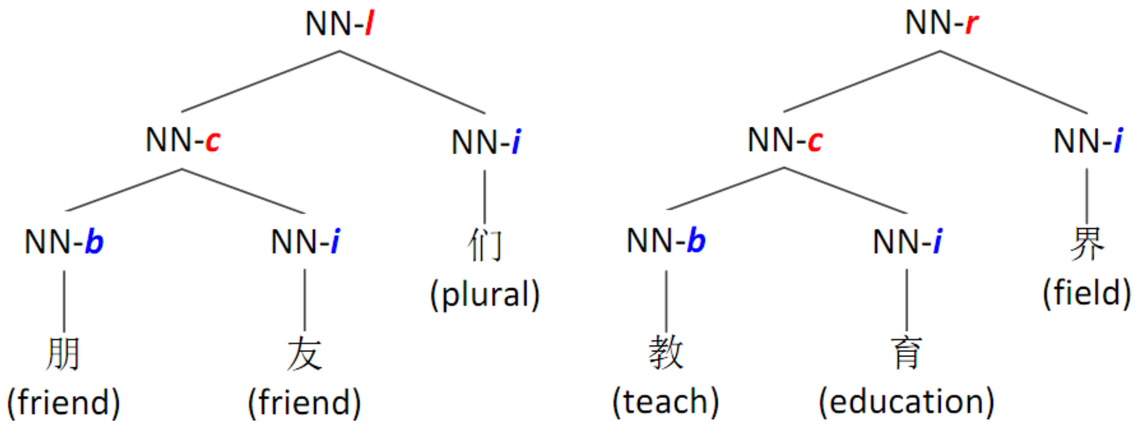


- **b, i denote whether the below character is at a word's beginning position.**
- **l, r, c denote the head direction of current node, respectively left, right and coordination.**

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

- Word structure annotation
  - Binarized tree structure for each word.



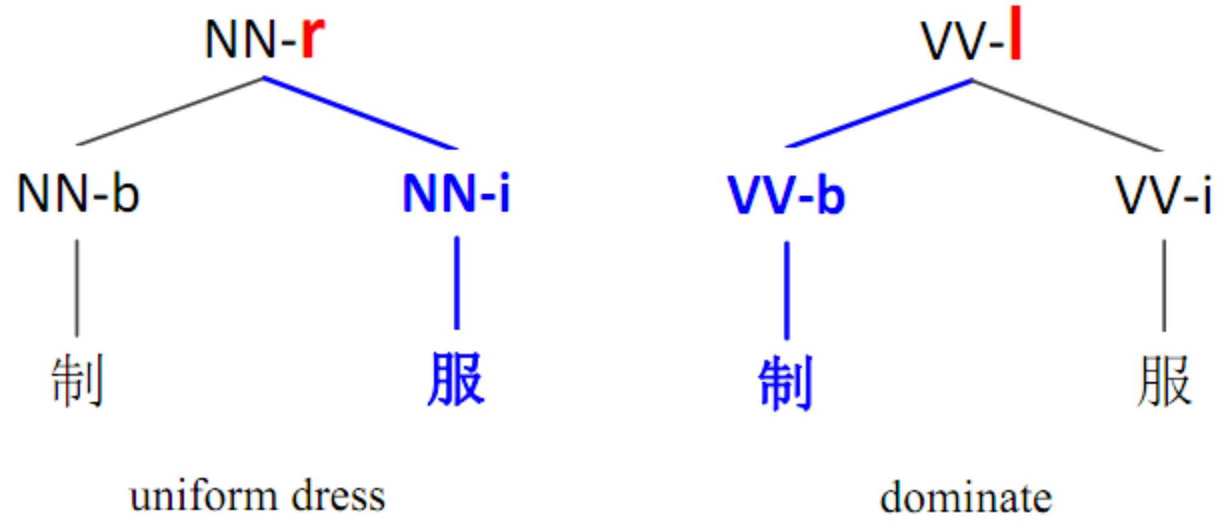
- **b, i denote whether the below character is at a word's beginning position.**
- **l, r, c denote the head direction of current node, respectively left, right and coordination.**

We extend word-based phrase-structures into character-based syntax trees using the word structures demonstrated above.

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

- Word structure annotation
  - Annotation input: a word and its POS.
    - A word may have different structures according to different POS.



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.

Joint Segmentation, POS-tagging and Constituent Parsing

- The character-based parsing model
  - A transition-based parser
    - Extended from Zhang and Clark (2009), a word-based transition parser.

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
        - Extended from Zhang and Clark (2009), a word-based transition parser.
    - Incorporating features of a word-based parser as well as a joint SEG&POS system.

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

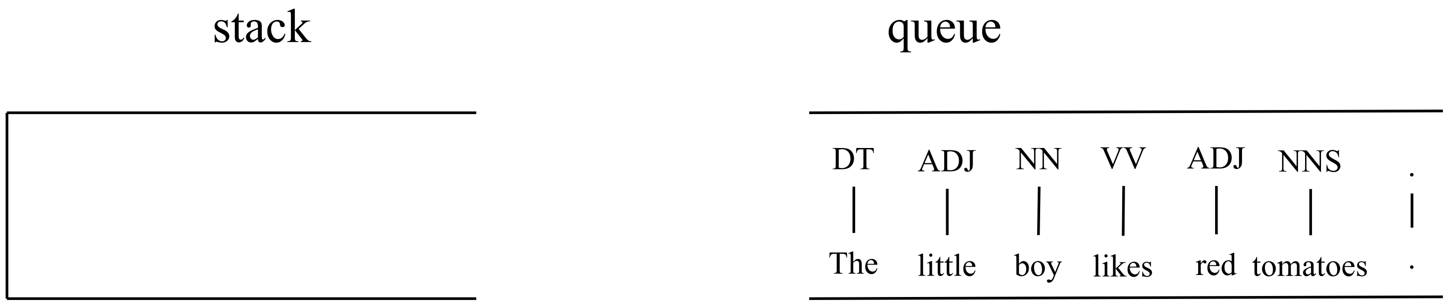
Joint Segmentation, POS-tagging and Constituent Parsing

- The character-based parsing model
  - A transition-based parser
    - Extended from Zhang and Clark (2009), a word-based transition parser.
  - Incorporating features of a word-based parser as well as a joint SEG&POS system.
  - Adding the deep character information from word structures.

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.

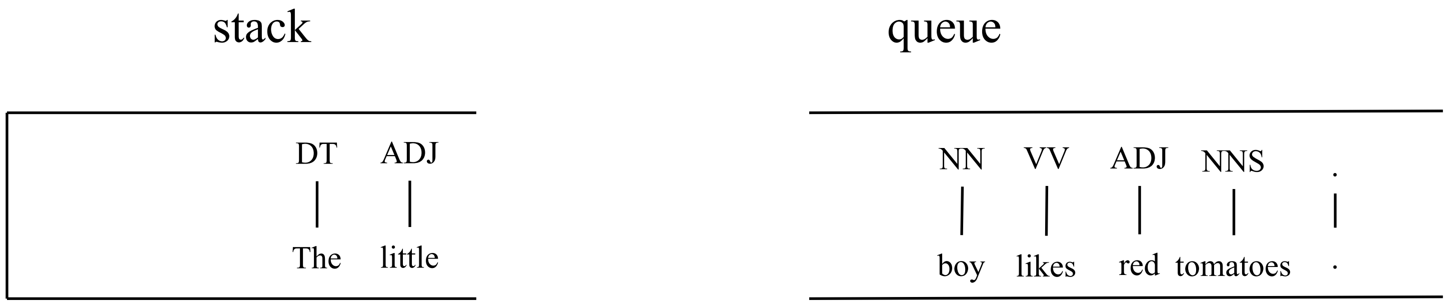
Transition-based Constituent Parsing

- 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

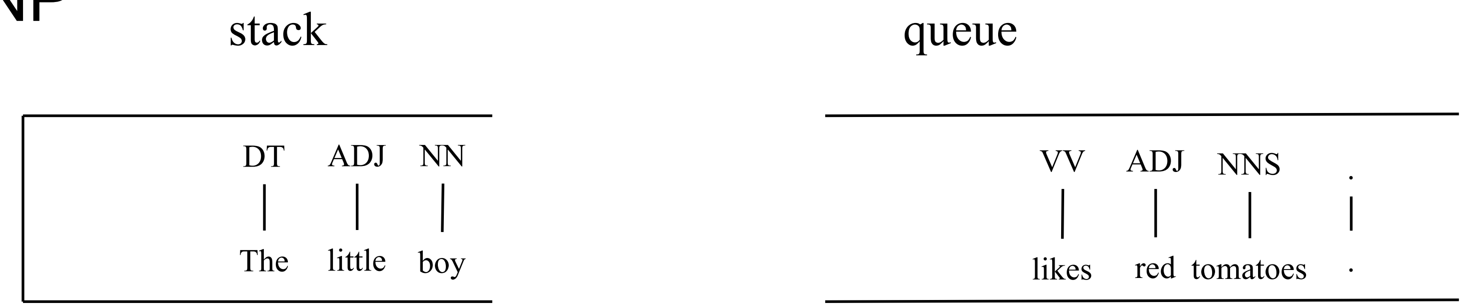
- Example
  - SHIFT



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

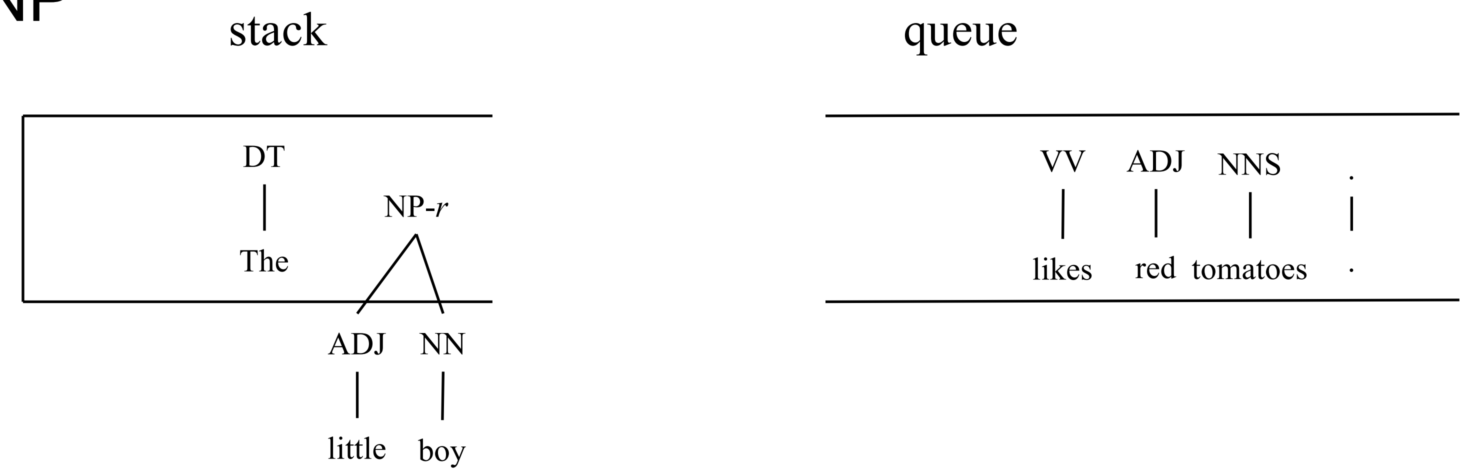
- Example
  - REDUCE-R-NP



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

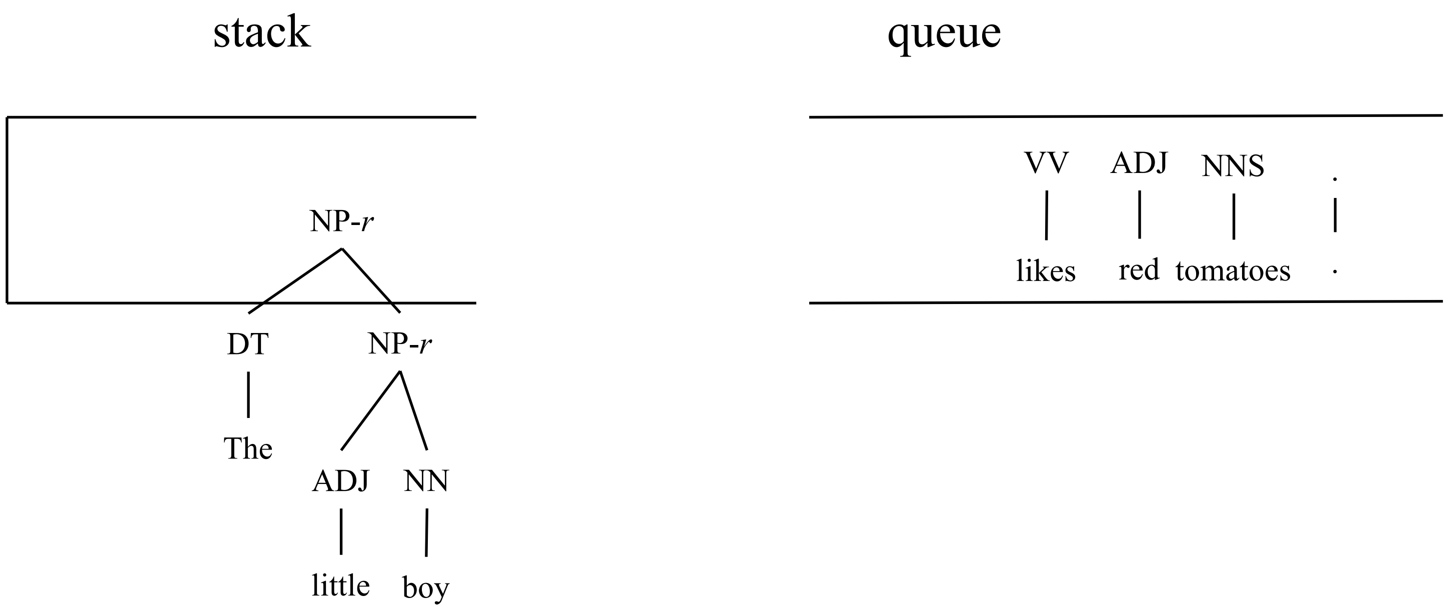
- Example
  - REDUCE-R-NP



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

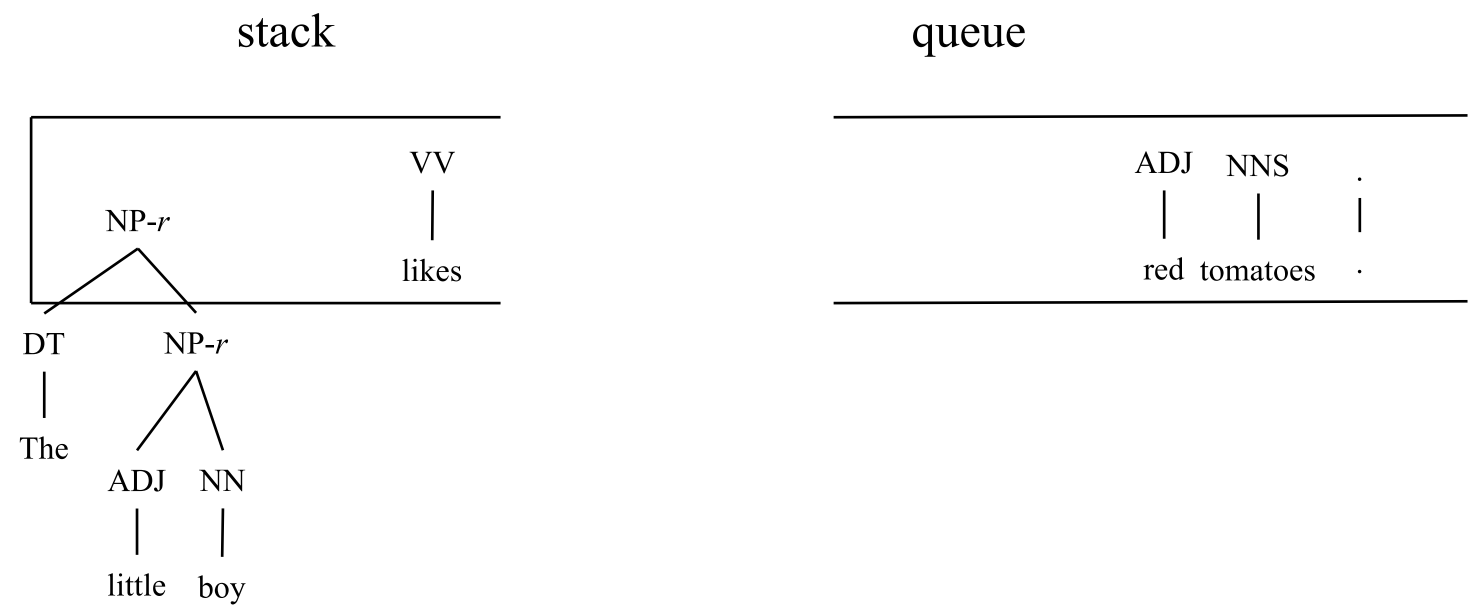
- Example
  - SHIFT



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

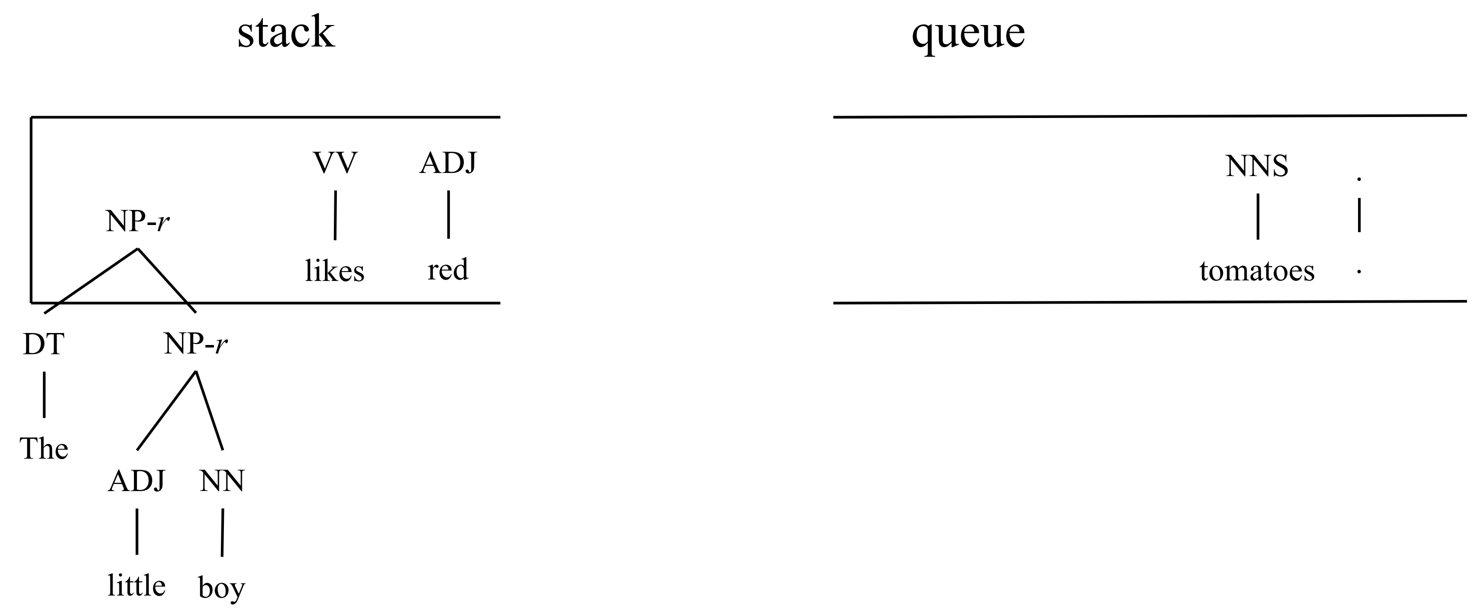
- Example
  - SHIFT



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

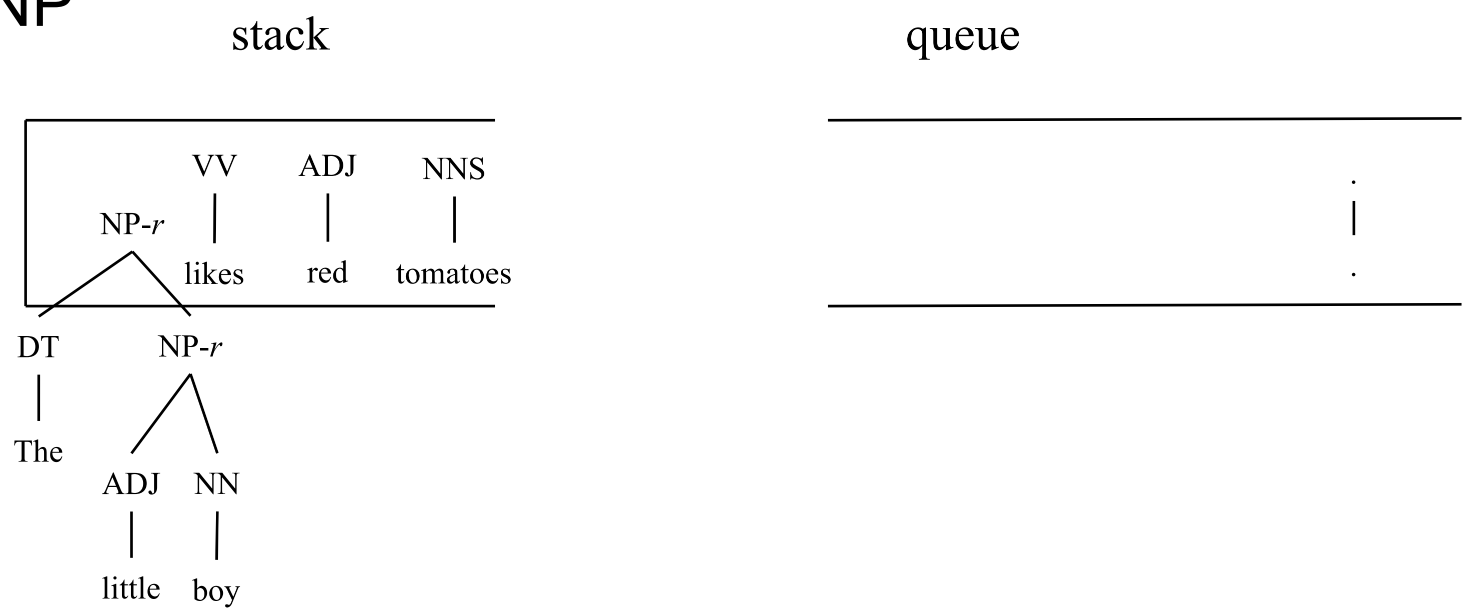
- Example
  - SHIFT



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

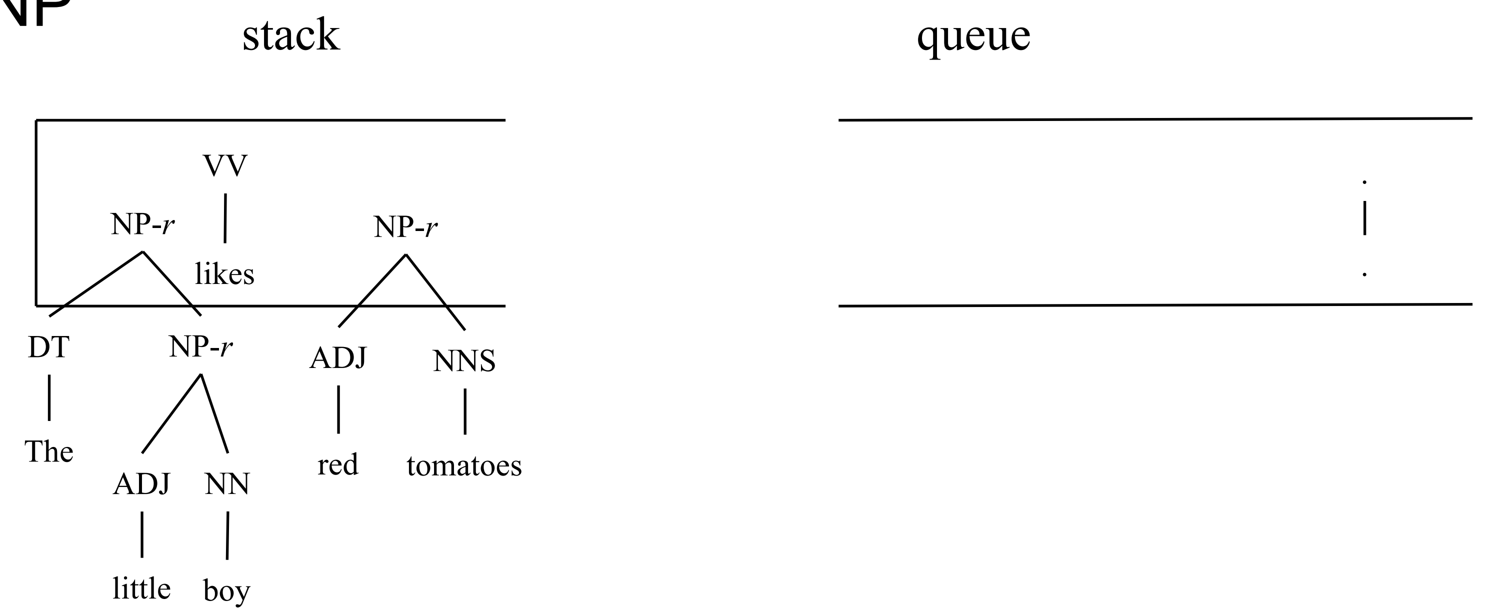
- Example
  - REDUCE-R-NP



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

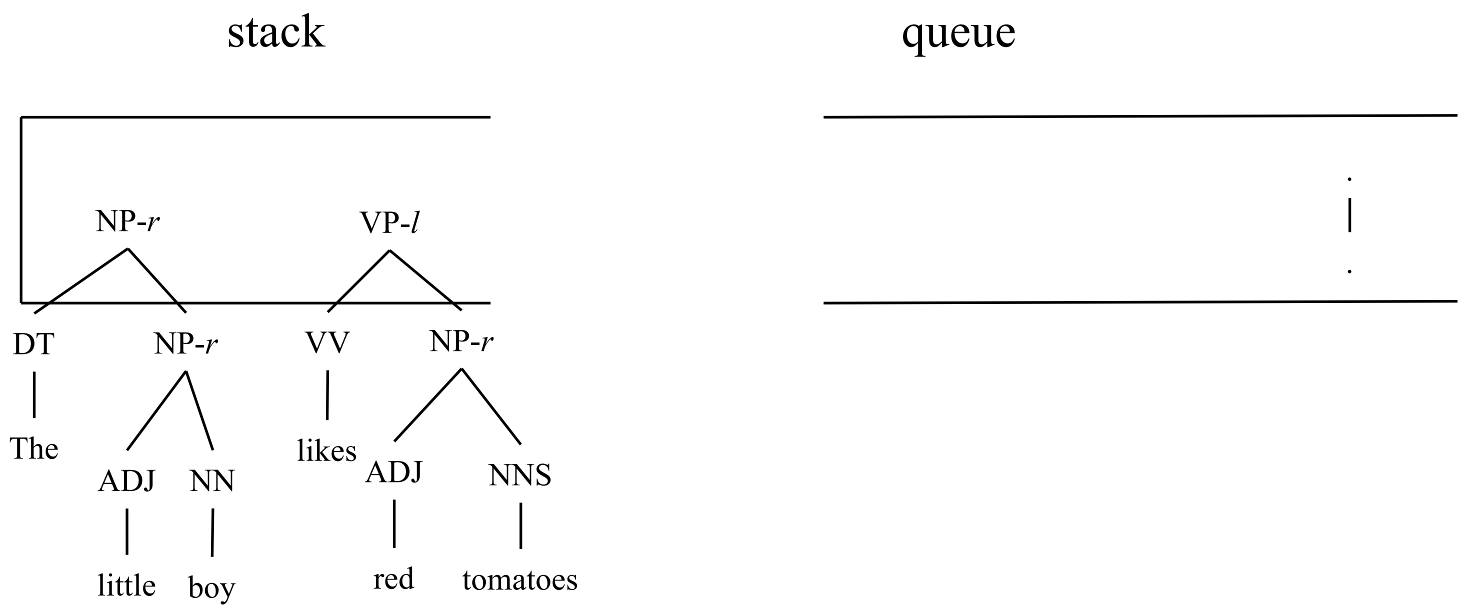
- Example
  - REDUCE-L-NP



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

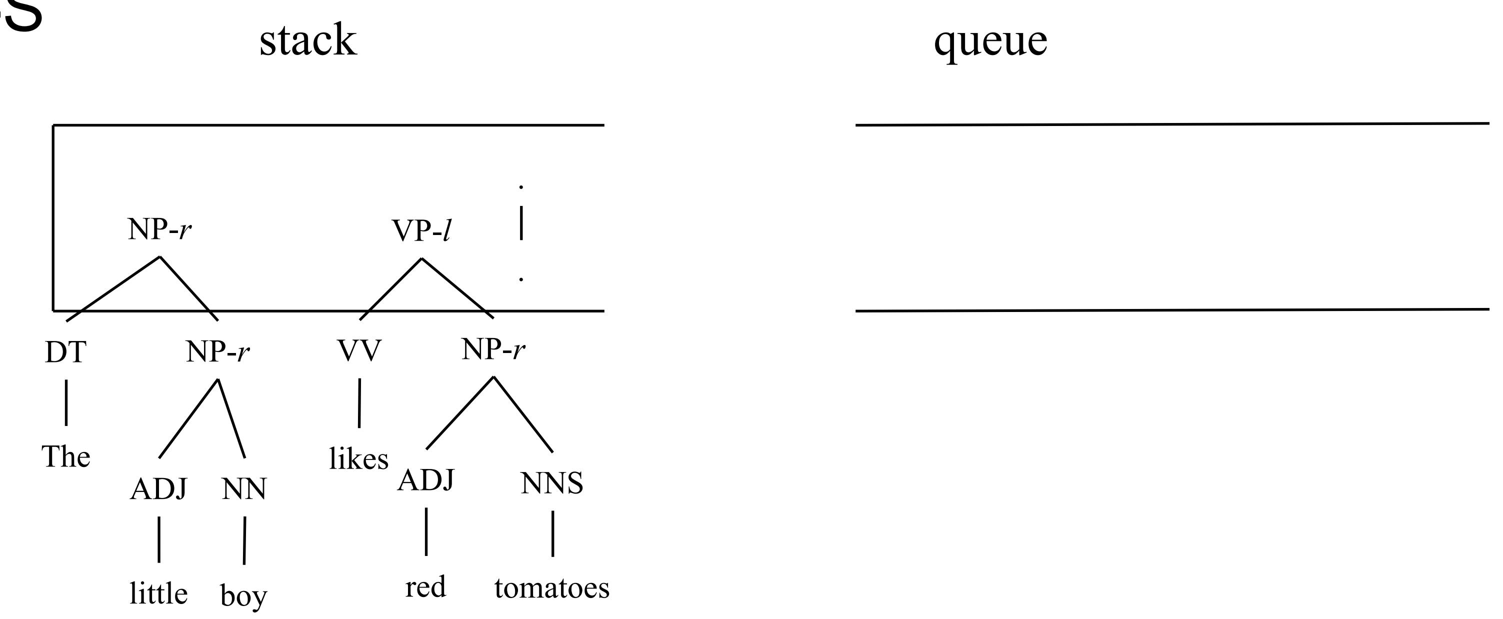
- Example
  - SHIFT



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

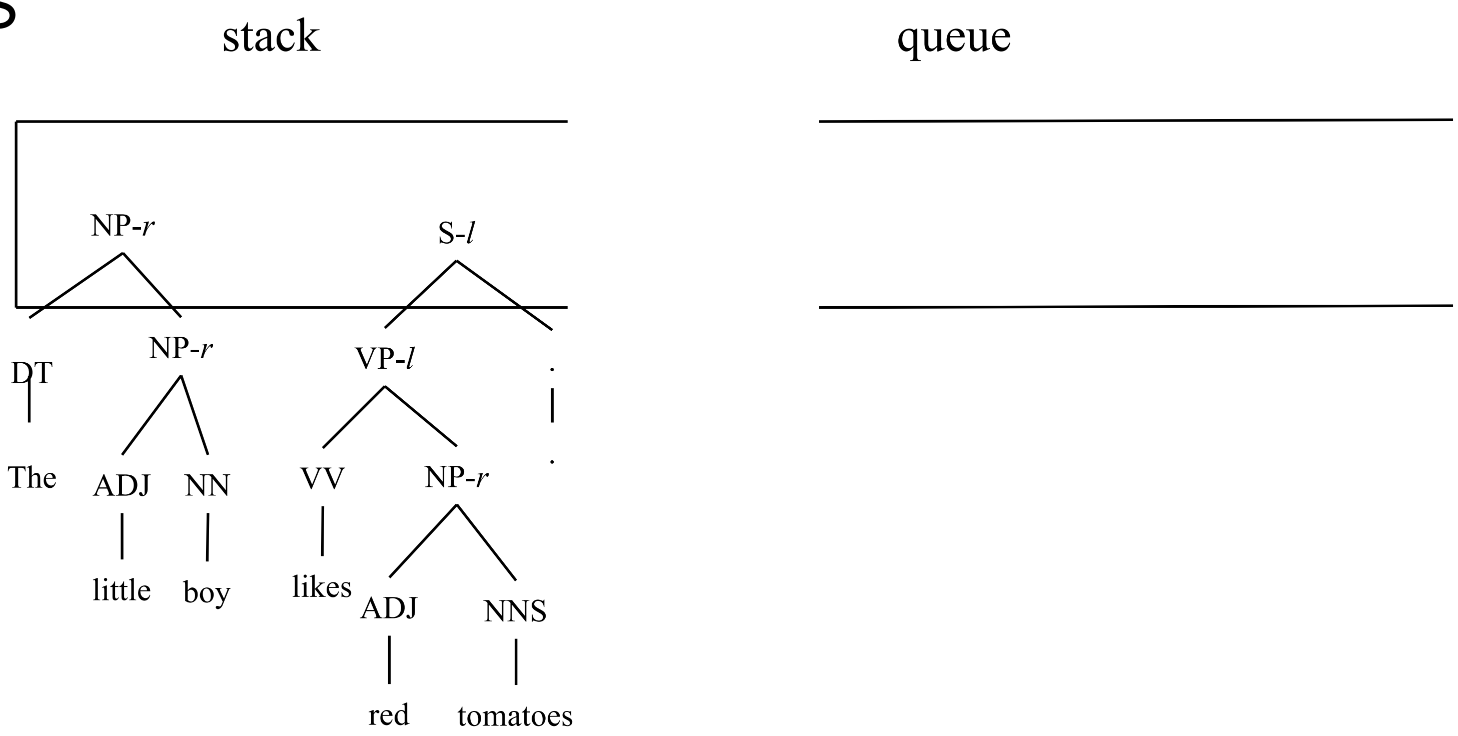
- Example
  - REDUCE-L-S



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

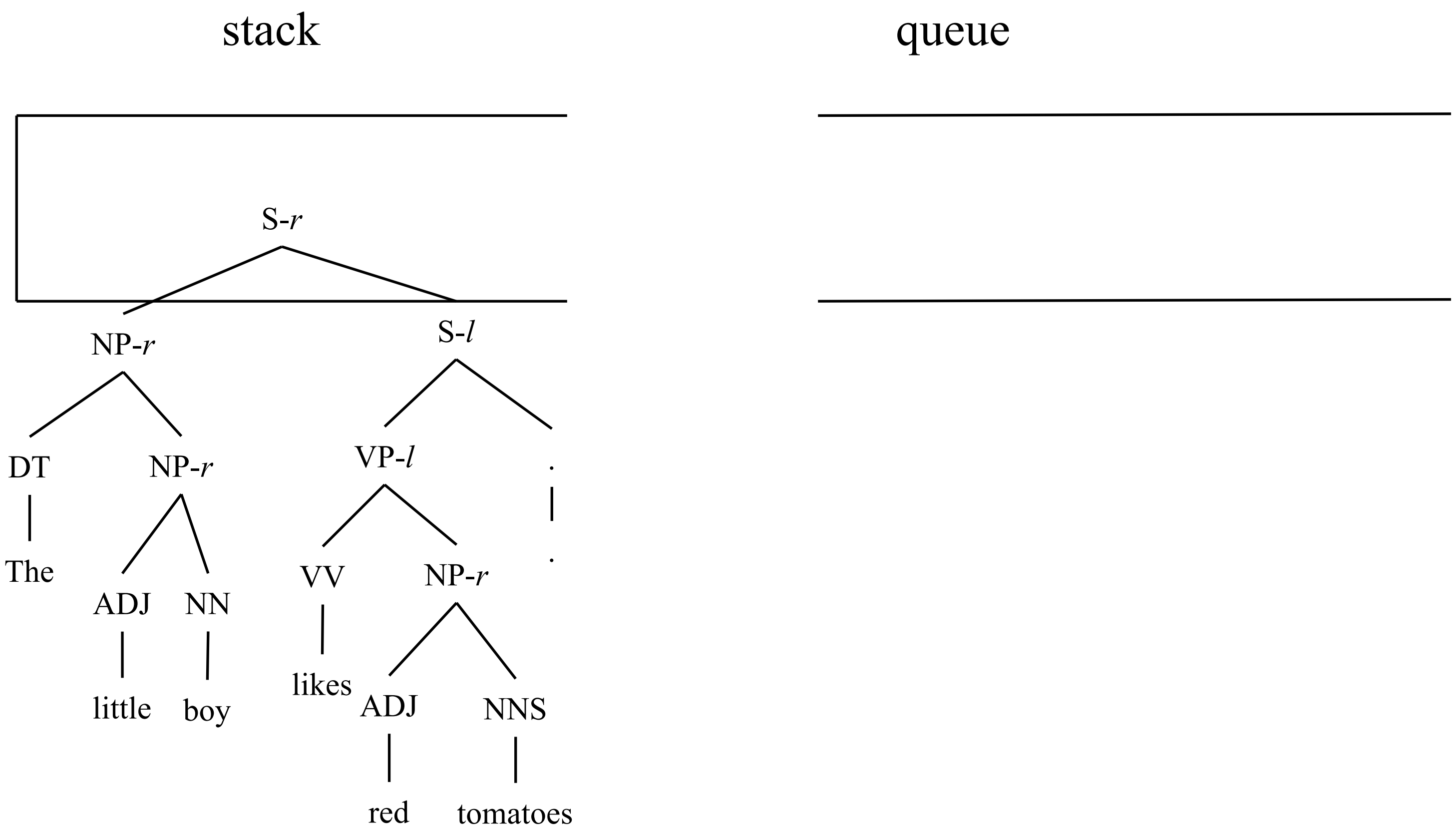
- Example
  - REDUCE-R-S



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

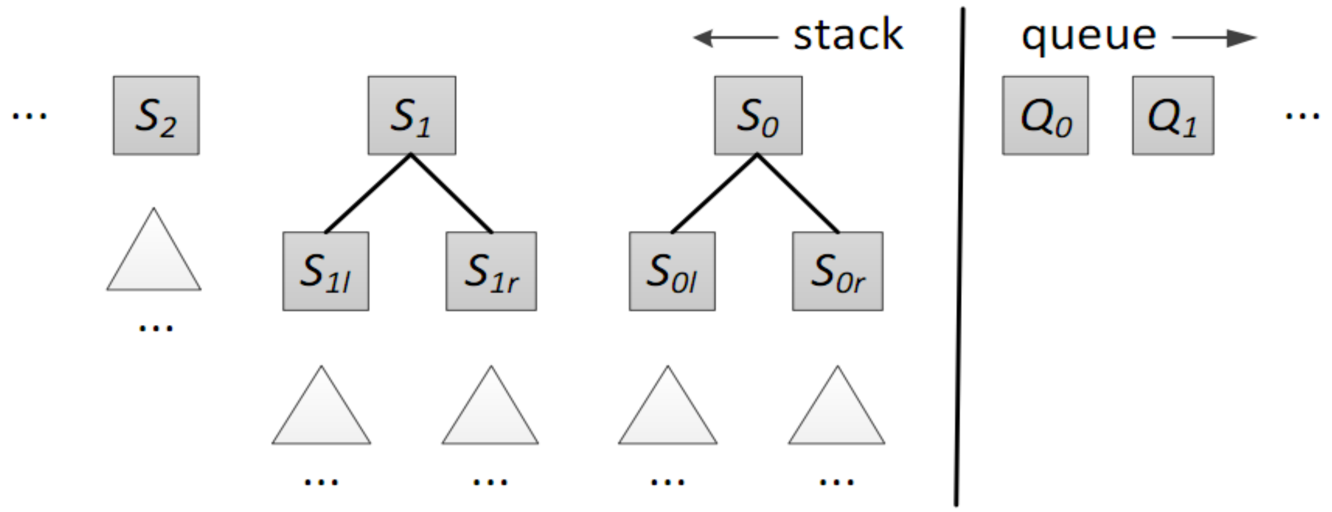
- Example
  - TERMINATE



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:

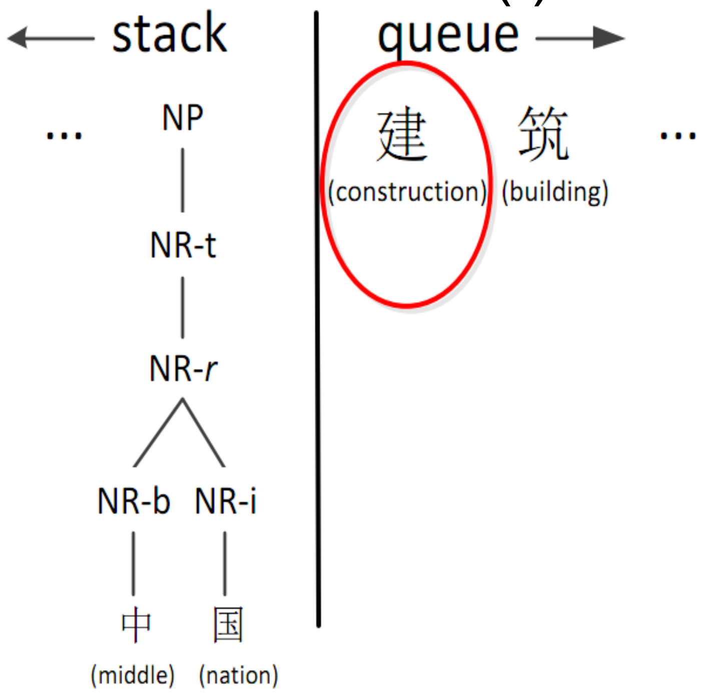


  - Actions:
    - SHIFT-SEPARATE(***t***), SHIFT-APPEND, REDUCE-SUBWORD(***d***), REDUCE-WORD, REDUCE-BINARY(***d***;***l***), REDUCE-UNARY(***l***), TERMINATE

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

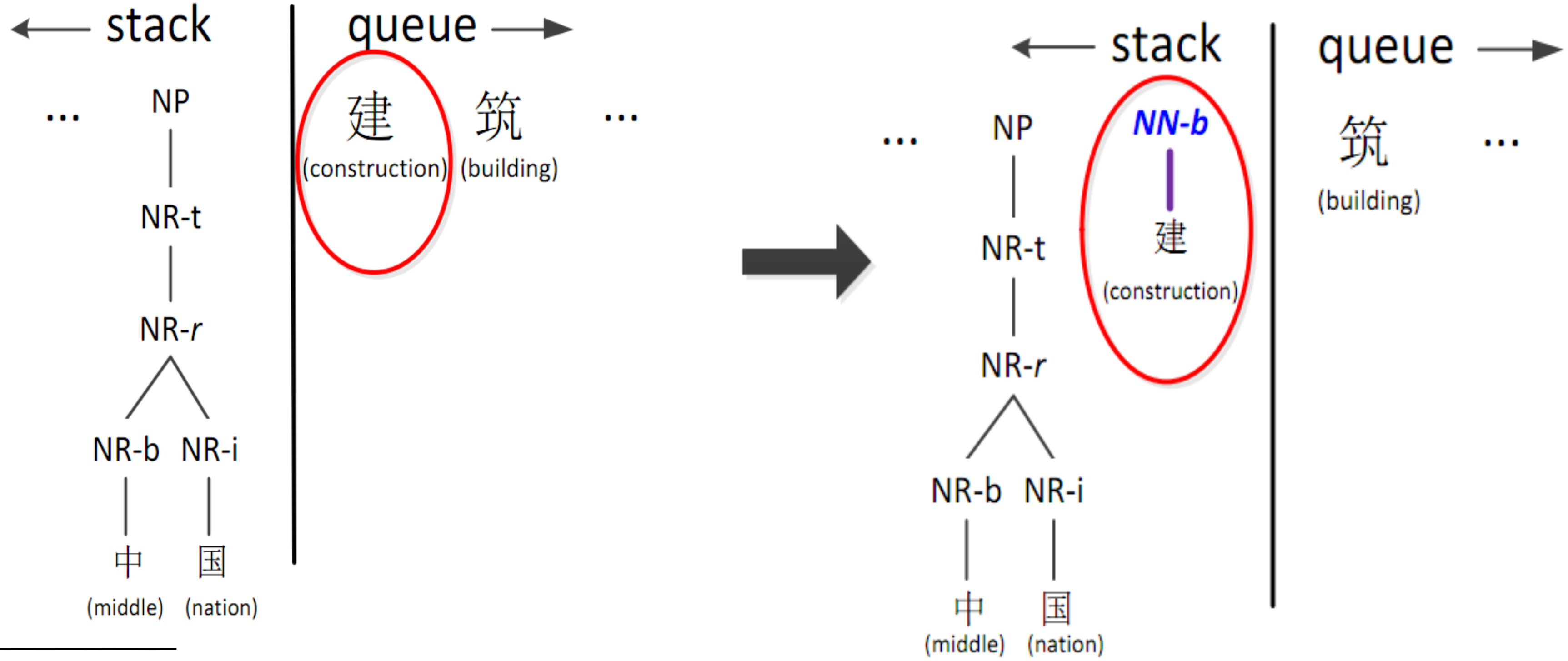
- Actions
  - SHIFT-SEPARATE(t)



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

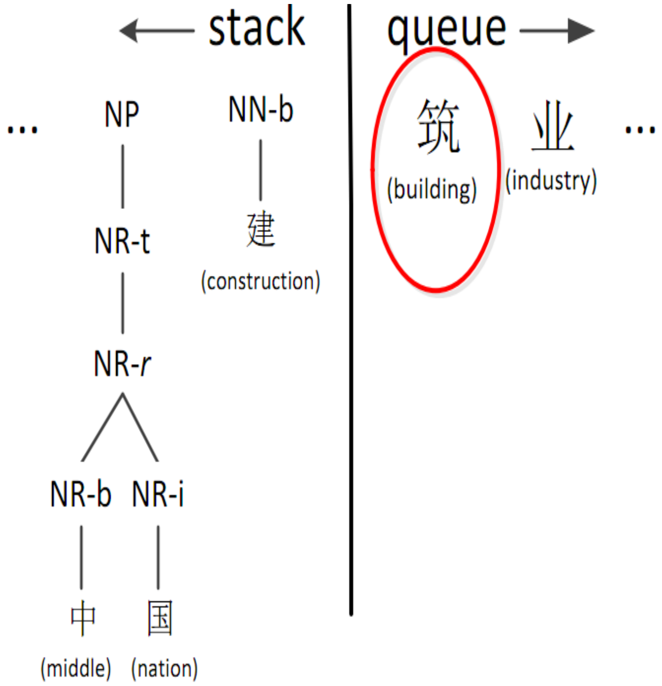
- Actions
  - SHIFT-SEPARATE(t)



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

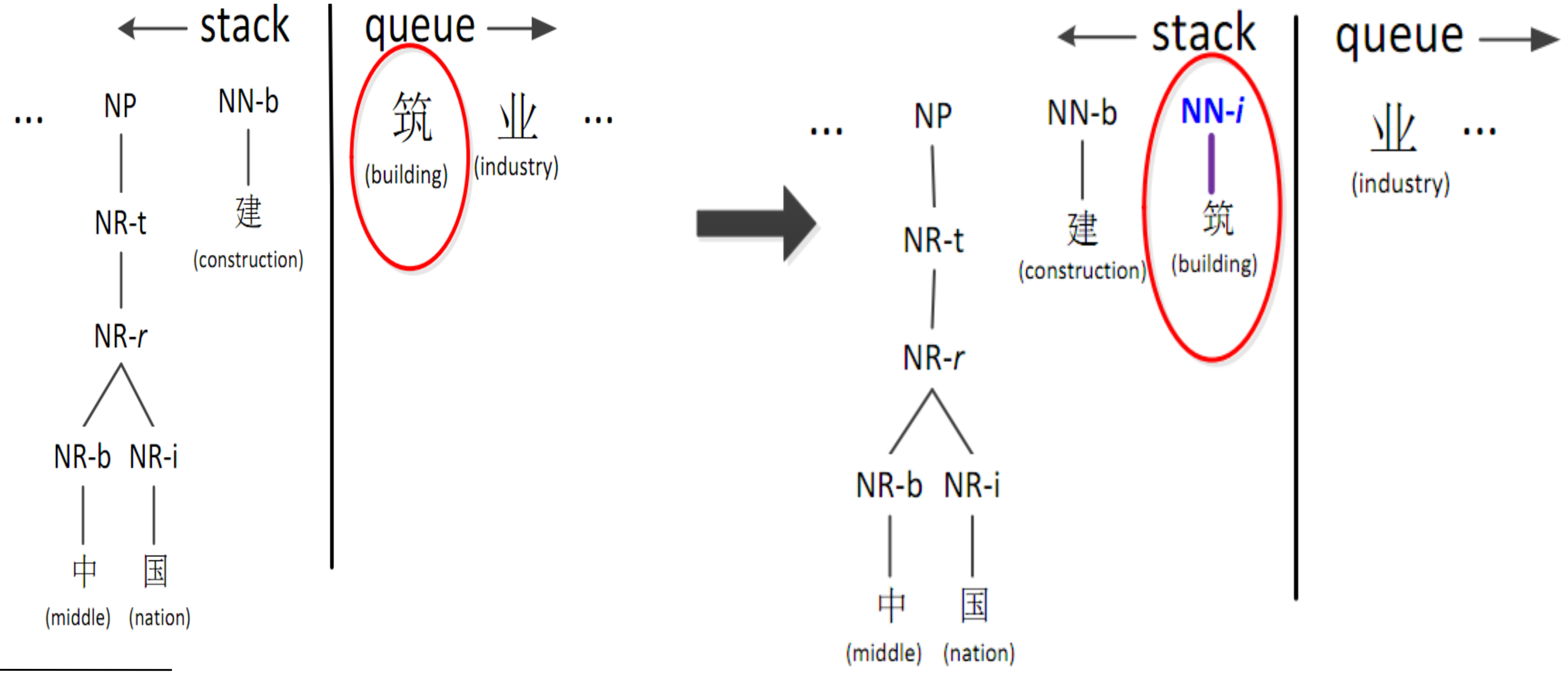
- Actions
  - SHIFT-APPEND



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

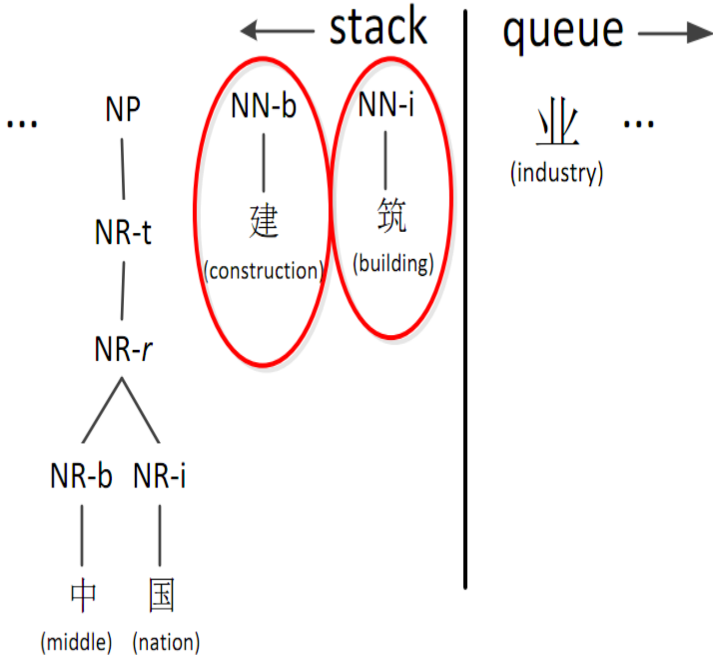
- Actions
  - SHIFT-APPEND



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

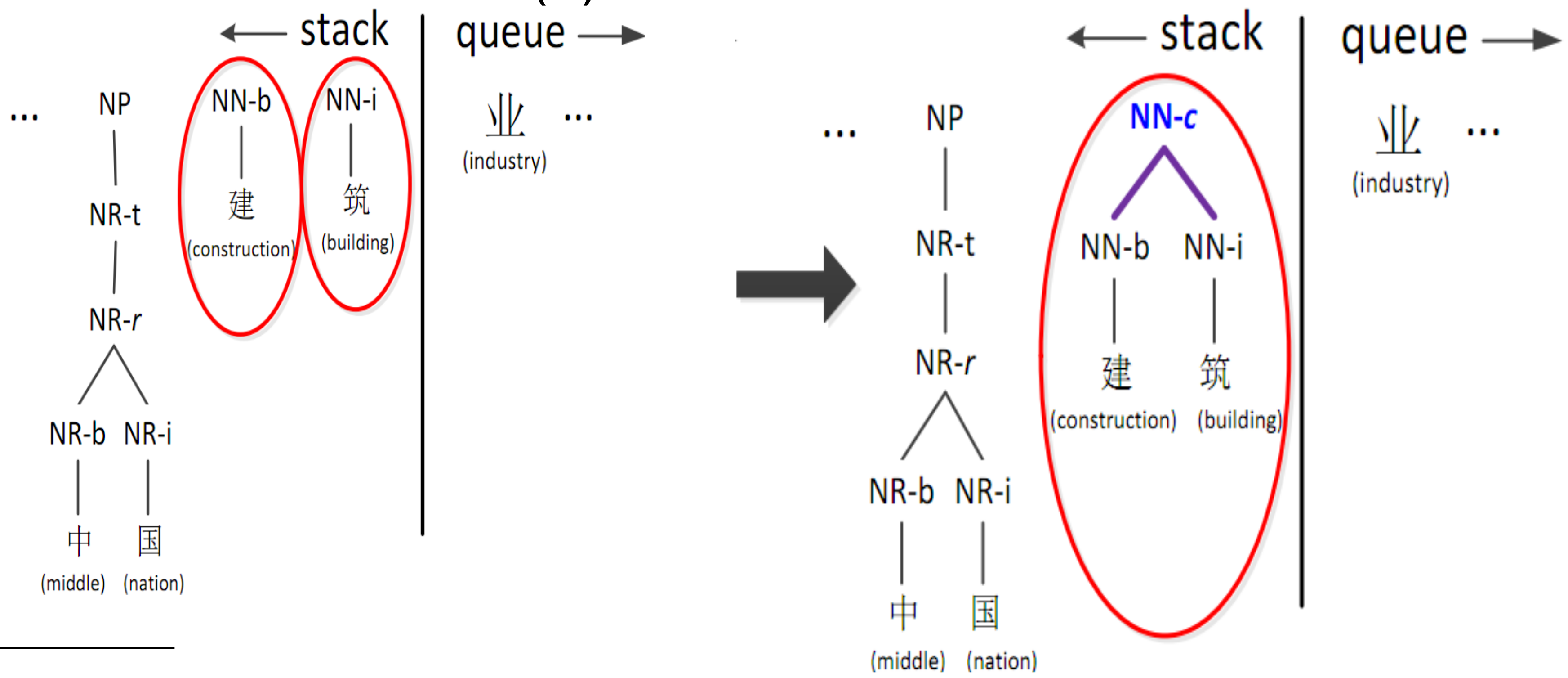
- Actions
  - REDUCE-SUBWORD(d)



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

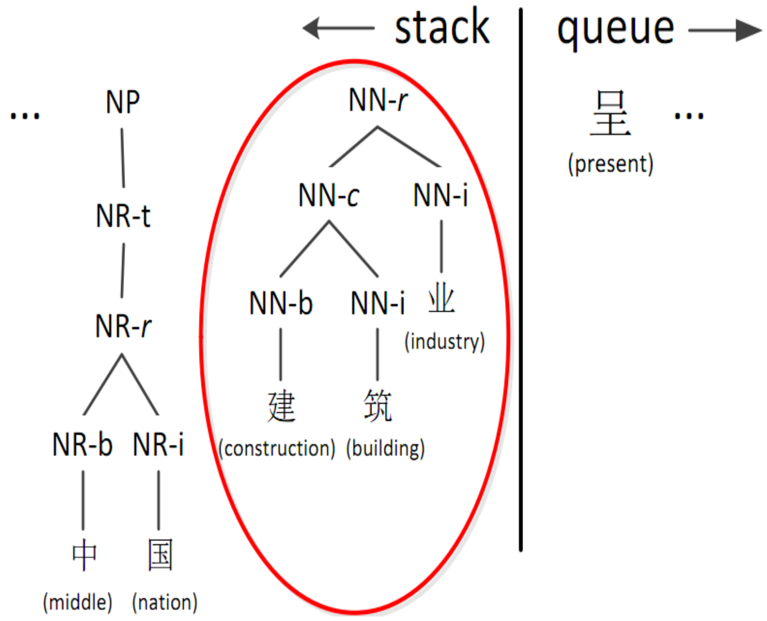
- Actions
  - REDUCE-SUBWORD(d)



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

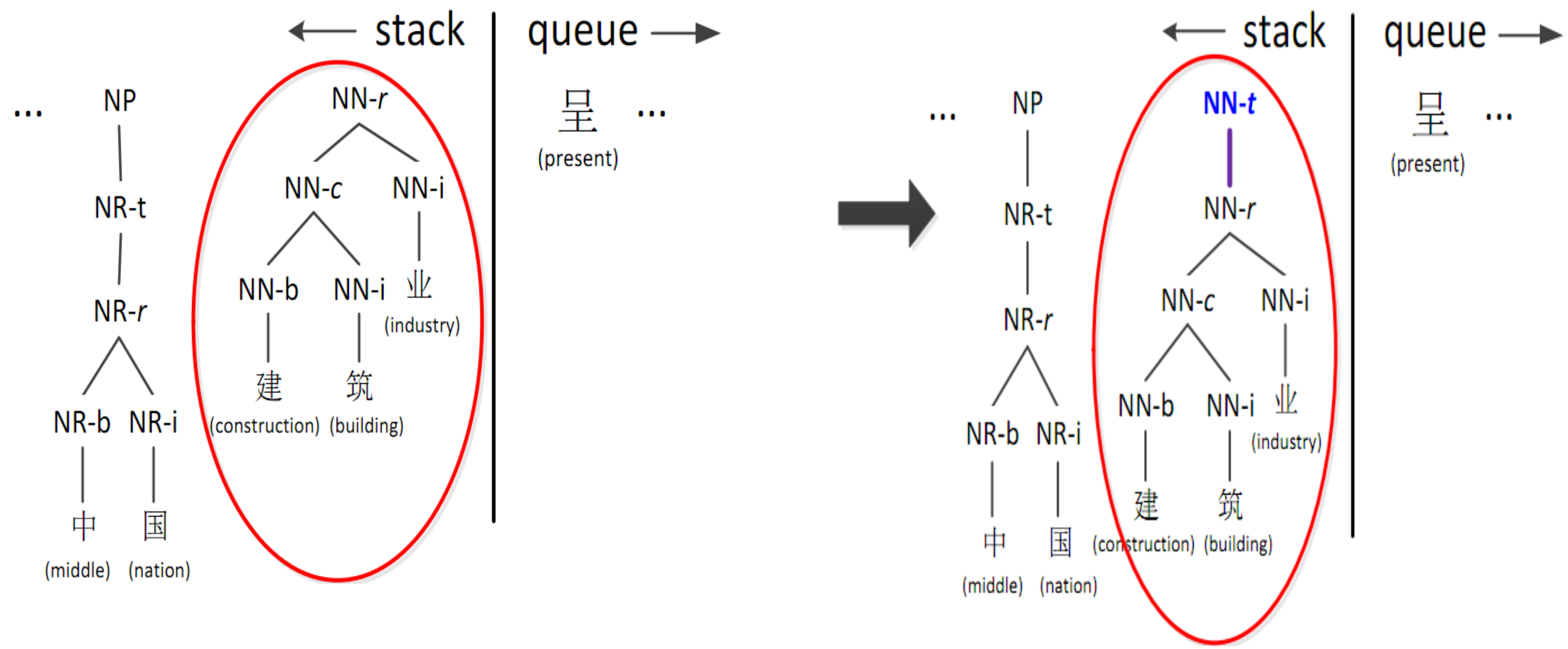
- Actions
  - REDUCE-WORD



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

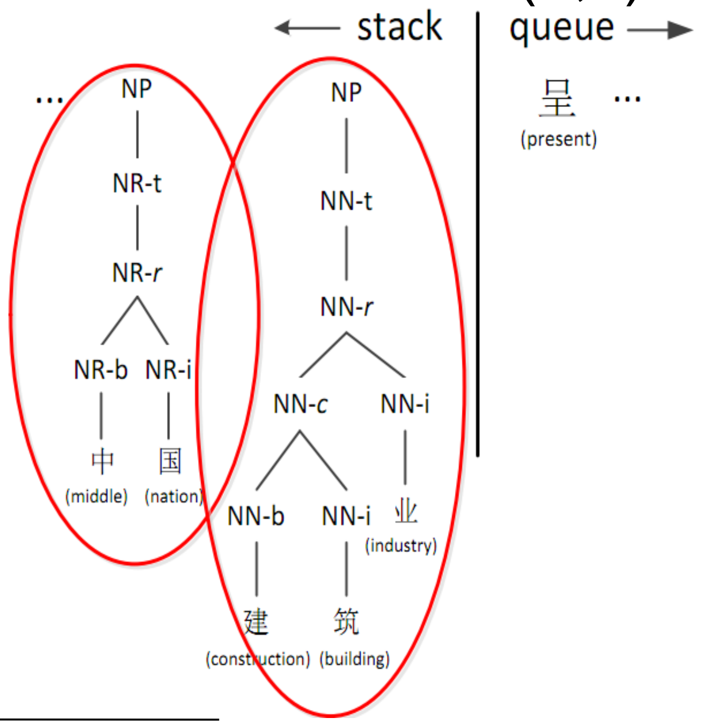
- Actions
  - REDUCE-WORD



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

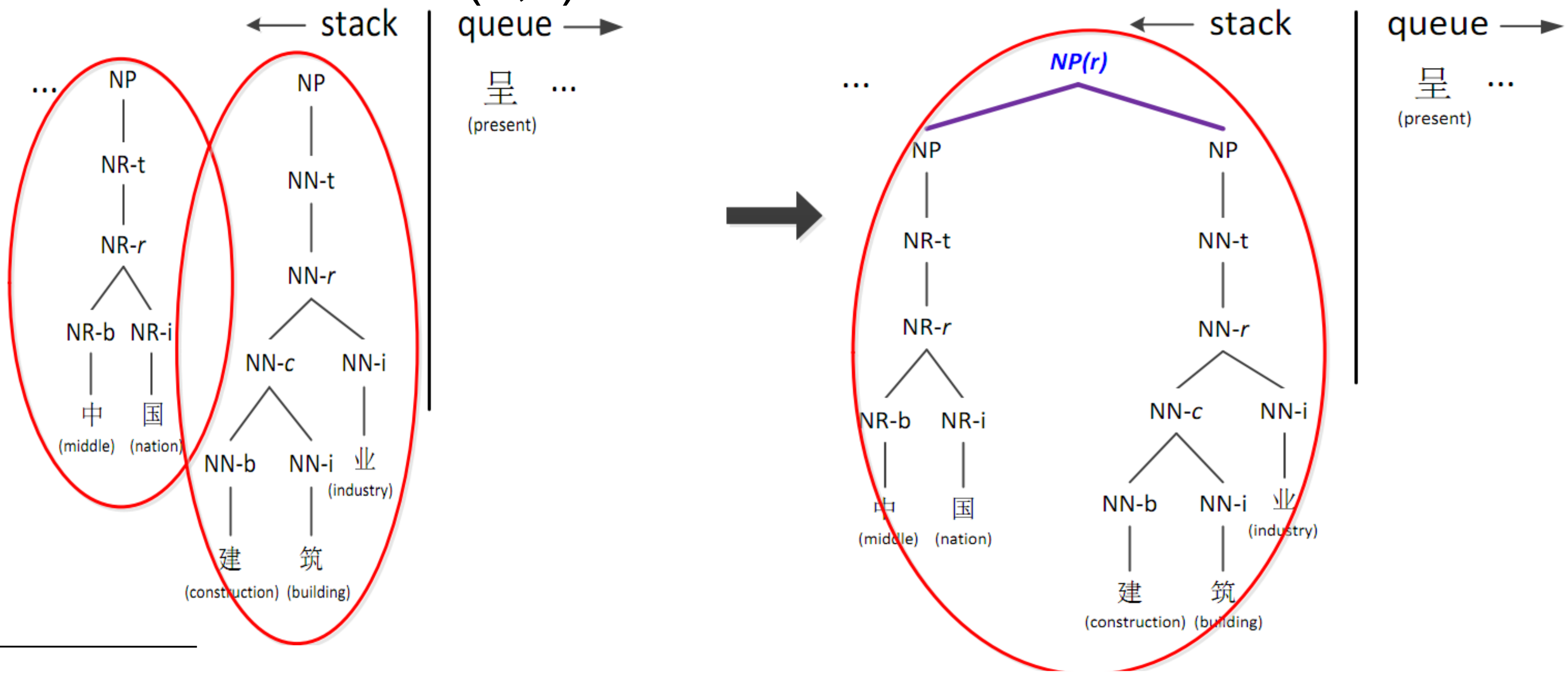
- Actions
  - REDUCE-BINARY(d; l)



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

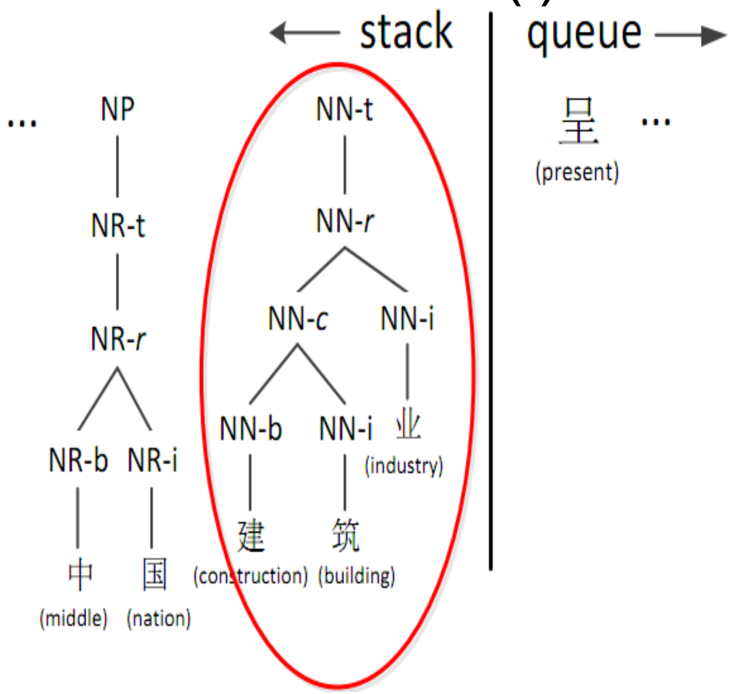
- Actions
  - REDUCE-BINARY(d; l)



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

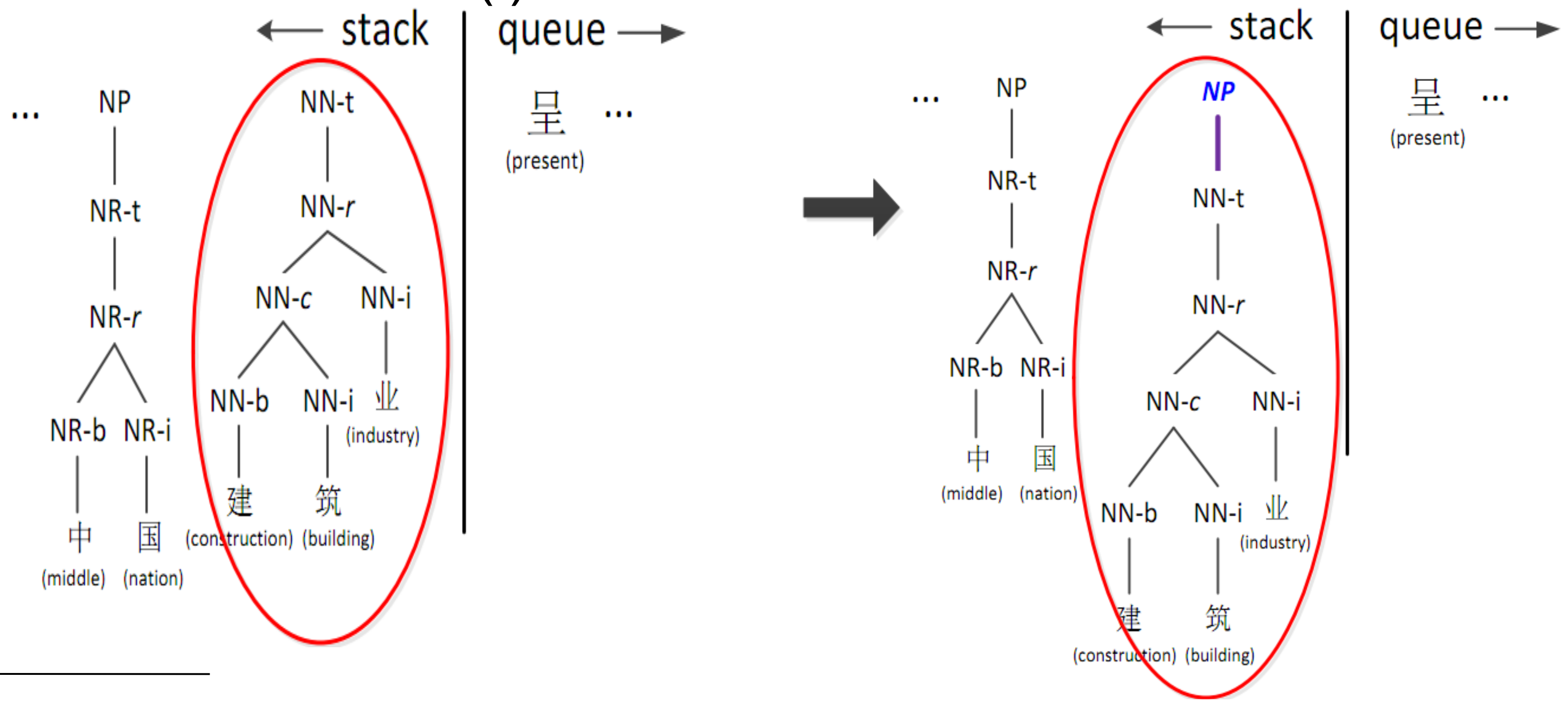
- Actions
  - REDUCE-UNARY(l)



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

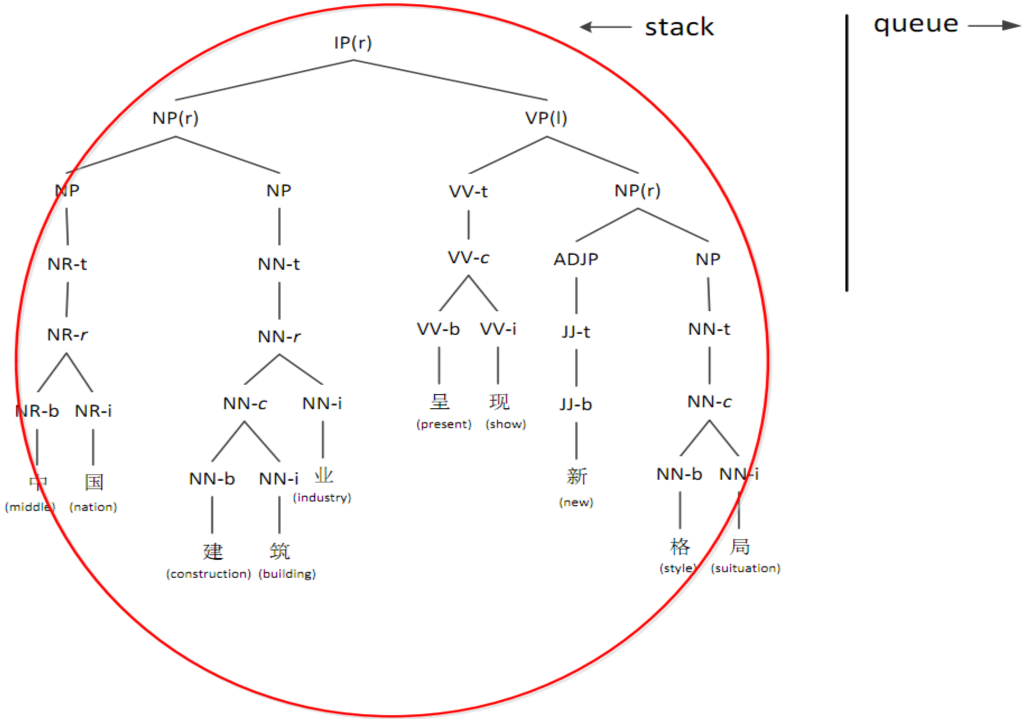
- Actions
  - REDUCE-UNARY(l)



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
  - From word-based parser (Zhang and Clark, 2009)
  - From joint SEG&POS-Tagging (Zhang and Clark, 2010)

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
  - From word-based parser (Zhang and Clark, 2009)
  - From joint SEG&POS-Tagging (Zhang and Clark, 2010)

  baseline 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
  - From word-based parser (Zhang and Clark, 2009)
  - From joint SEG&POS-Tagging (Zhang and Clark, 2010)

  baseline features

  - Deep character 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
  - From word-based parser (Zhang and Clark, 2009)
  - From joint SEG&POS-Tagging (Zhang and Clark, 2010)

  baseline features

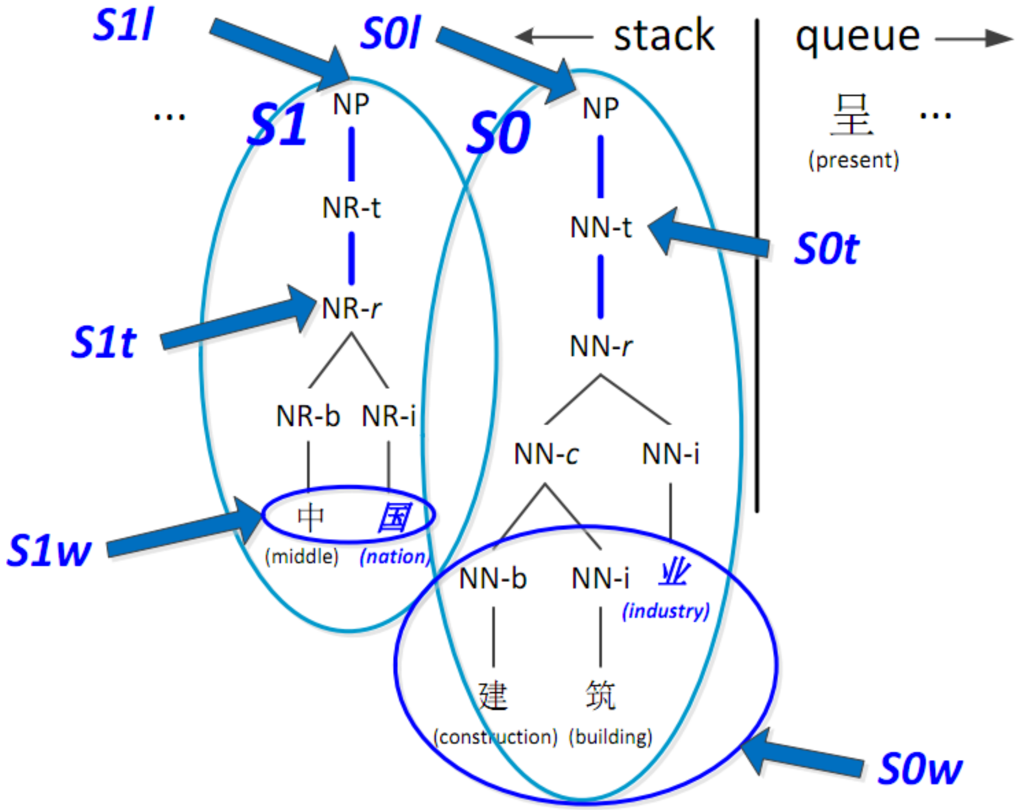
  - Deep character features

  new 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

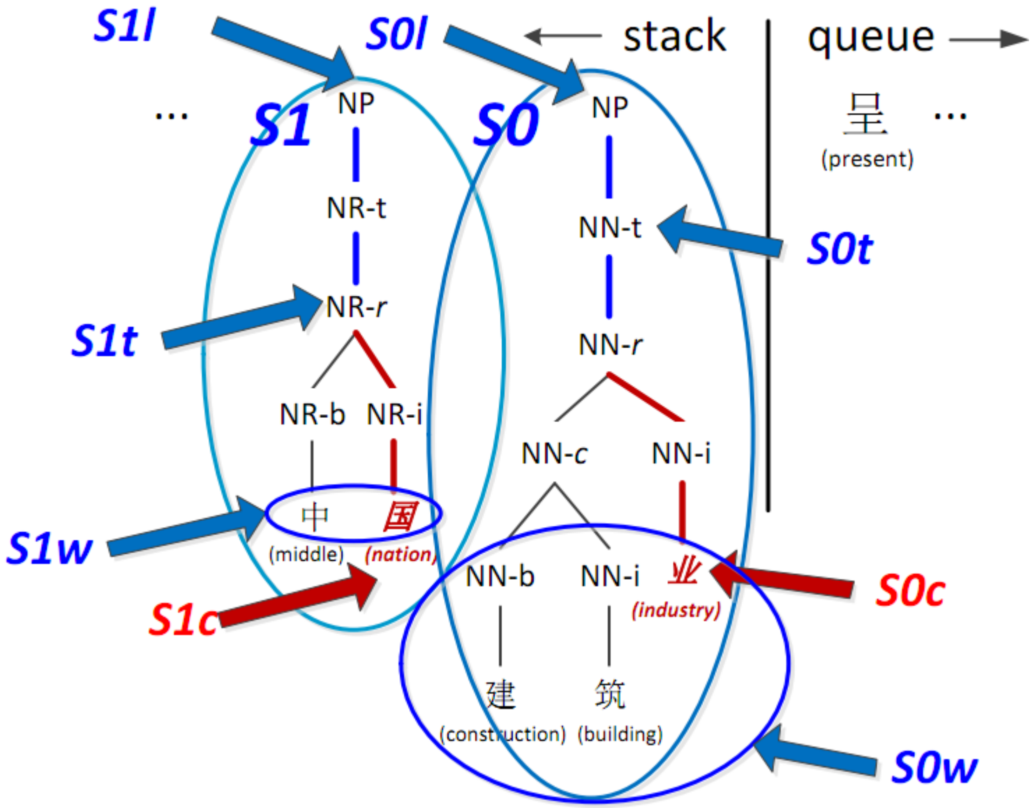
- Features



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

- 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

- Experiments
  - Penn Chinese Treebank 5 (CTB-5)

	CTB files	# sent.	# words
Training	1-270	18089	493,939
	400-1151		
Develop	301-325	350	6,821
Test	271-300	348	8,008

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

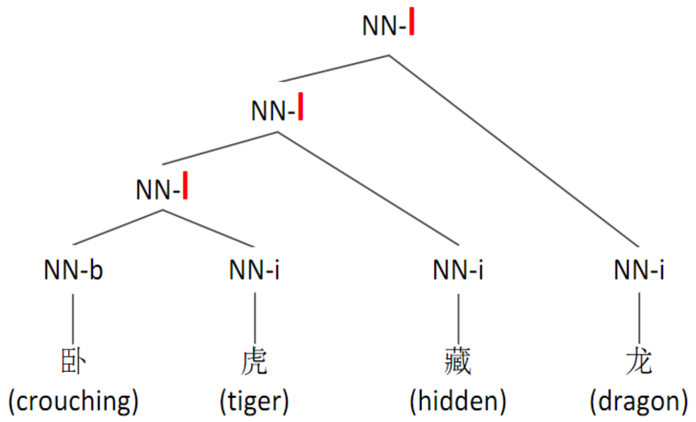
Joint Segmentation, POS-tagging and Constituent Parsing

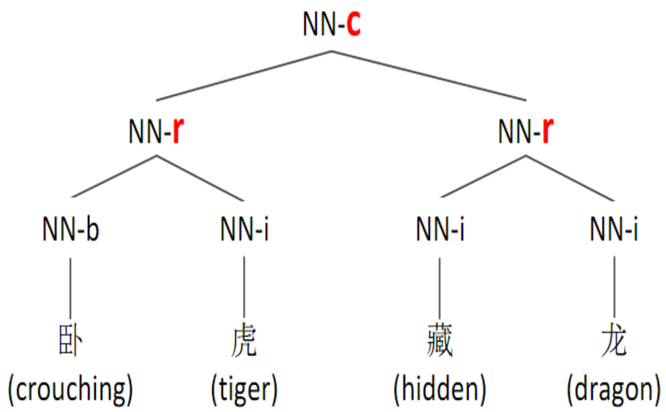
- Experiments
  - Baseline models
    - Pipeline model including:
      - Joint SEG&POS-Tagging model (Zhang and Clark, 2010).
      - Word-based CFG parsing model (Zhang and Clark, 2009).

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

- Experiments
  - Our proposed models
    - Joint model with flat word structures
    - Joint model with annotated 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

- Results

	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

- Compare with other systems

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.

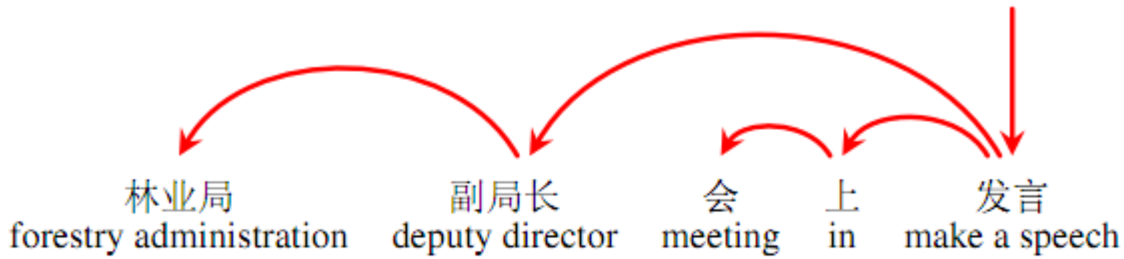
Joint Segmentation, POS-tagging and Dependency Parsing

- This paper investigate the problem of character-level Chinese dependency parsing, building dependency trees over characters.

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Segmentation, POS-tagging and Dependency Parsing

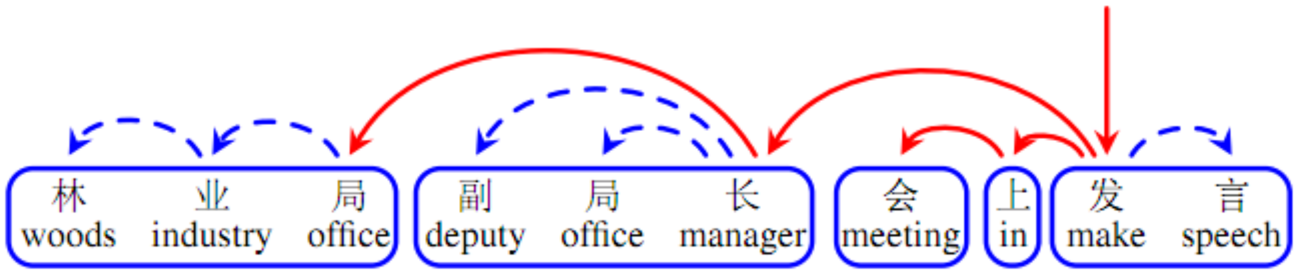
- Traditional word-based dependency parsing
    - Inter-word dependencies



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

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.

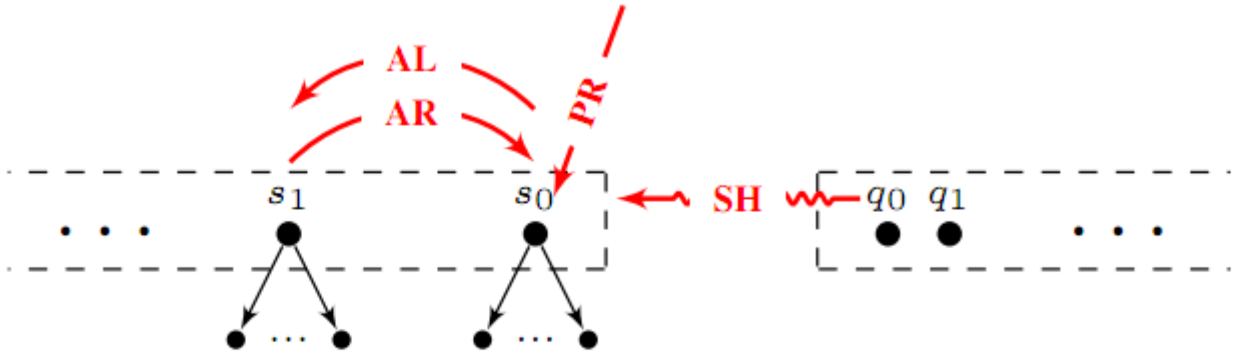
Joint Segmentation, POS-tagging and Dependency Parsing

- Main method
  - An overview
    - Transition-based framework with global learning and beam search (Zhang and Clark, 2011)
    - 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.

Joint Segmentation, POS-tagging and Dependency Parsing

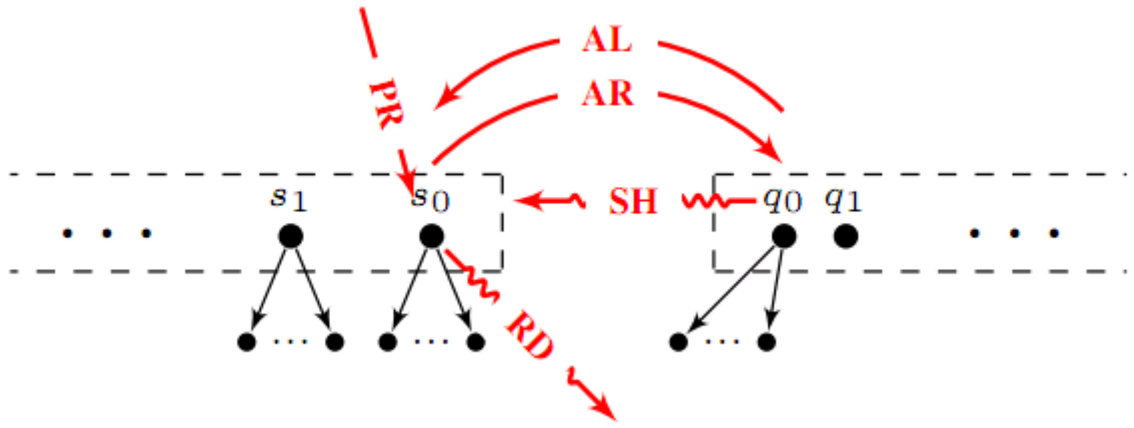
- Main method
  - Word-level transition-based dependency parsing
    - Arc-standard



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Segmentation, POS-tagging and Dependency Parsing

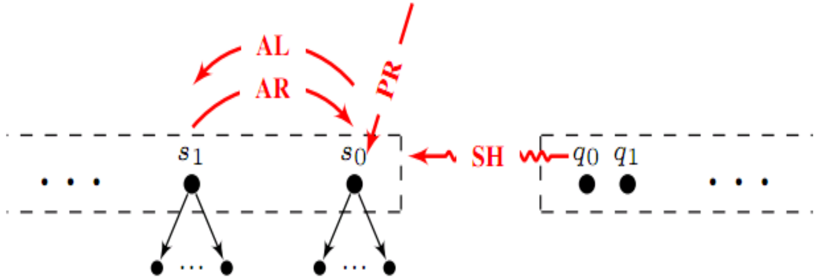
- Main method
  - Word-level transition-based dependency parsing
    - Arc-eager

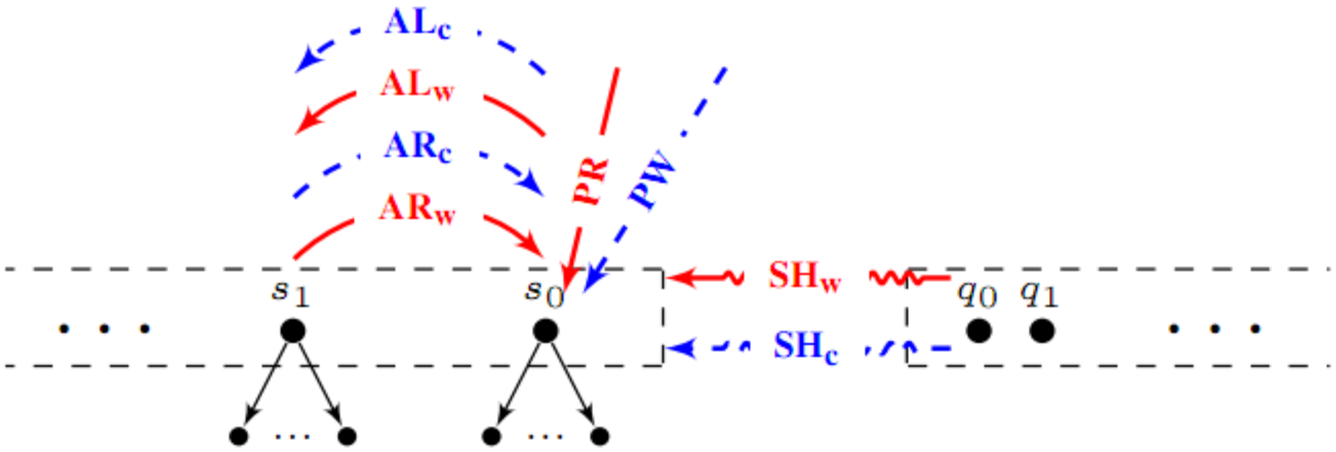


Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Segmentation, POS-tagging and Dependency Parsing

- Main method
  - 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.

Joint Segmentation, POS-tagging and Dependency Parsing

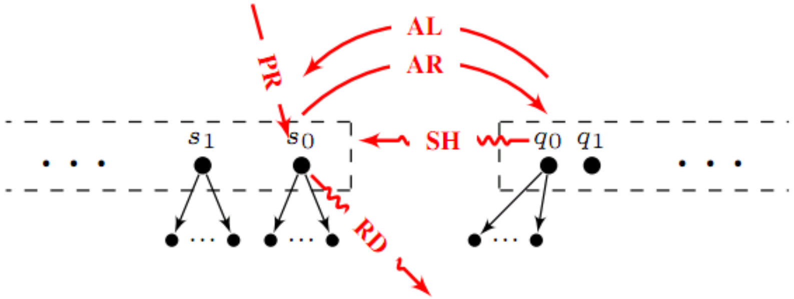
- Main method
  - Word-level to character-level
    - Arc-standard

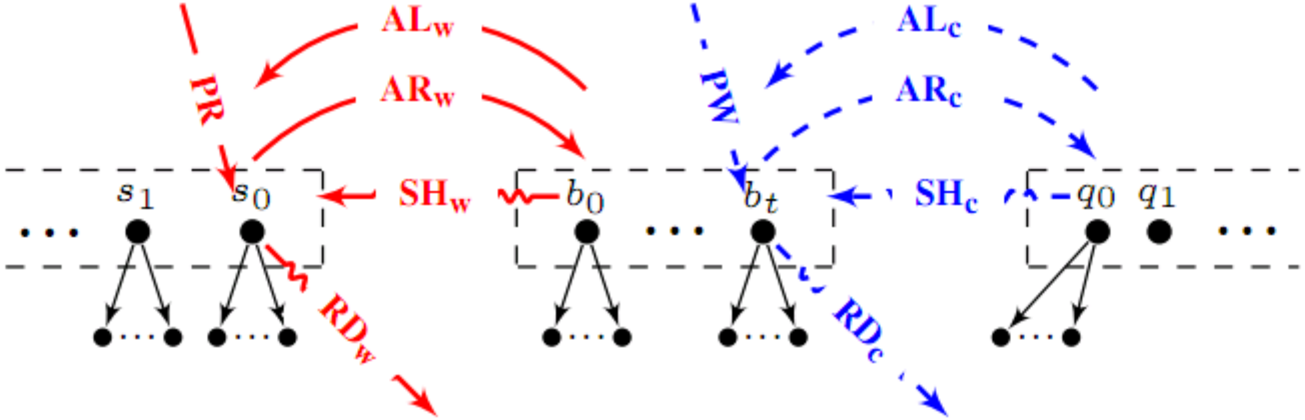
step	action	stack	queue	dependencies
0	-	ϕ	林 业 ···	ϕ
1	SH_w(NR)	林/NR	业 局 ···	ϕ
2	SH_c	林/NR 业/NR	局 副 ···	ϕ
3	AL_c	业/NR	局 副 ···	$A_1 = \{$林$\frown$业$\}$
4	SH_c	业/NR 局/NR	副 局 ···	A_1
5	AL_c	局/NR	副 局 ···	$A_2 = A_1 \bigcup \{$业$\frown$局$\}$
6	PW	林业局/NR	副 局 ···	A_2
7	SH_w(NN)	林业局/NR 副/NN	局 长 ···	A_2
···	···	···	···	···
12	PW	林业局/NR 副局长/NN	会 上 ···	A_i
13	AL_w	副局长/NN	会 上 ···	$A_{i+1} = A_i \bigcup \{$林业局/NR$\frown$副局长/NN$\}$
···	···	···	···	···

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Segmentation, POS-tagging and Dependency Parsing

- Main method
  - 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.

Joint Segmentation, POS-tagging and Dependency Parsing

- Main method
  - Word-level to character-level
    - Arc-eager

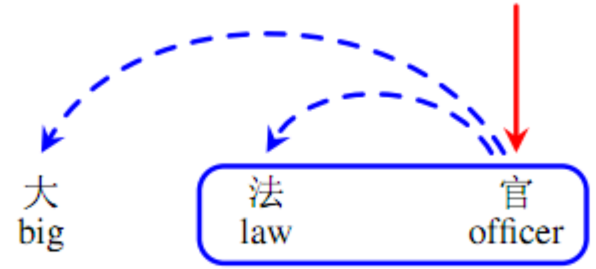
step	action	stack	deque	queue	dependencies
0	-	ϕ		林 业 $\cdots$	
1	$SH_c(NR)$	ϕ	林/NR	业 局 $\cdots$	ϕ
2	AL_c	ϕ	ϕ	业/NR 局 $\cdots$	$A_1 = \{$林$\curvearrowleft$业$\}$
3	SH_c	ϕ	业/NR	局 副 $\cdots$	A_1
4	AL_c	ϕ	ϕ	局/NR 副 $\cdots$	$A_2 = A_1 \bigcup \{$业$\curvearrowright$局$\}$
5	SH_c	ϕ	局/NR	副 局 $\cdots$	A_2
6	PW	ϕ	林业局/NR	副 局 $\cdots$	A_2
7	SH_w	林业局/NR	ϕ	副 局 $\cdots$	A_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
13	PW	林业局/NR	副局长/NN	会 上 $\cdots$	A_i
14	AL_w	ϕ	副局长/NN	会 上 $\cdots$	$A_{i+1} = A_i \bigcup \{$林业局/NR$\curvearrowright$副局长/NN$\}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

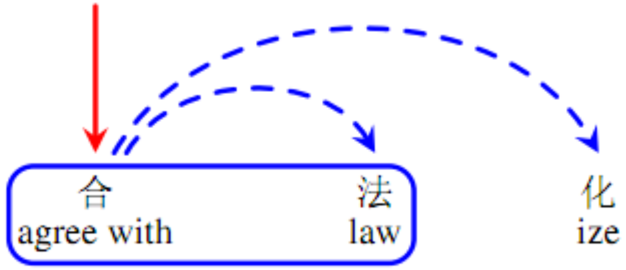
Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Segmentation, POS-tagging and Dependency Parsing

- Main method
  - New features

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





Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Segmentation, POS-tagging and Dependency Parsing

- Experiments
  - Data
    - CTB5.0, CTB6.0, CTB7.0

		CTB50	CTB60	CTB70
Training	#sent	18k	23k	31k
	#word	494k	641k	718k
Development	#sent	350	2.1k	10k
	#word	6.8k	60k	237k
	#oov	553	3.3k	13k
Test	#sent	348	2.8k	10k
	#word	8.0k	82k	245k
	#oov	278	4.6k	13k

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Segmentation, POS-tagging and Dependency Parsing

- Experiments
  - Proposed models
    - STD (real, pseudo)
      - Joint segmentation and POS-tagging with inner dependencies
    - STD (pseudo, real)
      - Joint segmentation, POS-tagging and dependency parsing
    - STD (real, real)
      - Joint segmentation, POS-tagging and dependency parsing with inner dependencies
    - EAG (real, pseudo)
      - Joint segmentation and POS-tagging with inner dependencies
    - EAG (pseudo, real)
      - Joint segmentation, POS-tagging and dependency parsing
    - EAG (real, real)
      - Joint segmentation, POS-tagging and dependency parsing with inner dependencies

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Segmentation, POS-tagging and Dependency Parsing

- Experiments
  - Final results

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.

Joint Segmentation, POS-tagging and Dependency Parsing

- Experiments
  - Analysis: word structure predication
    - OOV words
      - Overall

STD(real,real)	67.98%
EAG(real,real)	69.01%

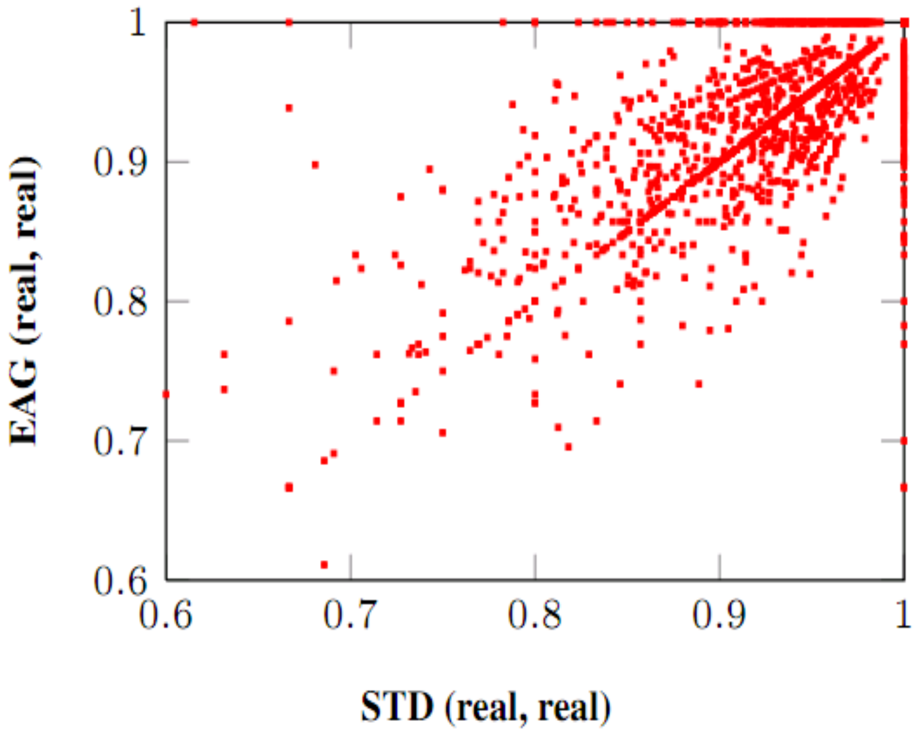
      - Assuming that the segmentation is correct

STD(real,real)	87.64%
EAG(real,real)	89.07%

Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Segmentation, POS-tagging and Dependency Parsing

- Experiments
  - Analysis: word structure predication
    - OOV words



Meishan Zhang, Yue Zhang, Wanxiang Che and Ting Liu. *Character-Level Chinese Dependency Parsing*. In Proceedings of ACL 2014. Baltimore, USA, June.

Joint Entity and Relation Extraction

- This paper investigate joint models for simultaneously extracting drugs, diseases and adverse drug events. The joint models have two main advantages.
  - They make use of information integration to facilitate performance improvement
  - They reduce error propagation in pipeline methods

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

- We define the action as:
  - 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

- We define the state of the joint model as a tuple <l, ds, dg, s>
  - l is a label sequence
  - ds is a list of readily-recognized disease entity mentions
  - dg is a list of readily-recognized drug entity mentions
  - s is a set of ADEs
- Two more actions are defined to achieve this
  - 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

- State transition examples
  - The sentence: Hepatitis caused by methotrexate and etretinate.
  - The action sequence: BD O O BC O Y BC O 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

state <l, ds, dg, s>

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

state <l, ds, dg, s>	next action
<[BD],[],[],[]>	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

state <l, ds, dg, s>	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

state <l, ds, dg, s>

next action

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

state <l, ds, dg, s>	next action
<[BD,O,O,BC],[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

state <l, ds, dg, s>

next action

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

state <l, ds, dg, s>

next action

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

state <l, ds, dg, s>

next action

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

state <l, ds, dg, s>

next action

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

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

state <l, ds, dg, s>

next action

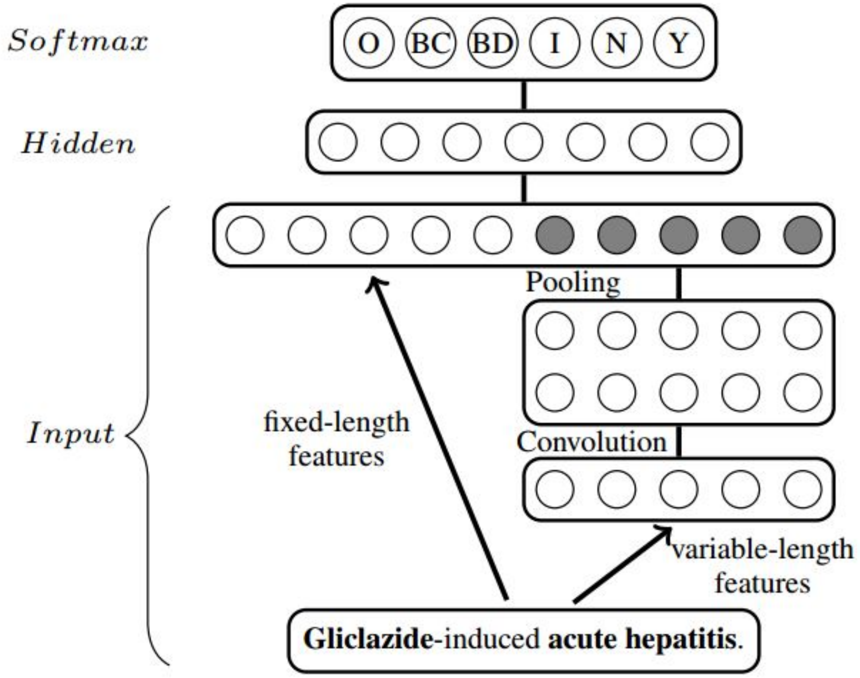
<[BD,O,O,BC,O,Y,BC,O,Y],[Hepatitis],[methotrexate,etretinate],[(Hepatiti
s,methotrexate),(Hepatitis,etretinate)]>

<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

- The neural joint 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

- Search and learning
  - Greedy
  - Local

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

- Experiments
  - Data: ADE corpus
  - Metrics: Standard precision (P), recall (R), F1-measure (F1) are used for evaluation
  - Preprocessing: The Stanford CoreNLP toolkit7 is utilized for preprocessing

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

- Experiments 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.

Outline

- Motivation
- Statistical Models
- Deep Learning Models

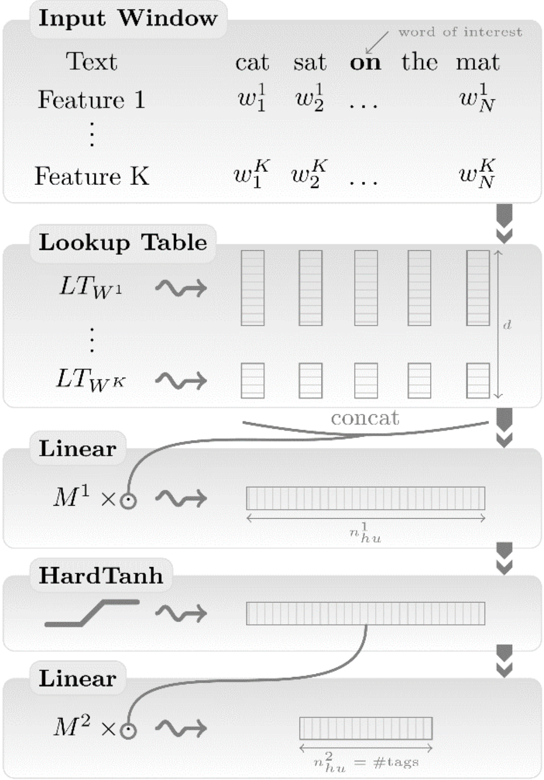
Joint Tagging, Chunking and NER

- Features trained for one task can be useful for related tasks. Multi-task learning (MTL) leverages this idea in a more systematic way. This paper trained jointly POS, CHUNK and NER using the window approach network.

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

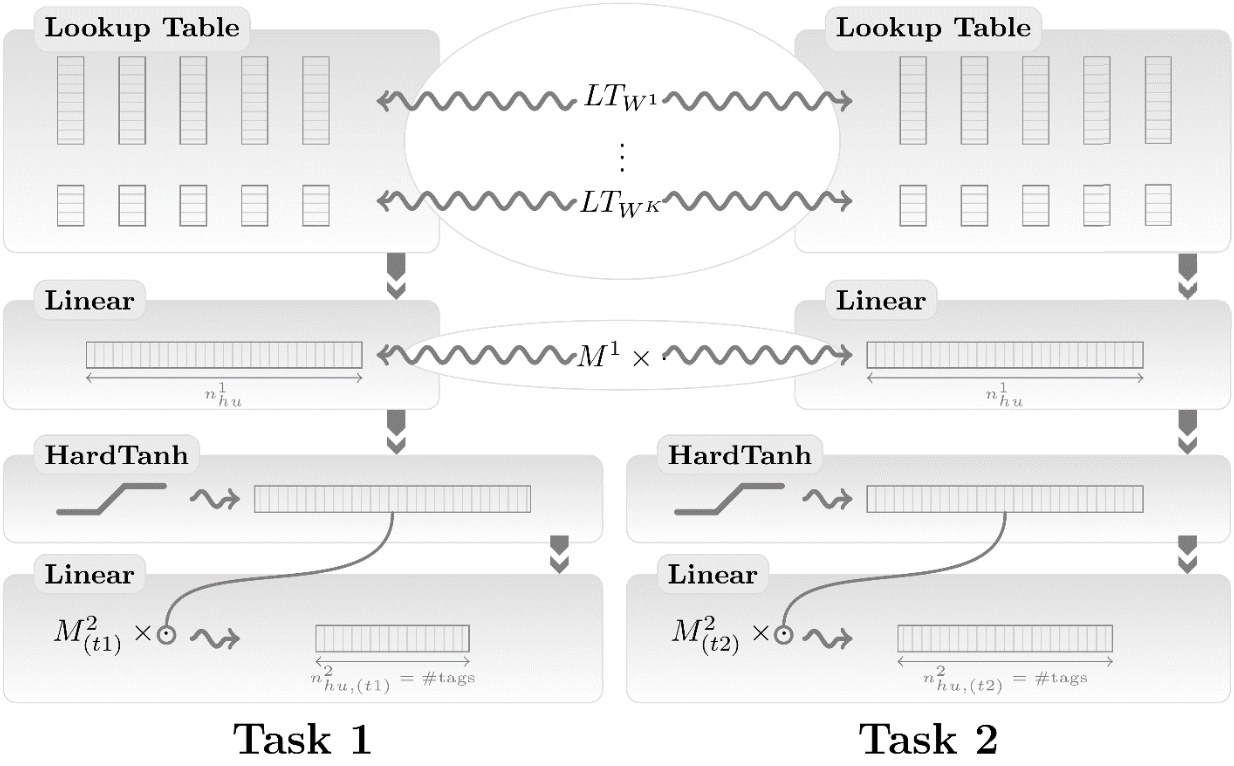
- window approach network.



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

- Example of multitasking with NN



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

- All models share the lookup table parameters
- The parameters of the first linear layers are shared in the window approach case
- Training is achieved by minimizing the loss averaged across all tasks

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

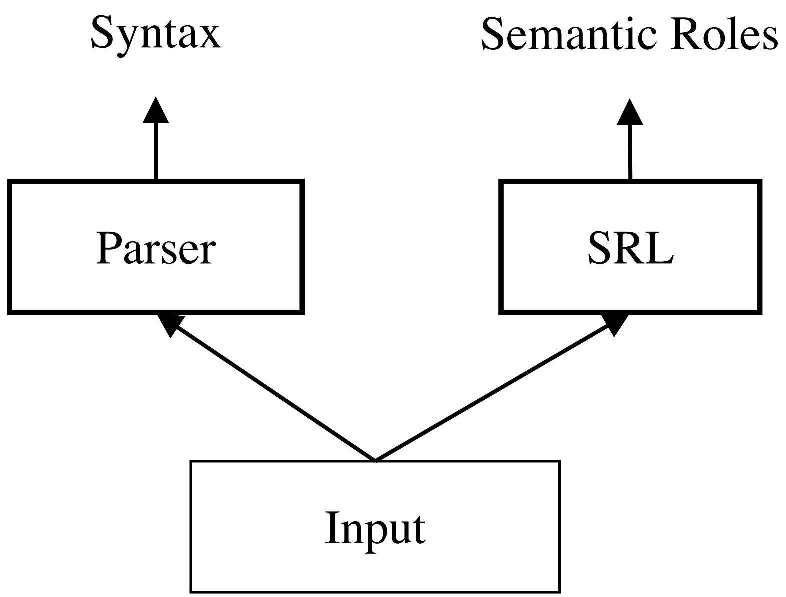
- Experiments

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.

Joint Parsing and SRL

- Model



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

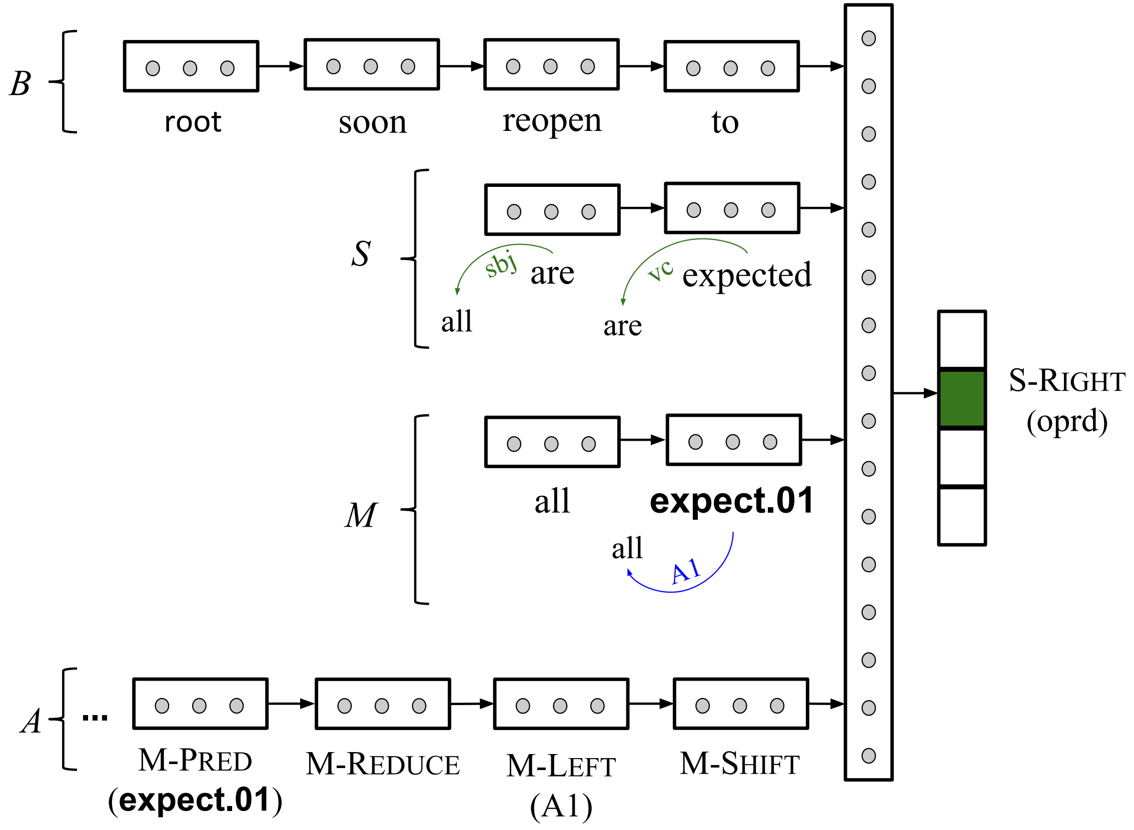
- Experiment Results

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

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

- 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

- Shift-Reduce

Transition	S	M	B	Dependency
	[]	[]	[all, are, expected, to, reopen, soon, root]	—
S-SHIFT	[all]	[]	[all, are, expected, to, reopen, soon, root]	—
M-SHIFT	[all]	[all]	[are, expected, to, reopen, soon, root]	—
S-LEFT(*sbj*)	[]	[all]	[are, expected, to, reopen, soon, root]	all $\xleftarrow{\text{sbj}}$ are
S-SHIFT	[are]	[all]	[are, expected, to, reopen, soon, root]	—
M-SHIFT	[are]	[all, are]	[expected, to, reopen, soon, root]	—
S-RIGHT(*vc*)	[are, expected]	[all, are]	[expected, to, reopen, soon, root]	are $\xrightarrow{\text{vc}}$ expected
M-PRED(**expect.01**)	[are, expected]	[all, are]	[expected, to, reopen, soon, root]	—
M-REDUCE	[are, expected]	[all]	[expected, to, reopen, soon, root]	—
M-LEFT(*A1*)	[are, expected]	[all]	[expected, to, reopen, soon, root]	all $\xleftarrow{\text{A1}}$ expect.01
M-SHIFT	[are, expected]	[all, expected]	[to, reopen, soon, root]	—
***S-RIGHT(oprd)	[are, expected, to]	[all, expected]	[to, reopen, soon, root]	expected $\xrightarrow{\text{oprd}}$ to
M-RIGHT(*C-A1*)	[are, expected, to]	[all, expected]	[to, reopen, soon, root]	expect.01 $\xrightarrow{\text{C-A1}}$ to
M-REDUCE	[are, expected, to]	[all]	[to, reopen, soon, root]	—
M-SHIFT	[are, expected, to]	[all, to]	[reopen, soon, root]	—
S-RIGHT(*im*)	[are, expected, to, reopen]	[all, to]	[reopen, soon, root]	to $\xrightarrow{\text{im}}$ reopen
M-PRED(**reopen.01**)	[are, expected, to, reopen]	[all, to]	[reopen, soon, root]	—
M-REDUCE	[are, expected, to, reopen]	[all]	[reopen, soon, root]	—
M-LEFT(*A1*)	[are, expected, to, reopen]	[all]	[reopen, soon, root]	all $\xleftarrow{\text{A1}}$ reopen.01
M-REDUCE	[are, expected, to, reopen]	[]	[reopen, soon, root]	—
M-SHIFT	[are, expected, to, reopen]	[reopen]	[soon, root]	—
S-RIGHT(*tmp*)	[are, expected, to, reopen, soon]	[reopen]	[soon, root]	reopen $\xrightarrow{\text{tmp}}$ soon
M-RIGHT(*AM-TMP*)	[are, expected, to, reopen, soon]	[reopen]	[soon, root]	reopen.01 $\xrightarrow{\text{AM-TMP}}$ soon
M-REDUCE	[are, expected, to, reopen, soon]	[]	[soon, root]	—
M-SHIFT	[are, expected, to, reopen, soon]	[soon]	[root]	—
S-REDUCE	[are, expected, to, reopen]	[soon]	[root]	—
S-REDUCE	[are, expected, to]	[soon]	[root]	—
S-REDUCE	[are, expected]	[soon]	[root]	—
S-REDUCE	[are]	[soon]	[root]	—
S-LEFT(*root*)	[]	[soon]	[root]	are $\xleftarrow{\text{root}}$ root
S-SHIFT	[root]	[soon]	[root]	—
M-REDUCE	[root]	[]	[root]	—
M-SHIFT	[root]	[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

- Compared with state-of-art

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).

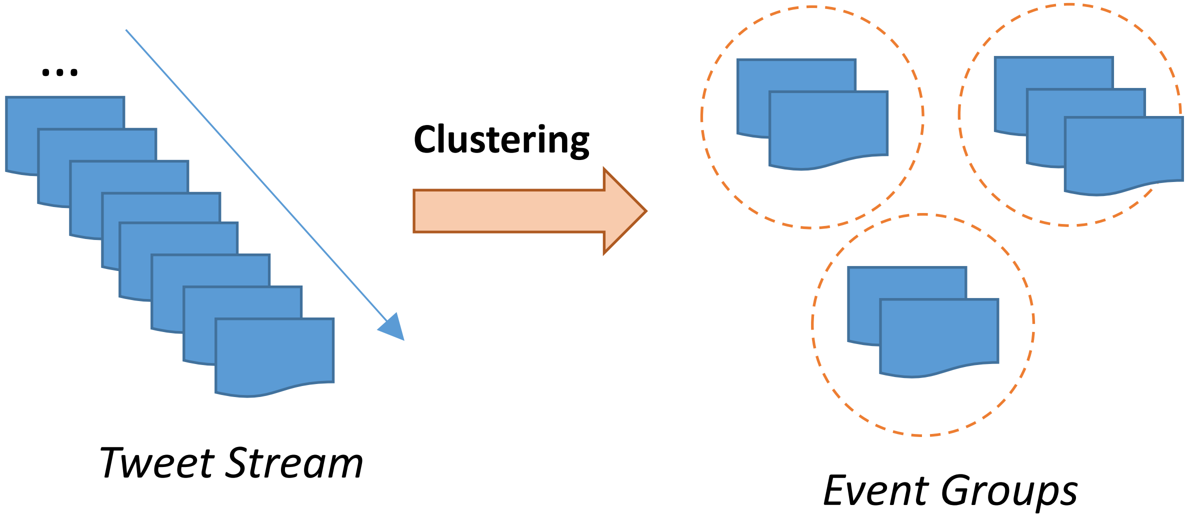
Joint Event Detection and Reporting

- This paper build a joint model to filter, cluster, and summarize the tweets for new events. In particular, deep representation learning is used to vectorize tweets, which serves as basis that connects tasks. A neural stacking model is used for integrating a pipeline of different sub tasks, and for better sharing between the predecessor and successors.

Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Tweet New Event Detection
  - Aims to identify **first stories** in a tweet stream
    - *Incremental clustering* is always used to cluster tweets into event groups.

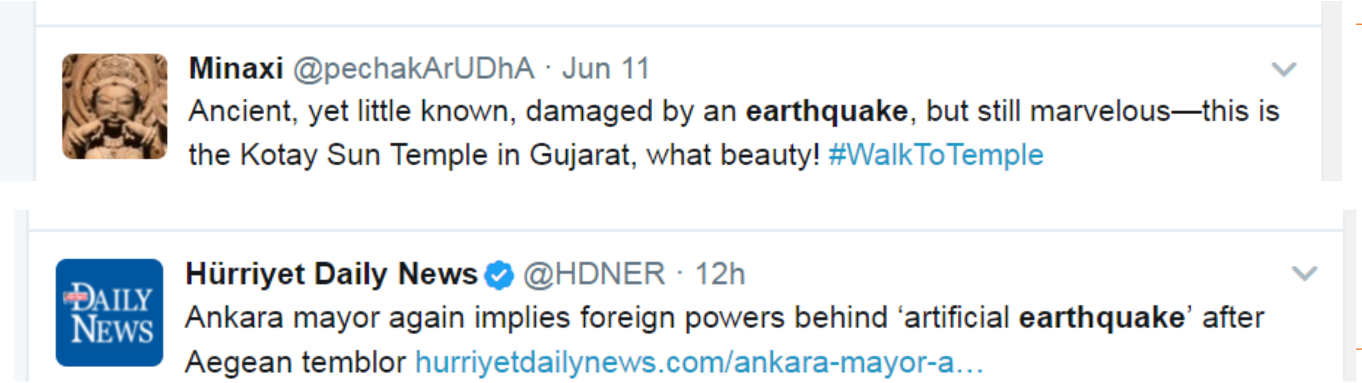


Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Challenges of Tweet New Event Detection
  - There are lots of noise tweets in the tweet stream

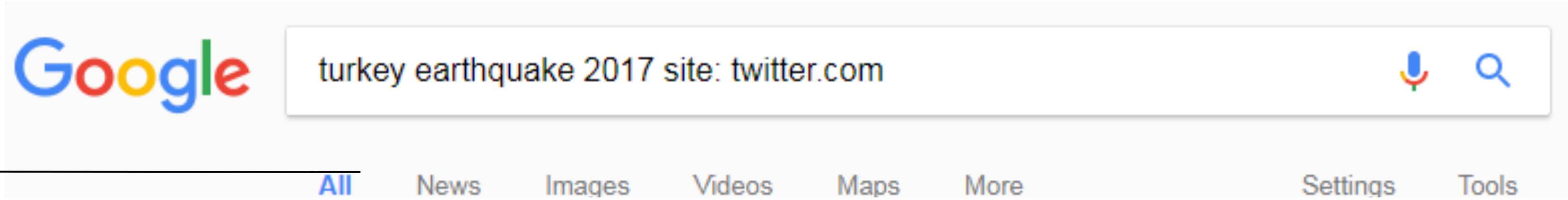
Need to filter



Contain earthquake keyword. But do not mention any earthquake event

  - Every event is mentioned by too many tweets

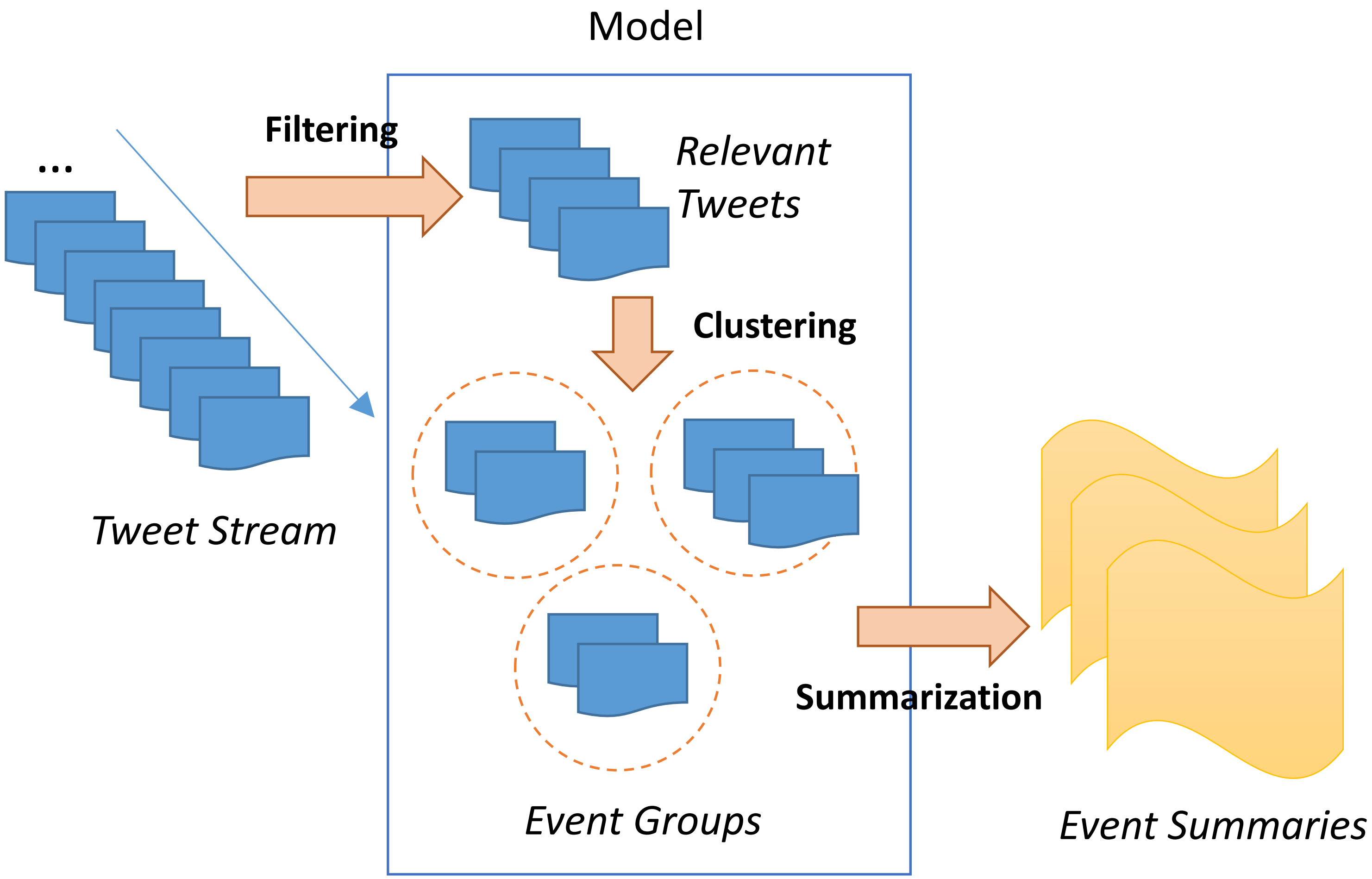
Need to summarize



Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Solution: Not only cluster events, but also filter tweets and summarize events.
  - Tweets Filtering
  - Event Clustering
  - Event Summarization



Joint Event Detection and Reporting

- Correlation between Different Stages
  - A tweet that comprehensively describes an event should be scored highly in both the *relevance-filtering* and the *extractive-summarization* steps.
  - Better understanding of a tweet is helpful for both *relevance-filtering* and *event-clustering*.

Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

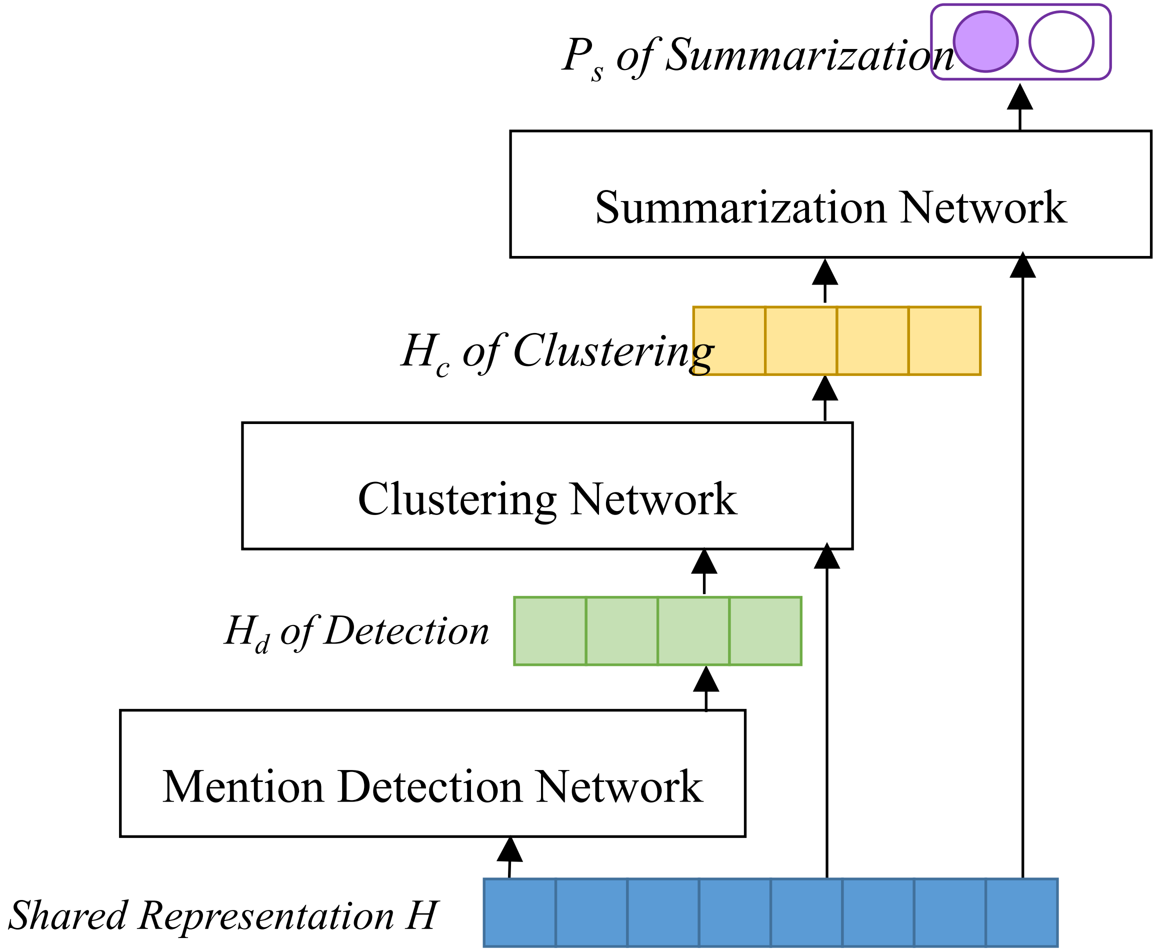
Joint Event Detection and Reporting

- Detect and Summarize Event Jointly
  - A deep neural network is used to model the three subtasks jointly
    - *Representation learning* is used to transform each incoming tweet into a dense low dimension vector
    - *Neural stacking* is used to integrate different subtasks.

Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Overview of Joint Event Detection and Summarization
  - Shared Representation
    - LSTM
  - Joint Model
    - Tweet Filtering
    - Event Clustering
    - Event Summarization



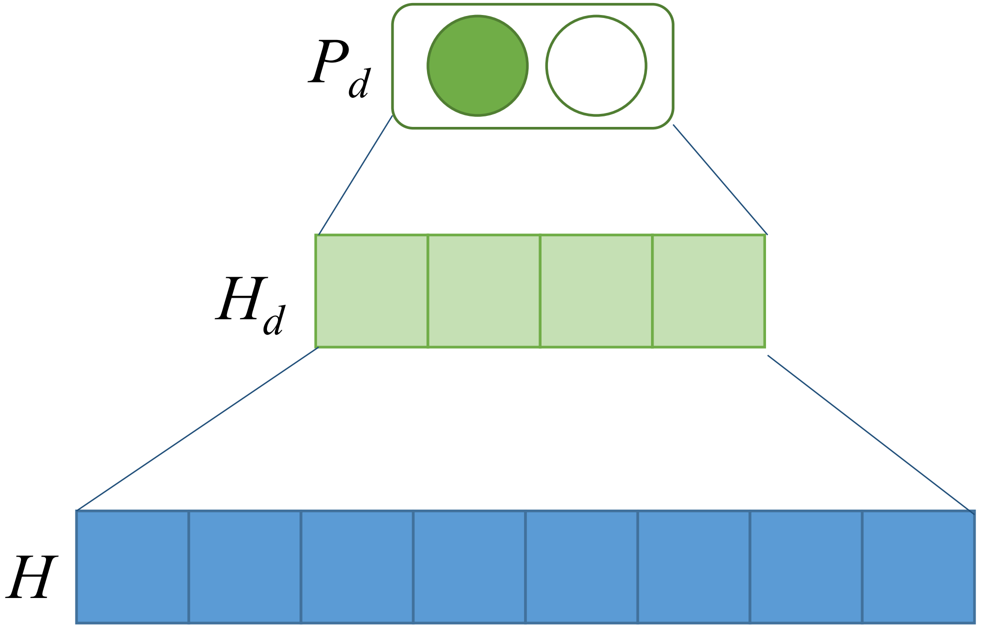
Joint Event Detection and Reporting

- Tweet Filtering
  - We classify each tweet in the stream as either being relevant or irrelevant to the events of concern.
    - A binary classification task
    - A multi-layer perceptron

$$H_d = \sigma(W_d^h \boxed{H} + b_d^h),$$

hidden variables of tweet

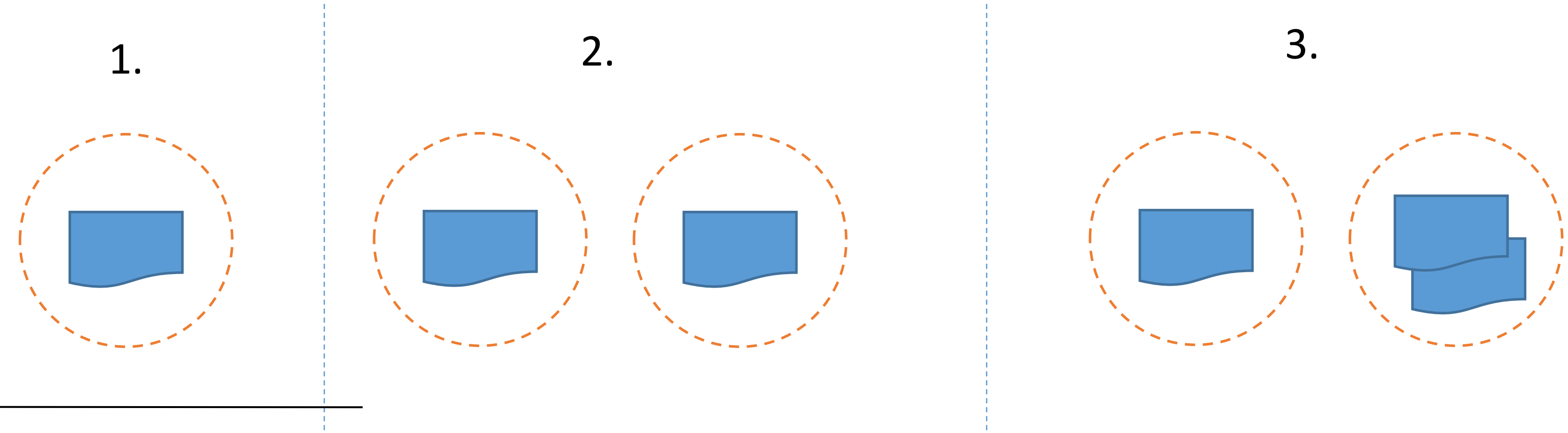
$$P_d = \mathrm{softmax}(W_d H_d + B_d)$$



Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Event Clustering
  - Incremental clustering of tweets [Aggarwal and Subbian, 2012].
    - Given a new tweet, decide whether it belongs to an existing event cluster, or describes a new event
    - A key issue is the calculation of *similarity between tweets*.



Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

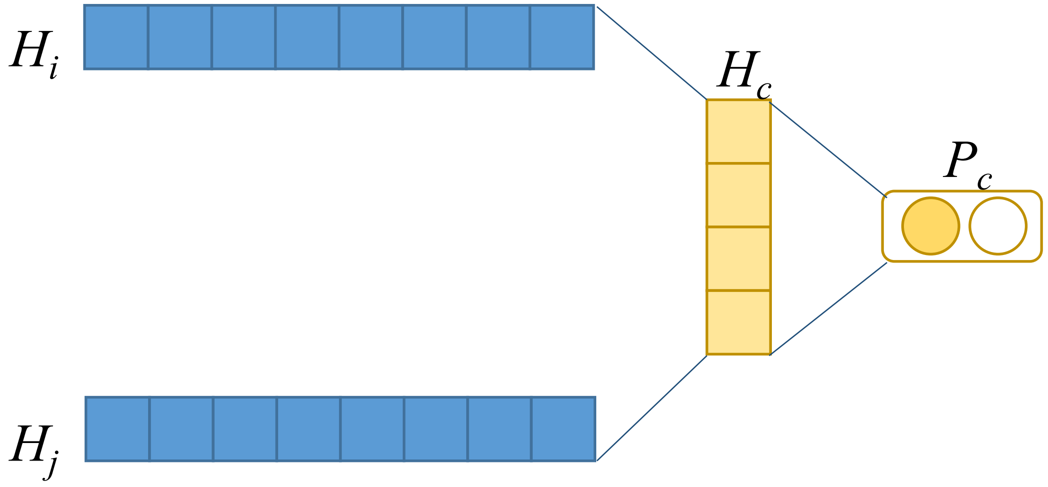
Joint Event Detection and Reporting

- Siamese Network for calculating similarity
  - Siamese Network

$$H_c = \sigma(W_c^h(\boxed{H_i \oplus H_j}) + b_c^h)$$

hidden variables of tweet *i* and *j*

$$P_c = \mathrm{softmax}(W_c H_c + B_c)$$



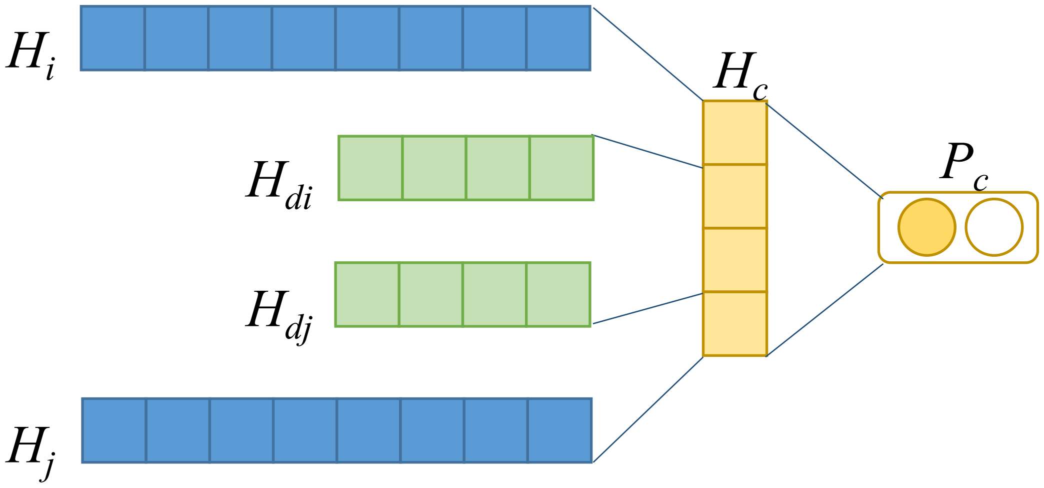
Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Integrating with tweet filtering

$$H_c = \sigma(W_c^h(H_i \oplus H_j) + b_c^h) \Rightarrow H_c = \sigma(W_c^h(H_i \oplus H_j \oplus \boxed{H_{d_i} \oplus H_{d_j}}) + b_c^h),$$

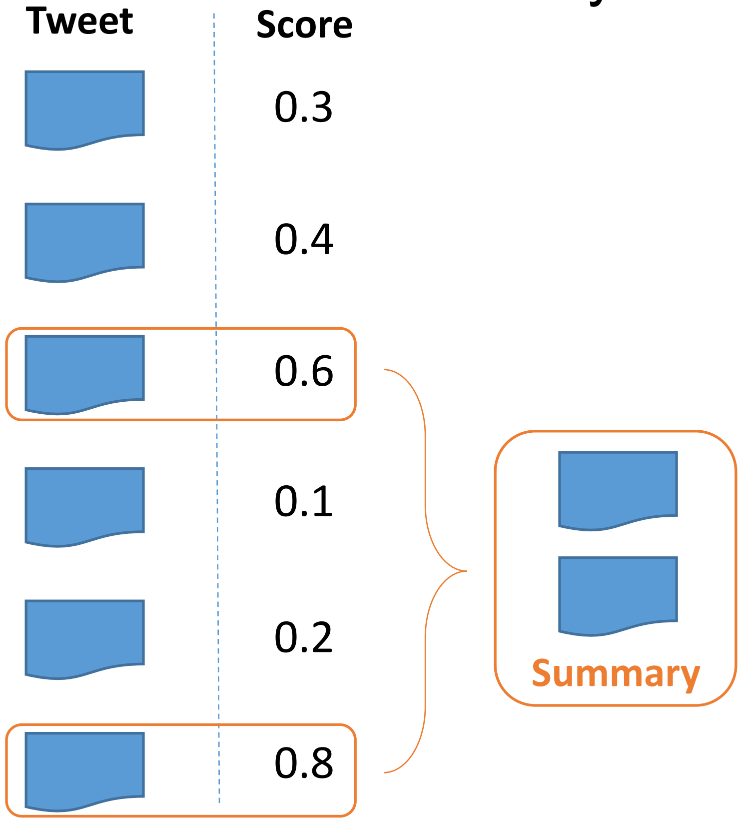
Hidden variables from tweet filtering



Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Event Summarization
  - We rank all the tweets in the cluster using a probability score, and select top-n to build the summary.



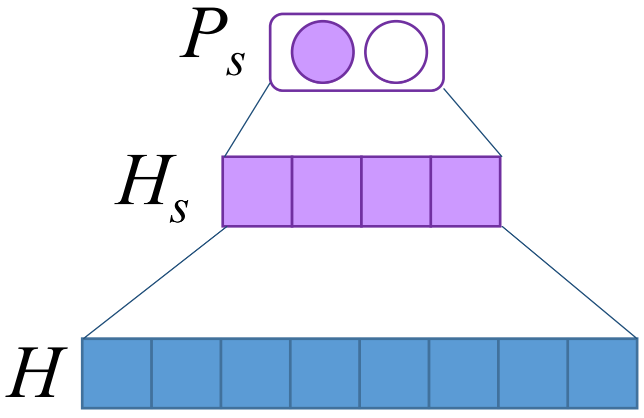
Joint Event Detection and Reporting

- Event Summarization (cont.)
  - A multi-layer perceptron

$$H_s = \sigma(W_s^h \boxed{H} + b_s^h)$$

hidden variables of tweet

$$P_s = \mathrm{softmax}(W_s H_s + B_s)$$



Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

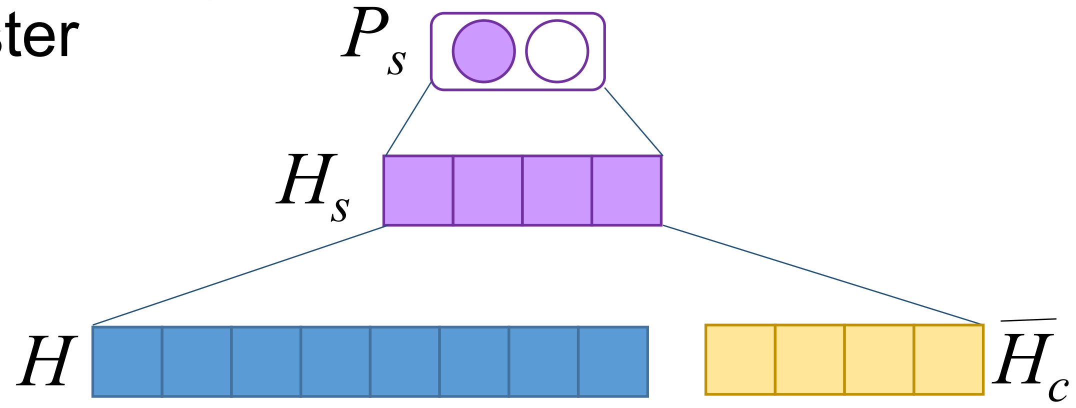
Joint Event Detection and Reporting

- Integrating with event clustering

$$H_s = \sigma(W_s^h(H \oplus \boxed{\overline{H_c^h}}) + b_s^h)$$

$$P_s = \mathrm{softmax}(W_s H_s + B_s)$$

  - $\overline{H_c^h}$ is the sum of H_c^h between the tweet X and all the other tweets in the same cluster



Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Data Collection
  - All data were collected by using the Twitter streaming API
    - consist of tweets from June 2013 until April 2016
  - The tweets are collected with relevant domain keywords
    - Earthquake:
      - earthquake, shake, refugees, victims
    - DDoS:
      - ddos, anonymous attack, spoofed attack, zombies host

Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Event Annotation
  - We adopt the approach employed by NIST in labeling TDT data [Allan, 2002]
    - A relevant tweet must explicitly mention the event
    - The main purpose of the tweet should be to inform of the event
  - Statistic of dataset

	Earthquake	DDoS
#Event	47	170
#Post	12090	17760
Vocabulary size	11462	15032

Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Evaluation Metrics
    - Clustering
        - We use the standard TDT evaluation procedure [Allan, 2002], where normalized *Topic Weighted Minimum Cost (C_{min})* is taken for evaluating clustering accuracy
    - Summarization
        - We use ROUGE-1.5.5 [Lin, 2004] for summary evaluation. We report *unigram overlap (ROUGE-1)* for assessing informativeness.
    - Firstly, we evaluate our proposed model on *earthquake* domain.

Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Effectiveness of Event Mention Detection
  - Below table indicates the event clustering performance with/without the event mention detection.
    - *Cosine* is a traditional strategy with bag-of-words as document representation [Aggarwal and Subbian, 2012]
    - *LSTM* means calculating the similarity using the LSTM based Siamese network [Mueller and Thyagarajan, 2016].

Method	C_{min}
Random	86.2
Cosine − filtering	65.8
Cosine + filtering	60.9
LSTM − filtering	64.4
LSTM + filtering	58.8

Event filtering always outperform those without event mention filtering

Neural Network is better than BOW model

Joint Event Detection and Reporting

- Effectiveness of Joint Modeling
  - The results of different ablation baselines

Method	Clustering	Summarization
LSTM-Pipeline	58.8	18.2
LSTM-Joint	52.2	19.4
+Detect	50.2	20.6
+Cluster	47.2	20.1
JEDS	**45.8**	**21.3**

Only integrate *filtering* for *clustering*

Only integrate *clustering* for *summarization*

Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

- Comparison with State-of-the-art
  - Comparison of clustering algorithms

State-of-the-art models for event clustering

Method	C_{min}
LSH	66.7
AS12	60.9
JEDS	**45.8**

  - Comparison of summarization algorithms

State-of-the-art model for event clustering and summarization

Method	ROUGE-1
AS12+LexRank	18.8
AS12+CL16	19.6
LSH+LexRank	17.2
LSH+CL16	19.1
JEDS	**21.3**

Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Joint Event Detection and Reporting

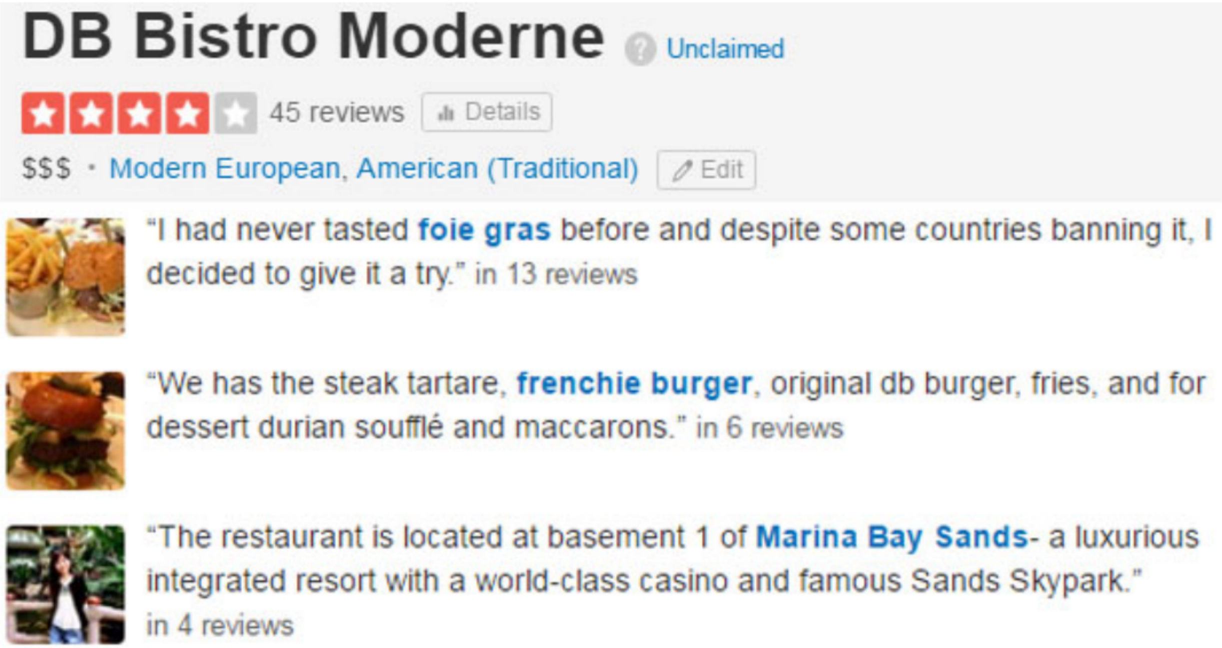
- Results on DDoS Domain
  - Comparison with state-of-the-art

Method	Clustering	Summarization
AS12+LexRank	64.4	15.5
LSH+CL16	57.8	16.5
JEDS	**38.3**	**18.7**

Wang, Zhongqing, et al. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*, 2017.

Opinion Recommendation

- A restaurant review on Yelp.com



Wang, Zhongqing, et al. "Opinion Recommendation Using A Neural Model." *EMNLP*, 2017.

Opinion Recommendation

- Opinion Recommendation: a novel task of jointly predicting a custom review with a rating score that a certain user would give to a certain product or service, given existing reviews and rating scores to the product or service by other users, and the reviews that the user has given to other products and services.

	Product 1	Product 2	Product 3
User A	Review + 3.0	Review + 4.5	Review + 4.5
User B		Review + 2.5	Review + 3.5
User C	Review + 4.0		Review + ?

Wang, Zhongqing, et al. "Opinion Recommendation Using A Neural Model." *EMNLP*, 2017.

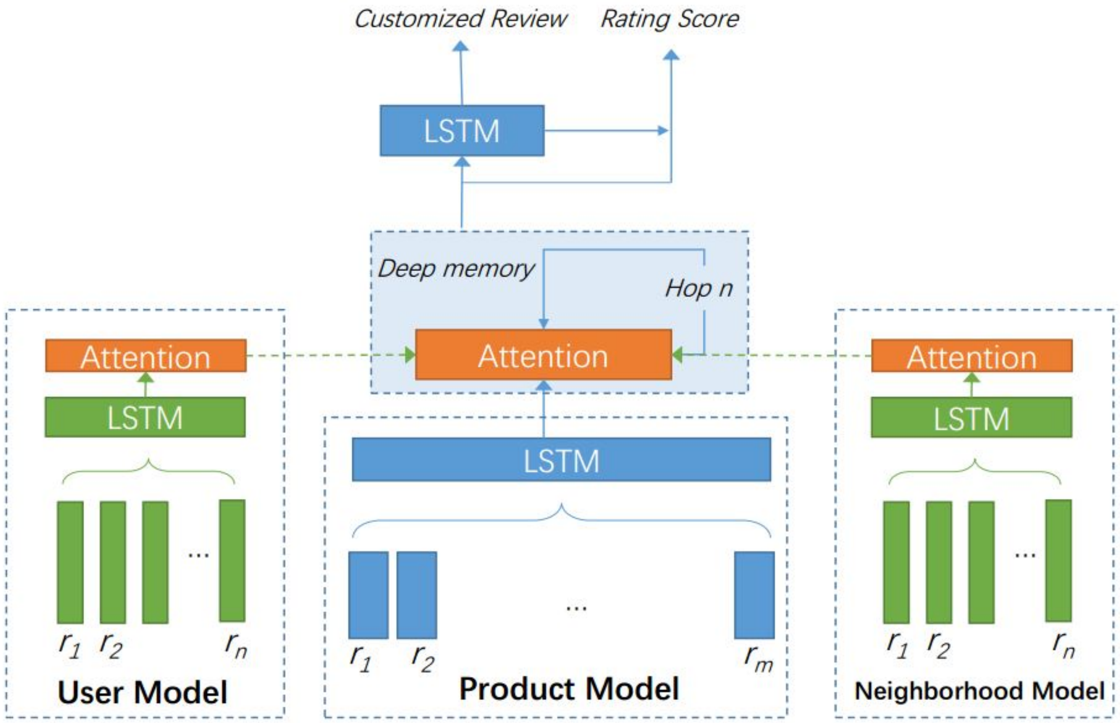
Opinion Recommendation

- This paper use a single neural network to model users and products, capturing their correlation and generating customised product representations using a deep memory network, from which customised ratings and reviews are constructed jointly.

Wang, Zhongqing, et al. "Opinion Recommendation Using A Neural Model." *EMNLP*, 2017.

Opinion Recommendation

- Overview of proposed model



Wang, Zhongqing, et al. "Opinion Recommendation Using A Neural Model." *EMNLP*, 2017.

Opinion Recommendation

- Experiments
  - Data: collected from the yelp academic dataset, provided by Yelp.com
  - Evaluation: use the ROUGE-1.5.5 toolkit for evaluating the performance of customized review generation, and report unigram overlap (ROUGE-1) as a means of assessing informativeness.; Mean Square Error (MSE) is used as the evaluation metric for measuring the performance of customized rating score prediction.

Wang, Zhongqing, et al. "Opinion Recommendation Using A Neural Model." *EMNLP*, 2017.

Opinion Recommendation

- Results

	Rating	Generation
RS-Average	1.280	-
RS-Linear	1.234	-
RS-Item	1.364	-
RS-MF	1.143	-
Sum-Opinosis	-	0.183
Sum-LSTM-Att	-	0.196
Joint	**1.023**	**0.250**

Wang, Zhongqing, et al. "Opinion Recommendation Using A Neural Model." *EMNLP*, 2017.

Joint Entity and Sentiment Extraction

- Open domain targeted sentiment is the joint information extraction task that finds target mentions together with the sentiment towards each mention from a text corpus.

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

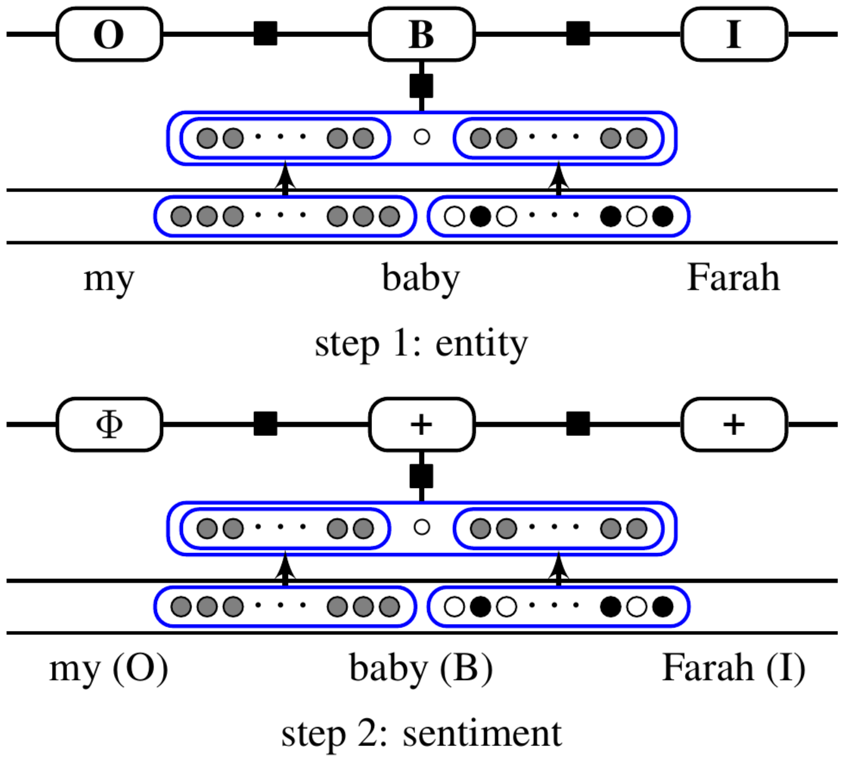
Joint Entity and Sentiment Extraction

- This paper
  - make an empirical comparison between discrete and neural CRF models, and further combine the strengths of each model via feature integration.
  - compare the effects of the pipeline, joint and collapsed models for open targeted sentiment analysis under the neural model settings.

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

Joint Entity and Sentiment Extraction

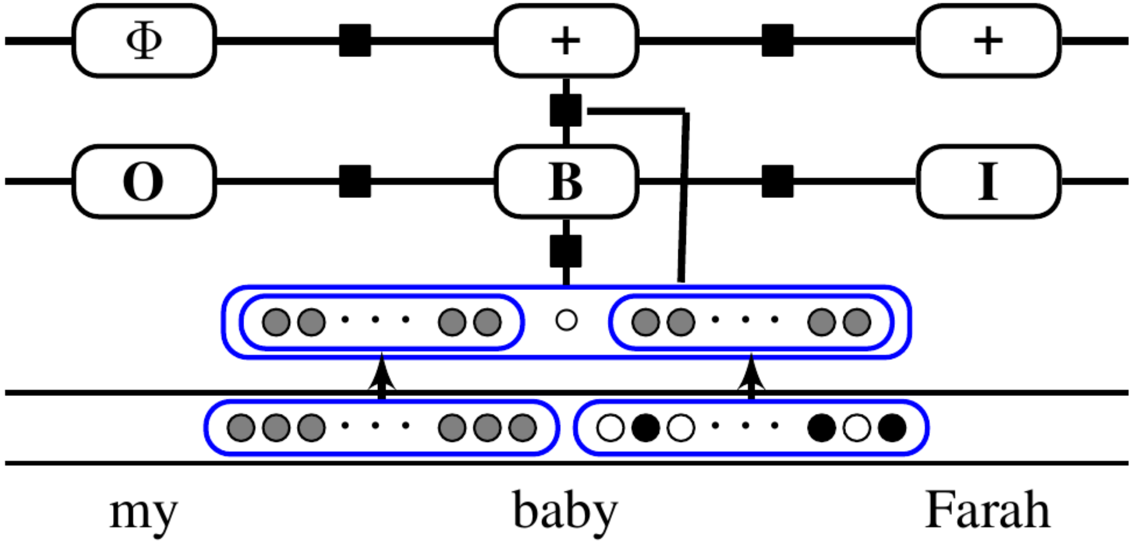
- Integrated models for pipeline



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

Joint Entity and Sentiment Extraction

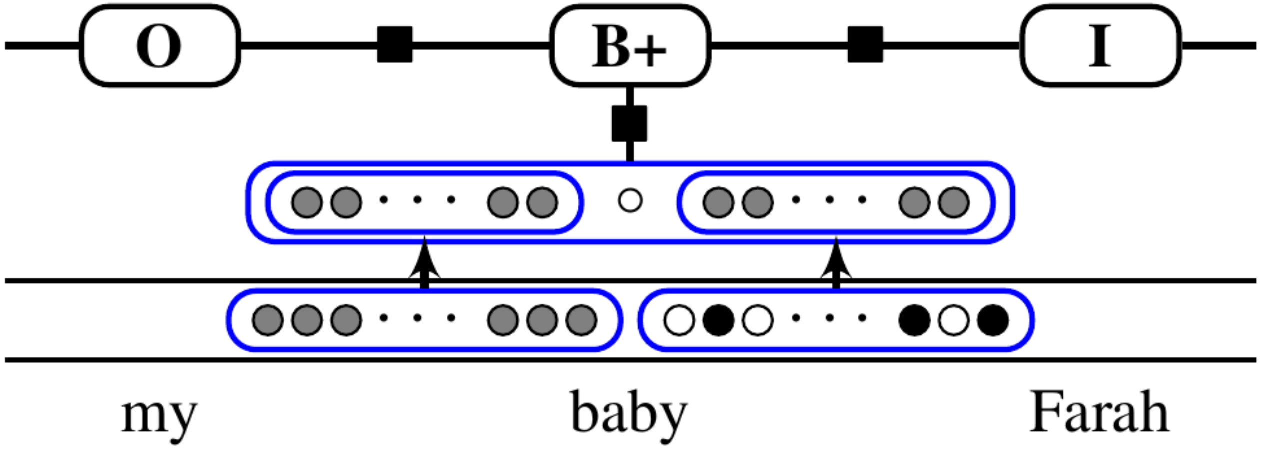
- Integrated models for joint



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

Joint Entity and Sentiment Extraction

- Integrated models for collapsed



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

Joint Entity and Sentiment Extraction

- Data of Mitchell et al. (2013)

Domain	#Sent	#Entities	#+	#-	#0
English	2,350	3,288	707	275	2,306
Spanish	5,145	6,658	1,555	1,007	4,096

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

Joint Entity and Sentiment Extraction

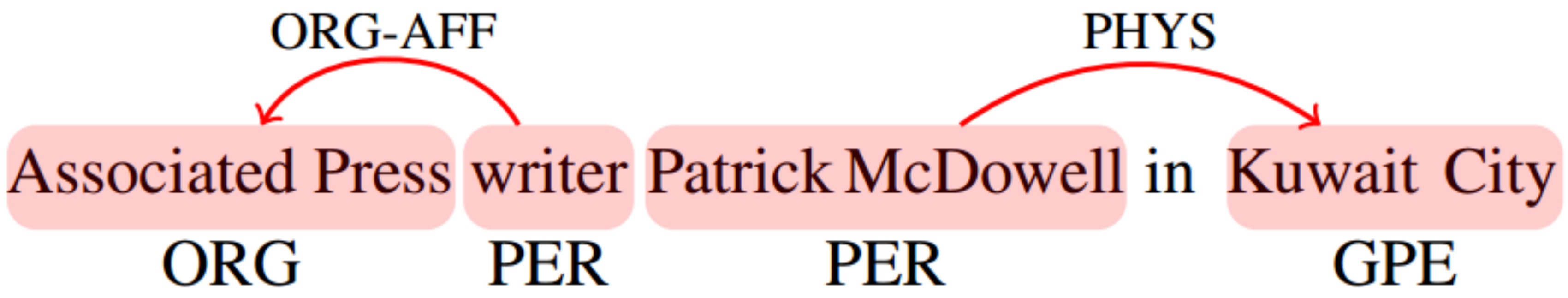
- Results

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.

Joint Entity and Relation Extraction

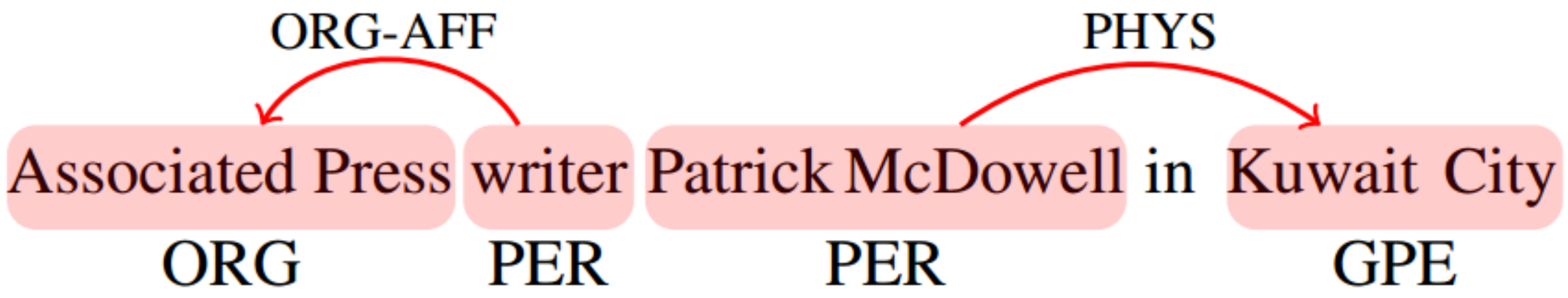
- Background
  - Relation Extraction



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

- Background
  - Relation Extraction

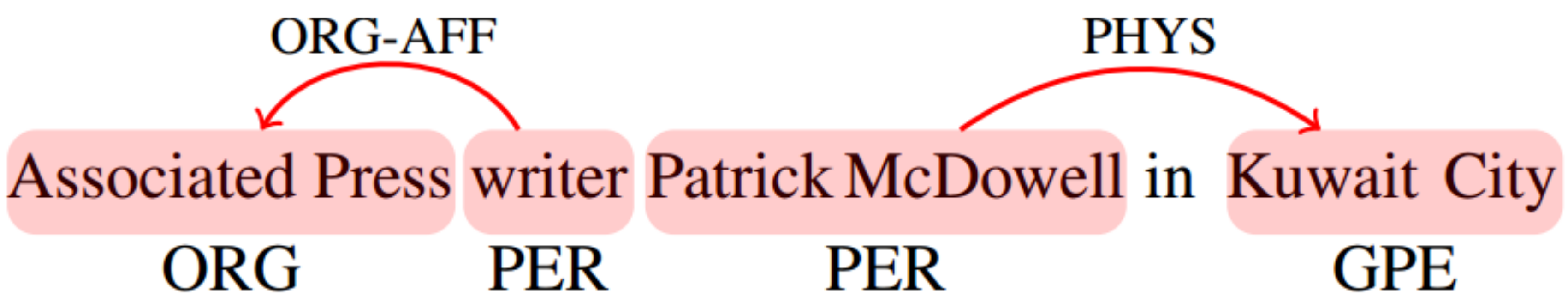


  - Entity Recognition
  - Relation Classification

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

- Background
  - Relation Extraction



  - Entity Recognition
  - Relation Classification

Single Model
Joint & End to End

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

- Background
  - Relation Extraction

Single Model (Joint & End to End)
Approach: Table Filling
Related work:
- **Miwa and Sasaki (2014)**
- **Miwa and Bansal (2016)**

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

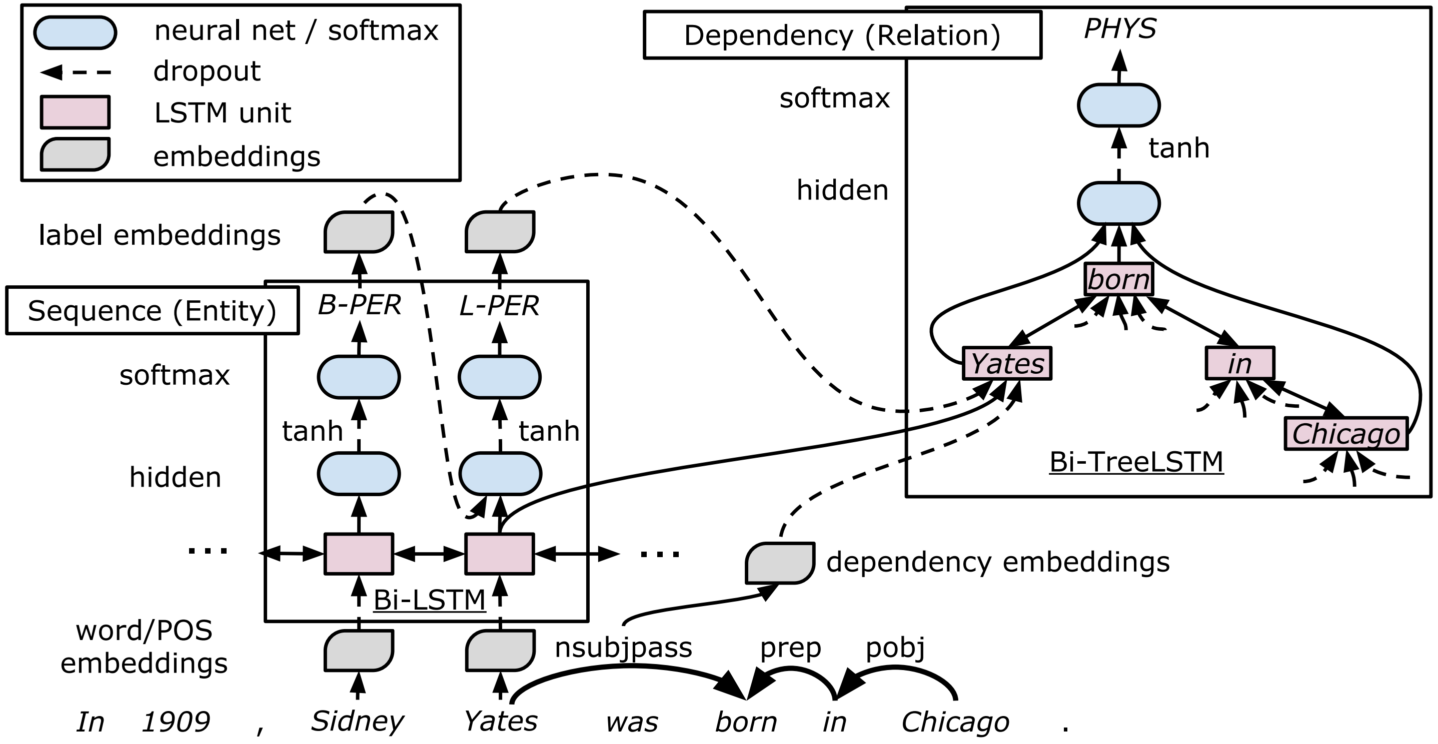
- Background
  - Relation Extraction
    - Table-Filling Sequence
    - Miwa and Bansal (2016)

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

- Background
  - Relation Extraction
    - Table-Filling Sequence
    - Miwa and Bansal (2016)



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

- Background
  - Relation Extraction
    - Table-Filling Sequence
    - Miwa and Bansal (2016)

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

- This paper build a globally optimized neural model for end-to-end relation extraction, proposing novel LSTM features in order to better learn context representation. In addition, this paper present a novel method to integrate syntactic information to facilitate global learning, yet requiring little background on syntactic grammars thus being easy to extend

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

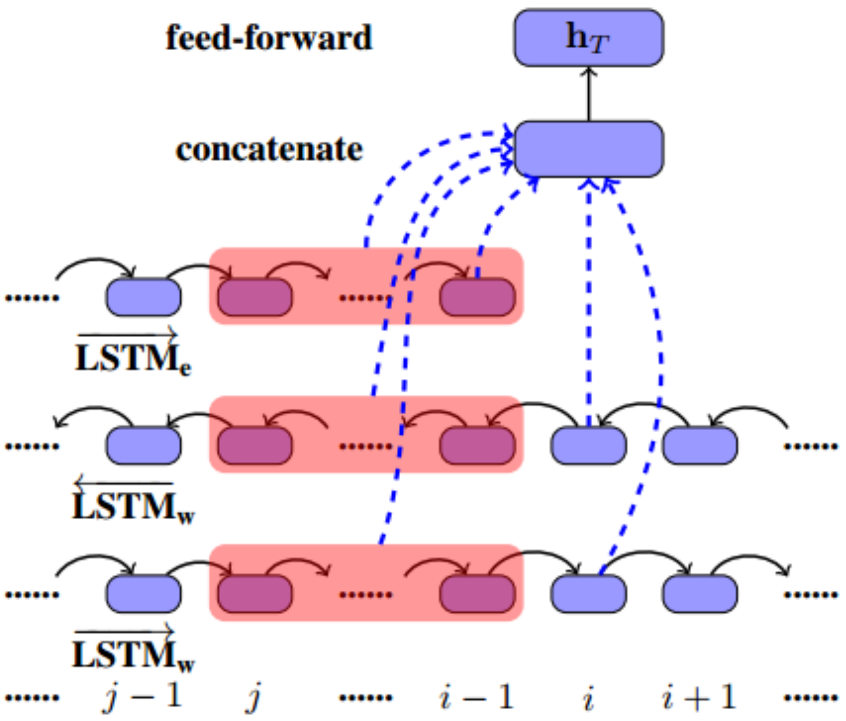
Joint Entity and Relation Extraction

- Our Contributions
  - Beam Search with Global Learning
  - Novel Syntactic Features
    - Without any background on syntactic grammars

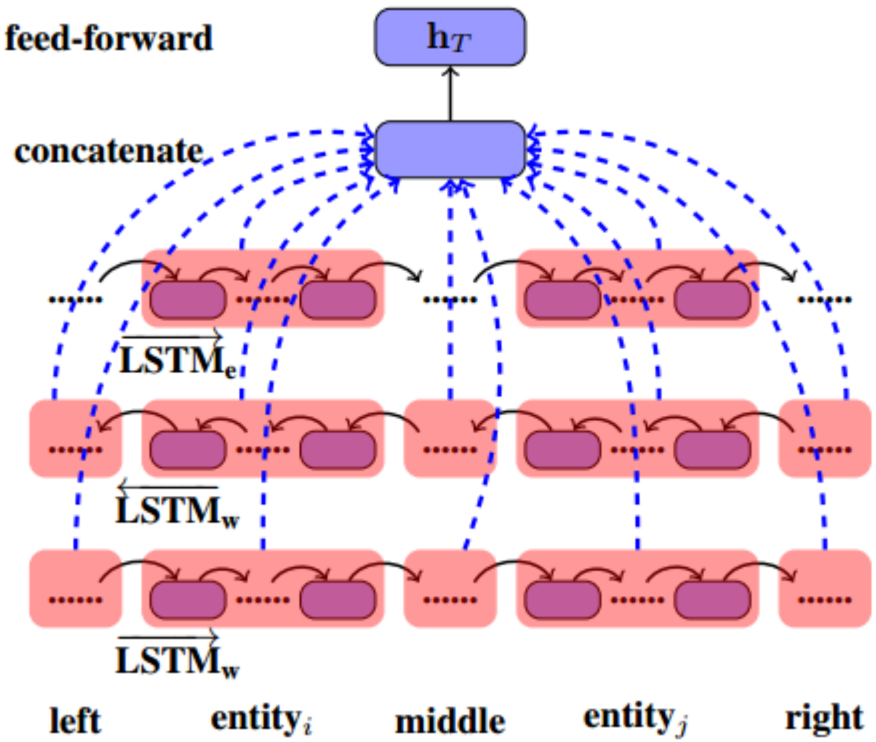
Zhang, Meishan, et al. "End-to-End Neural Relation Extraction with Global Optimization." *EMNLP*, 2017.

Joint Entity and Relation Extraction

- Baseline



6 features



12 features

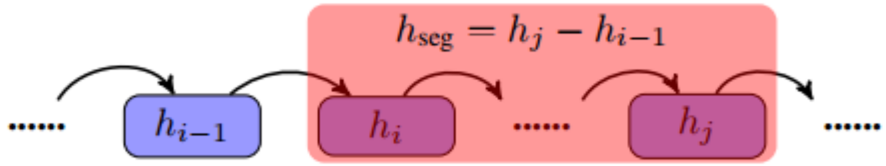
Zhang, Meishan, et al. "End-to-End Neural Relation Extraction with Global Optimization." *EMNLP*, 2017.

Joint Entity and Relation Extraction

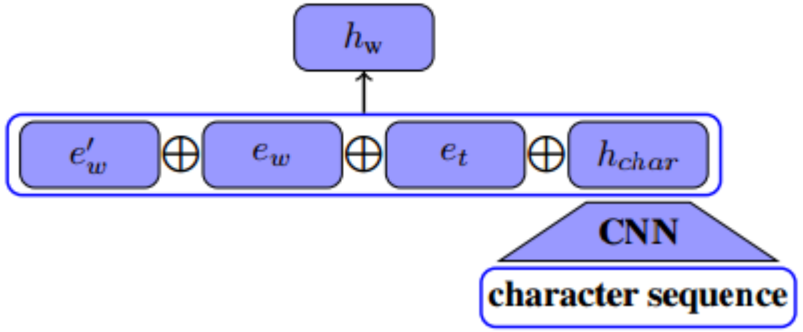
- Baseline

Details

segment representation



word representation



Joint Entity and Relation Extraction

- Baseline
  - Classification
    - Greedy Search
    - Objective

$$\mathrm{loss}(T, l_i^g, \Theta) = -\log p_{l_i^g}$$

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

Joint Entity and Relation Extraction

- Beam Search

Algorithm 1 Beam-search.

```
agenda ← { (empty table, score=0.0) }
for i in 1 ··· max-step
    next_scored_tables ← { }
    for scored_table in agenda
        labels ← NEXTLABELS(scored_table)
        for next_label in labels
            new ← FILL(scored_table, next_label)
            ADDITEM(next_scored_tables, new)
    agenda ← TOP-B(next_scored_tables, B)
```

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

Joint Entity and Relation Extraction

- Beam Search

Local: classification

$$\text{loss}(T, l_i^g, \Theta) = -\log p_{l_i^g}$$

Global: beam search

$$\text{loss}(x, T_i^g, \Theta) = -\log p_{T_i^g} = -\log \frac{\text{score}(T_i^g)}{\sum_{T_i'} \text{score}(T_i')}$$

$$\text{score}(T_i) = \sum_{j=0}^{i} \text{score}(T_{j-1}, l_j)$$

Joint Entity and Relation Extraction

- Comparative Experiments(ACE05 dataset，development dataset)

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.

Joint Entity and Relation Extraction

- Syntactic Features
  - Why not dependency path?
    - many paths caused dynamic outputting entities
    - requiring background on dependency grammar

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

Joint Entity and Relation Extraction

- Syntactic Features
  - Encoder-Decoder Framework
    - Encoder : Sentence Representation
      - Usually Bi-LSTM(multi-layer)
    - Decoder : Parsing Decoding
      - Transition-based, Graph-based or other

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

Joint Entity and Relation Extraction

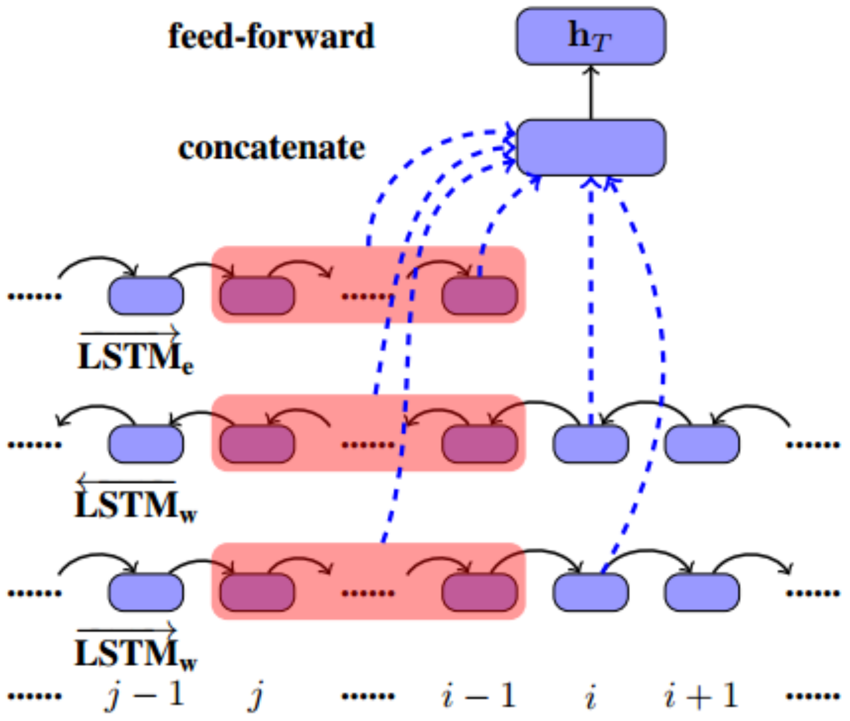
- Syntactic Features
  - Encoder-Decoder Framework
    - Encoder : Sentence Representation
      - Usually Bi-LSTM(multi-layer)
    - Decoder : Parsing Decoding
      - Transition-based, Graph-based or

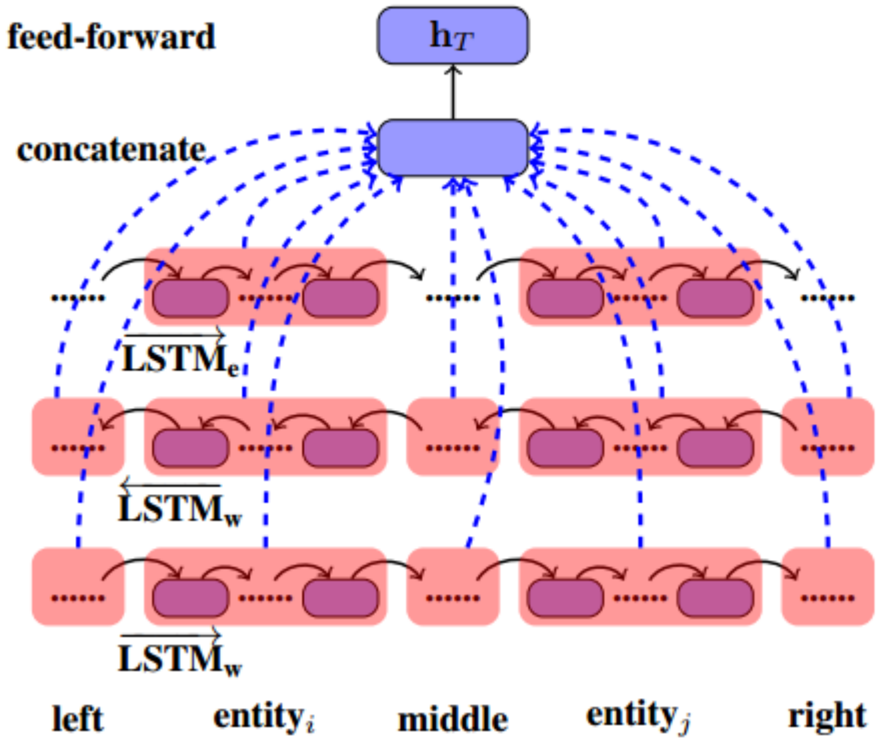
Simply dumping and build lstms based on the output!

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

Joint Entity and Relation Extraction

- Syntactic Features





6 features → 10 features **12 features → 22 features**

Joint Entity and Relation Extraction

- Syntactic Features
  - Comparative Experiments(ACE05 dataset，development dataset)

Model	Features	Entity F1	Relation F1
Local	all	**81.6**	**53.0**
	-syn	81.5	50.9
Global	all	**81.9**	**54.2**
	-syn	81.6	52.6

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

Joint Entity and Relation Extraction

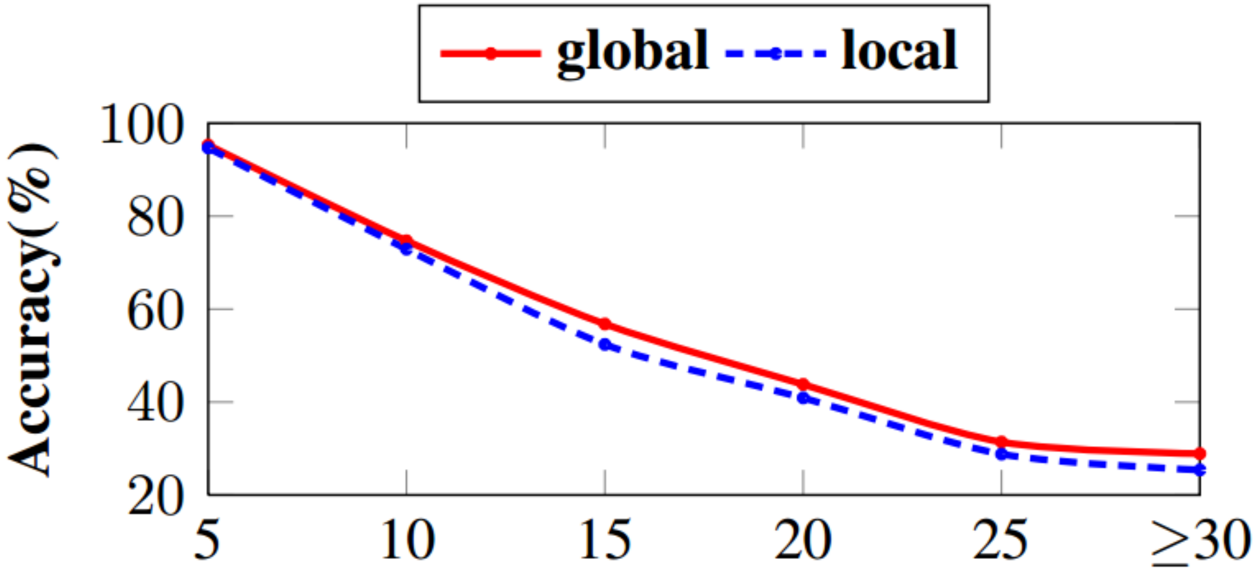
- Final Results
  - Comparative Experiments(test dataset)

model	ACE05		CONLL04	
	Entity	Relation	Entity	Relation
Our Model	**83.6**	**57.5**	**85.6**	**67.8**
M&B (2016)	83.4	55.6	—	—
L&J (2014)	80.8	49.5	—	—
M&S (2014)	—	—	80.7	61.0

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

Joint Entity and Relation Extraction

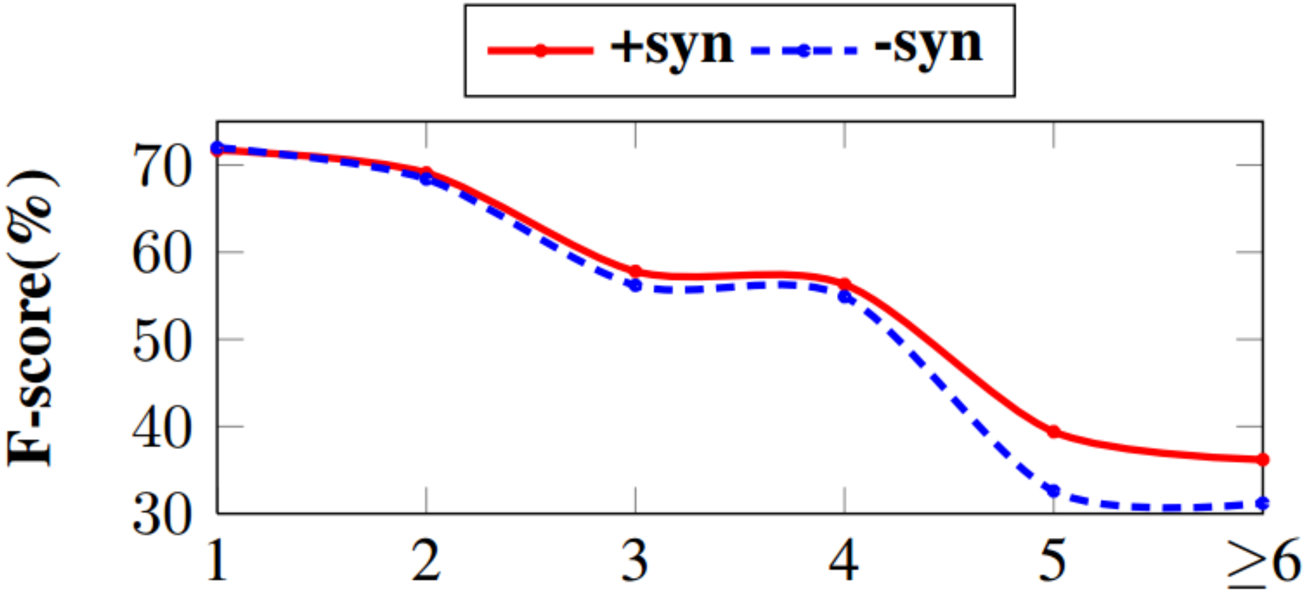
- Analysis
  - Global Learning



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

Joint Entity and Relation Extraction

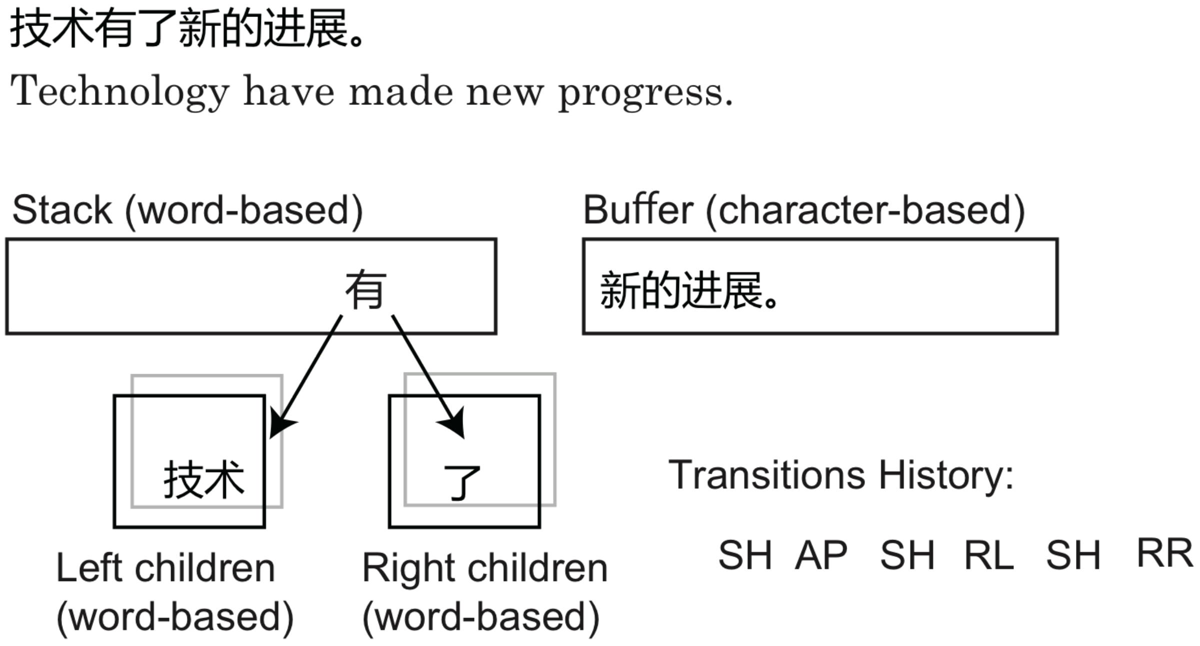
- Analysis
  - Syntactic Feature (Relation)



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

Joint Word Segmentation, POS tagging and Dependency Parsing

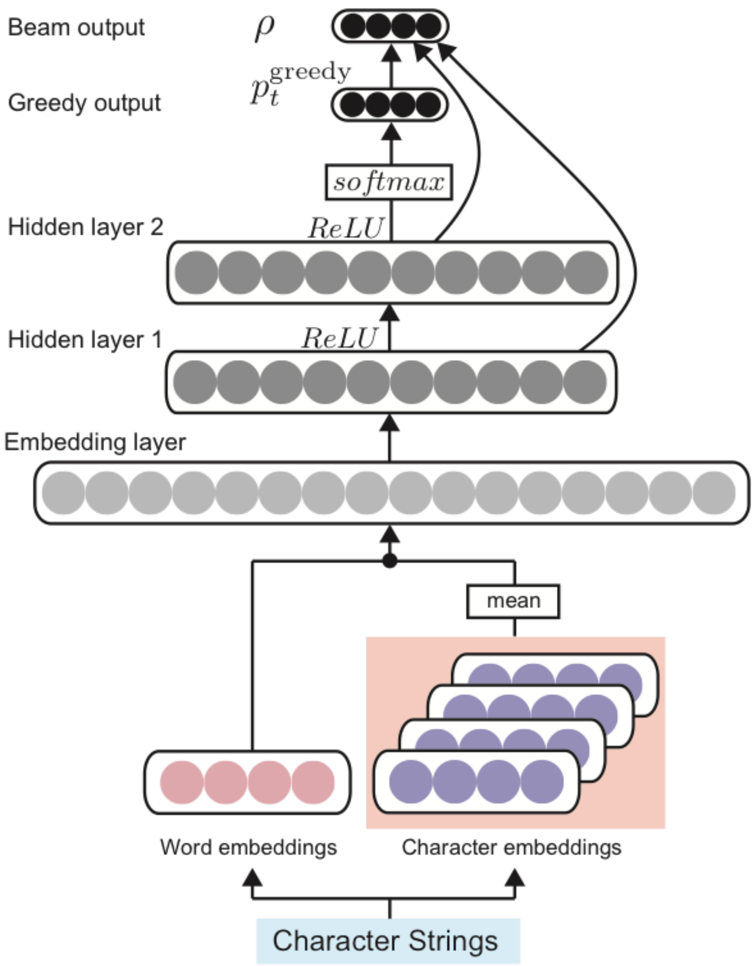
- Model



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

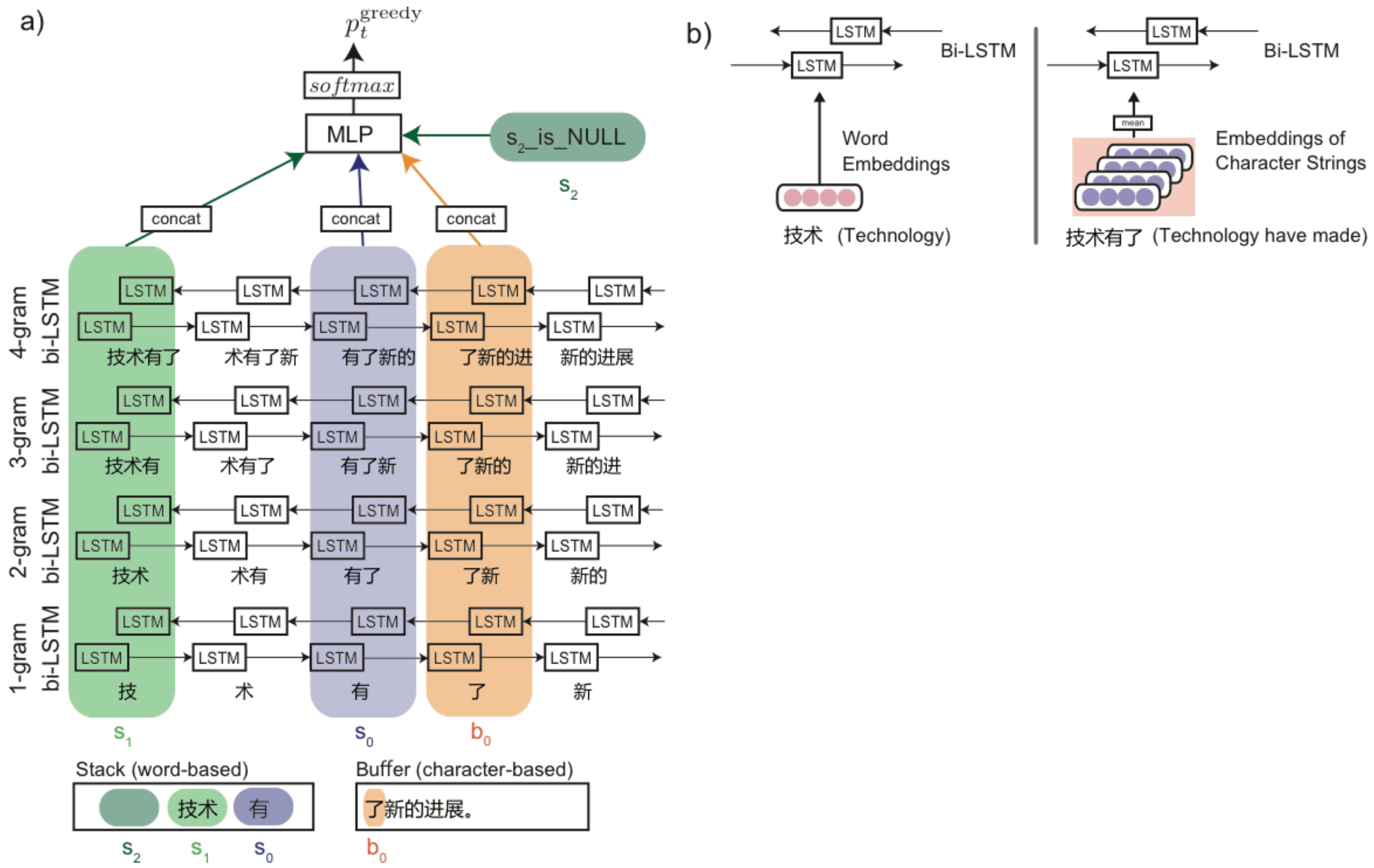
- Feed-forward NN model



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

- The SegTag+Dep model

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 Word Segmentation, POS tagging and Dependency Parsing

- Bi-LSTM feature extraction model

Model	Seg	POS	Dep
Hatori+12	97.75	94.33	81.56
M. Zhang+14 EAG	97.76	94.36	81.70
SegTagDep (g)	98.24	94.49	80.15
Bi-LSTM 4feat.(g)	97.72	93.12	79.03
Bi-LSTM 8feat.(g)	97.70	93.37	79.38

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.

Other instances of multitask learning

- Cross-Lingual
- Cross-Standard

Cross-Lingual

- Motivation
  - **Singlish** is one of the major **creole** languages and has been increasingly used in written forms on web media.
  - **Little** NLP research has been focused on the creoles and **poor performance** on Singlish using English POS taggers and dependency parsers.

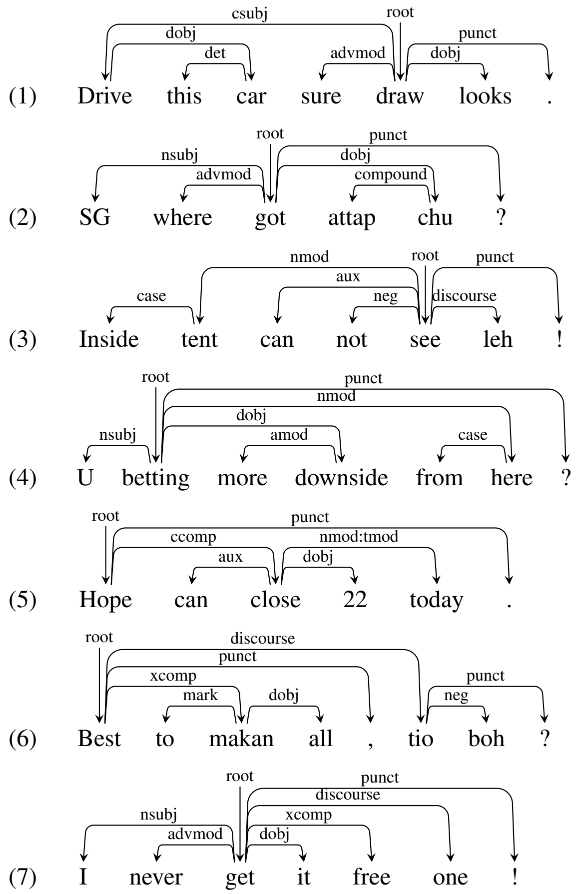
Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

Cross-Lingual

- Singlish Dependency Treebank
  - **Lexical Differences**: Extensive vocabularies borrowed from major local languages including Malay, Tamil, and Chinese dialects such as Hokkien, Cantonese and Teochew.
  - **Grammatical Variations**: 5 syntactical constructions. Topic Prominence (1-3) ; Copula Deletion (4) ; NP Deletion (5) ; Inversion (6) ; Discourse Particles (3,7)
  - **Universal Dependencies**: Cross-lingual consistency that facilitates transfer-learning for multilingual parsers.

Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

Cross-Lingual



Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

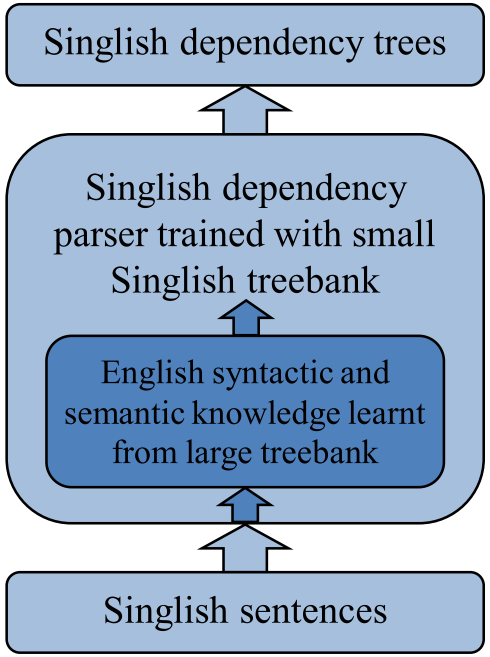
Cross-Lingual

- Knowledge Transfer using Neural Stacking
  - **English basic syntax** : state-of-the-art neural dependency parser with biaffine attentions (Dozat and Manning, 2017)
  - **Singlish specific syntax**: stacked neural layers capturing unique syntactical constructions (Chen et al., 2016)

Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

Cross-Lingual

- Knowledge Transfer using Neural Stacking



Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

Cross-Lingual

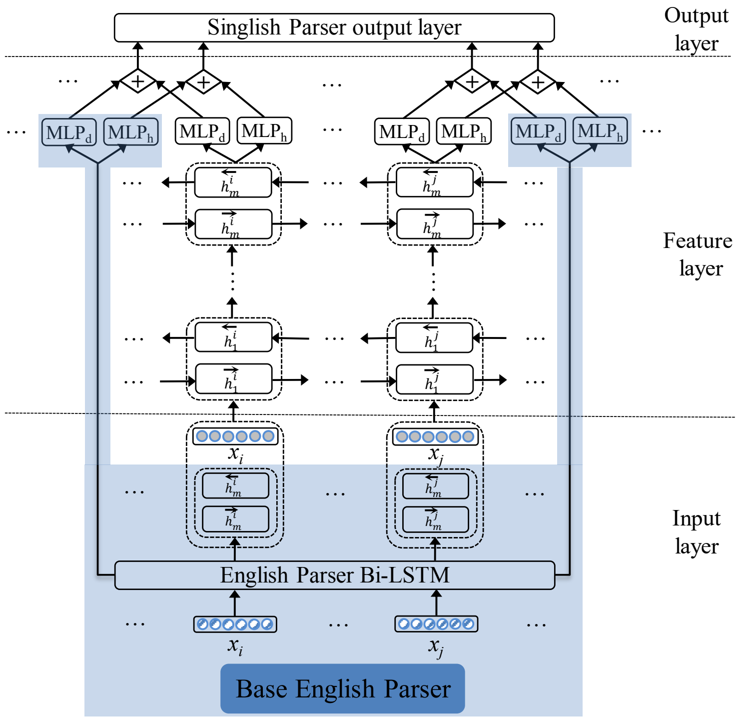
- Neural Stacking Parser with Biaffine Attentions
  - Distributed lexical semantics encoded in pre-trained word embeddings trained on English and Singlish respectively
  - **Feature level neural stacking** by concatenations of word embedding with last bi-LSTM layer from the base model

---

Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

Cross-Lingual

- Neural Stacking Parser with Biaffine Attentions



Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

Cross-Lingual

- **Significant improvement using neural stacking** over the state-of-the-art dependency parser (Dozat and Manning, 2017) trained on English, Singlish and their combination.

Trained on	System	UAS	LAS
English	ENG-on-SIN	75.89	65.62
	Baseline	75.98	66.55
Singlish	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**

Table 4: Dependency parser performances

Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

Cross-Lingual

- **Consistent improvements over all grammar types** by successful incorporation of English knowledge.

	Topic Prominence		Copula Deletion		NP Deletion		Discourse Particles		Others	
Sentences	15		19		21		51		67	
	UAS	LAS	UAS	LAS	UAS	LAS	UAS	LAS	UAS	LAS
ENG-on-SIN	78.15	62.96	66.91	56.83	72.57	64.00	70.00	59.00	78.92	68.47
Base-Giga100M	77.78	68.52	71.94	61.15	76.57	69.14	85.25	77.25	73.13	60.63
Base-ICE	81.48	72.22	74.82	63.31	**80.00**	73.71	85.25	77.75	75.56	64.37
Stack-ICE	**87.04**	**76.85**	**77.70**	**71.22**	**80.00**	**75.43**	**88.50**	**83.75**	**84.14**	**76.49**

Table 6: Error analysis with respect to grammar types

Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

Cross-Lingual

- Contributions
    - Annotation of a Singlish dependency treebank of 10,986 words using Universal Dependencies and POS tags.
    - Application of neural stacking for knowledge transfer to enhance POS tagging and dependency parsing for Singlish.

Wang, Hongmin, et al. "Universal Dependencies Parsing for Colloquial Singaporean English." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2017.

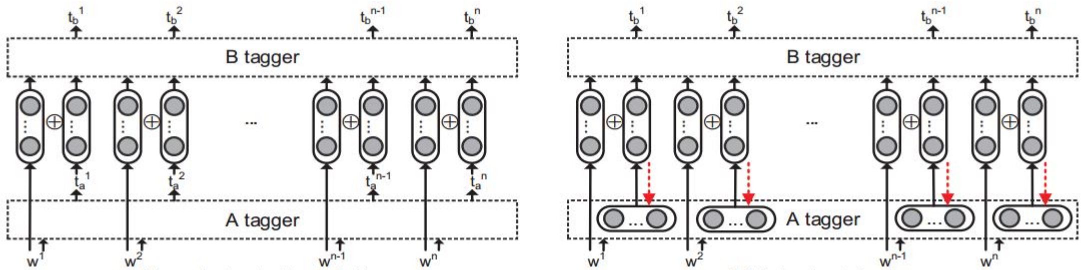
Cross-Standard

- This paper empirically investigate heterogeneous annotations using neural network models, building a neural network counterpart to discrete stacking and multi-view learning, respectively, finding that neural models have their unique advantages thanks to the freedom from manual feature engineering.
- CTB standard
- PD standard

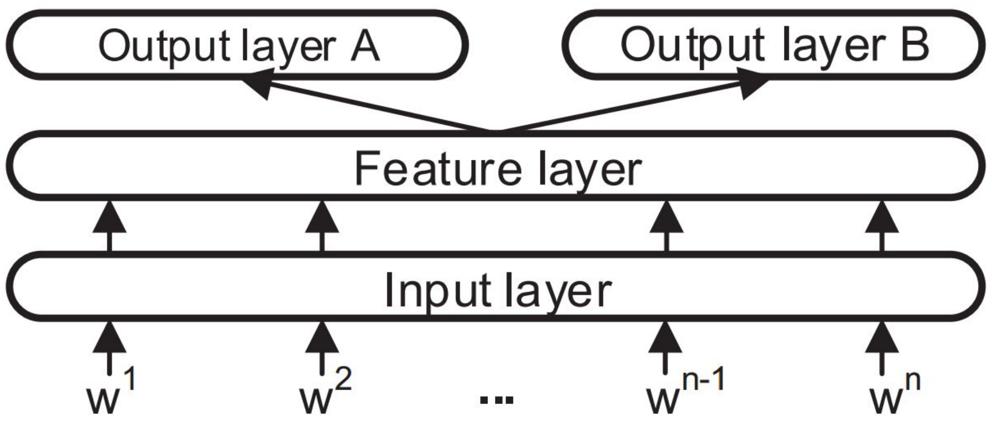
Chen, Hongshen, Yue Zhang, and Qun Liu. "Neural Network for Heterogeneous Annotations." *EMNLP*. 2016.

Cross-Standard

- Neural Stacking and Neural multi-view Model



(a) one-best-output level stacking (b) feature level stacking



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