A Shift-Reduce Parsing Algorithm for Phrase-based String-to-Dependency Translation

Yang Liu

Tsinghua University

SMT: Phrase Vs. Syntax

- Modern SMT approaches can be roughly divided into two broad categories:
  - phrase-based [Koehn et al., 2003; Och and Ney, 2004]
  - syntax-based
    - string-to-string [Wu, 1997; Chiang, 2007]
    - string-to-tree [Galley et al., 2006; Shen et al., 2008]
    - tree-to-string [Liu et al., 2006; Huang et al., 2006]
    - tree-to-tree [Eisner et al., 2003; Quirk et al., 2005]

SMT: Phrase Vs. Syntax

- Phrase-based
  - pros: efficient to integrate *n*-gram LM
  - cons: reordering is hard
- Syntax-based
  - pros: linguistically-motivated reordering
  - cons: expensive to integrate *n*-gram LM

SMT: Phrase Vs. Syntax

- Phrase-based
  - pros: efficient to integrate *n*-gram LM
  - cons: reordering is hard
- Syntax-based
  - pros: linguistically-motivated reordering
  - cons: expensive to integrate *n*-gram LM

Is it possible to combine the advantages of both?

Related Work

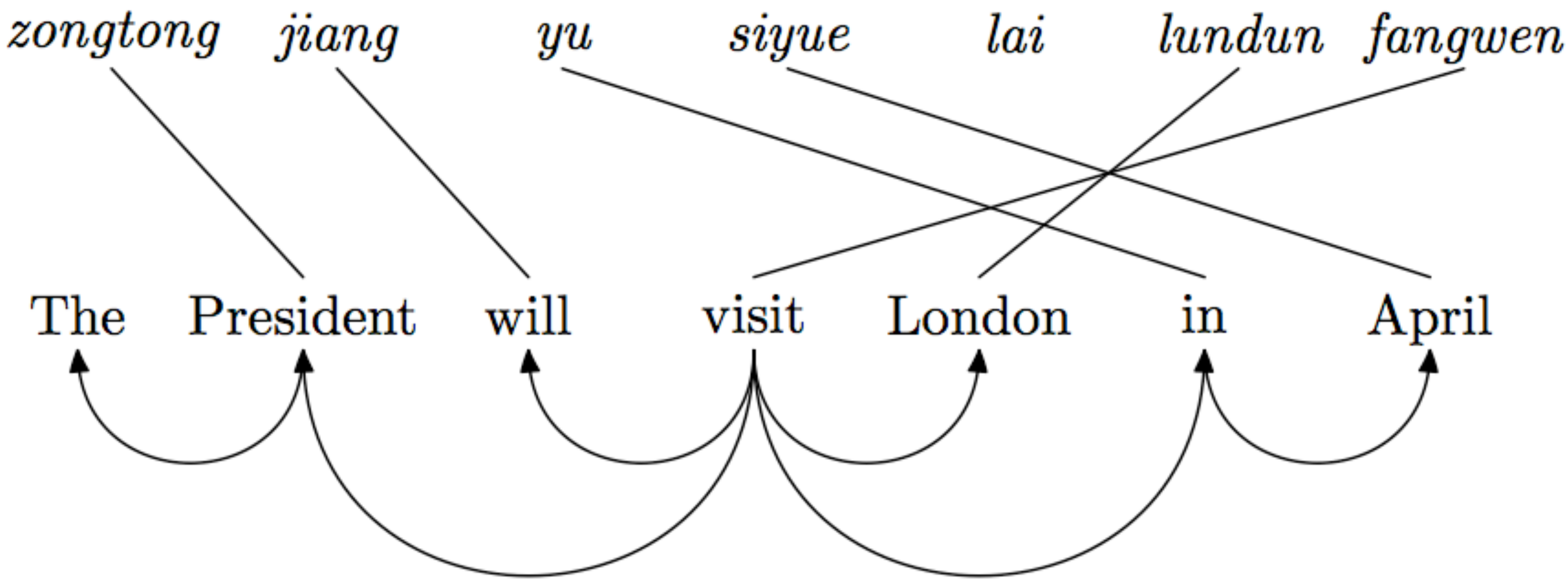
- Adding syntax to phrase-based decoding
  - hierarchical phrase reordering [Galley and Manning, 2008]
  - quadratic-time dependency parsing [Galley and Manning, 2009]
  - shift-reduce parsing for phrase-based models [Feng et al., 2010]
- incremental decoding for syntax-based models
  - left-to-right generation for SCFG [Watanabe et al., 2006]
  - incremental decoding for tree-to-string translation [Huang and Mi, 2010]
  - context-free reordering, finite-state translation [Dyer and Resnik, 2010]
  - incremental decoding with prediction [Feng et al., 2012]

This Work

- This work tries to combine the advantages of phrase-based and string-to-dependency models.

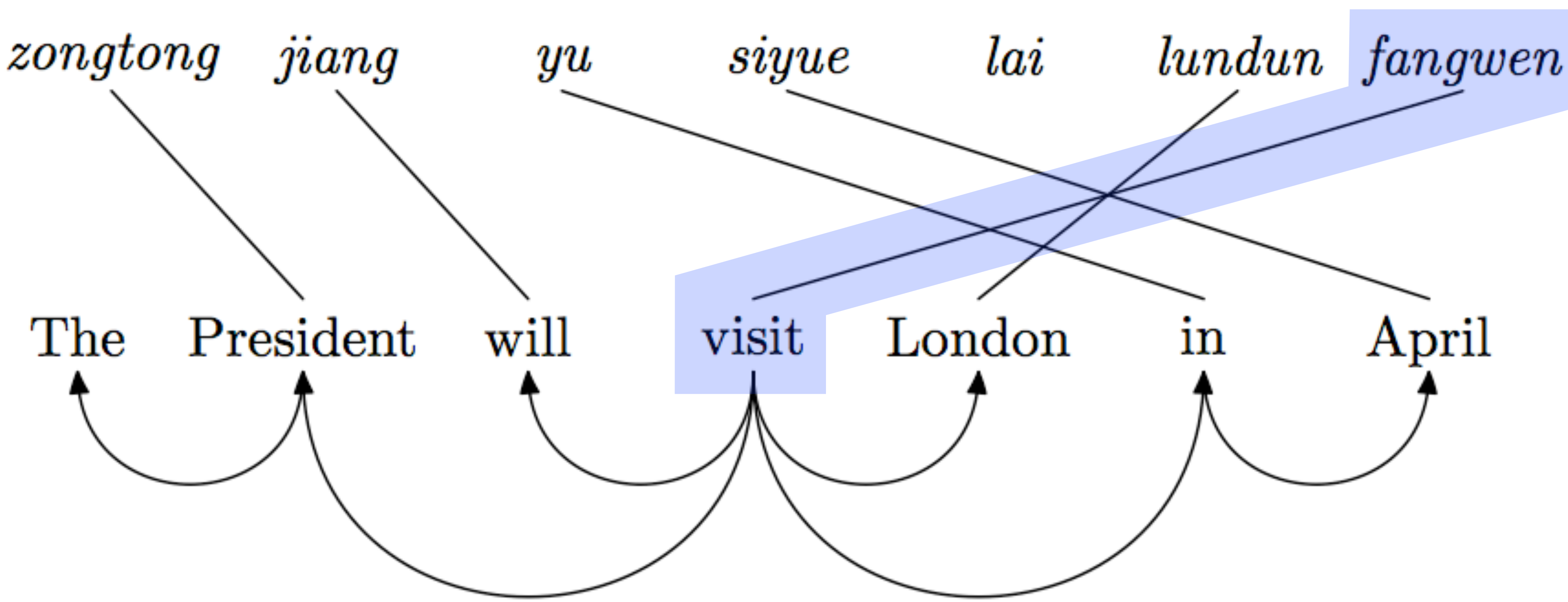
	phrase (Koehn et al., 2003)	s2d (Shen et al., 2008)	this work
rule table size	compact	large	compact
rule table coverage	high	low	high
n-gram LM	efficient	expensive	efficient
dep LM	N/A	yes	yes

Rule Extraction



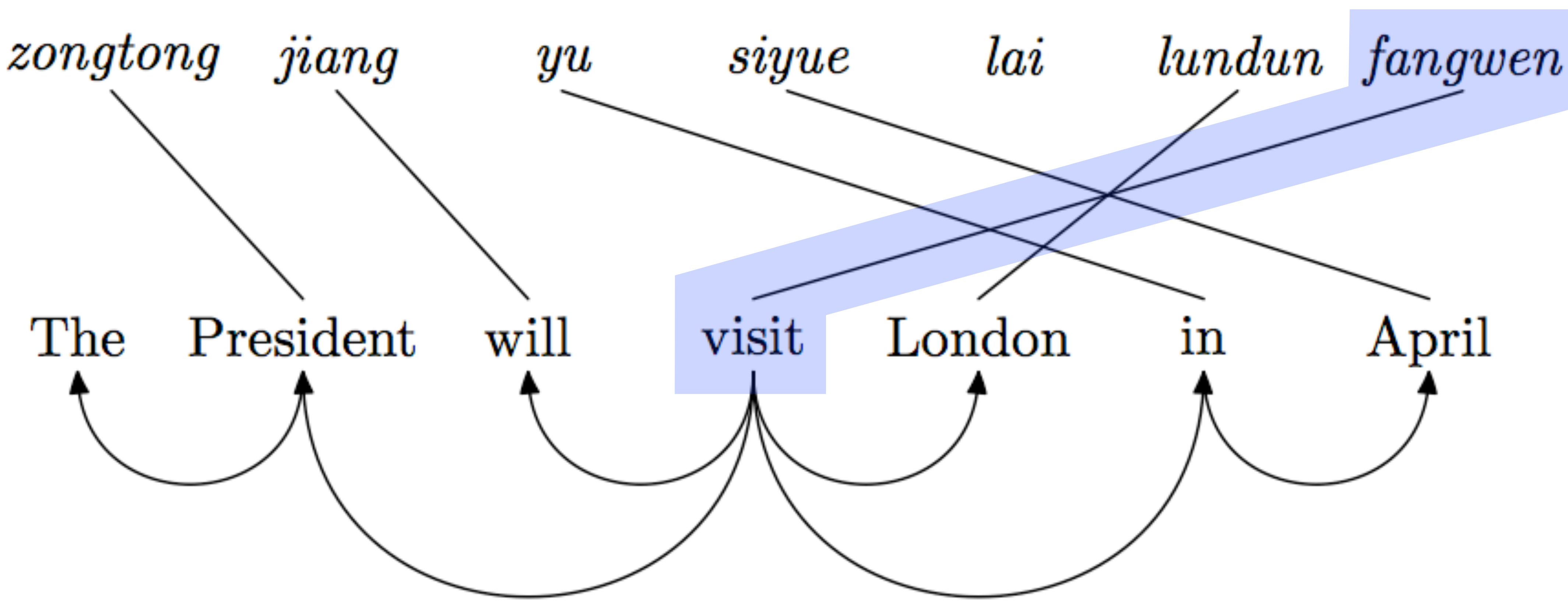
(Shen et al., 2008)

Rule Extraction



(Shen et al., 2008)

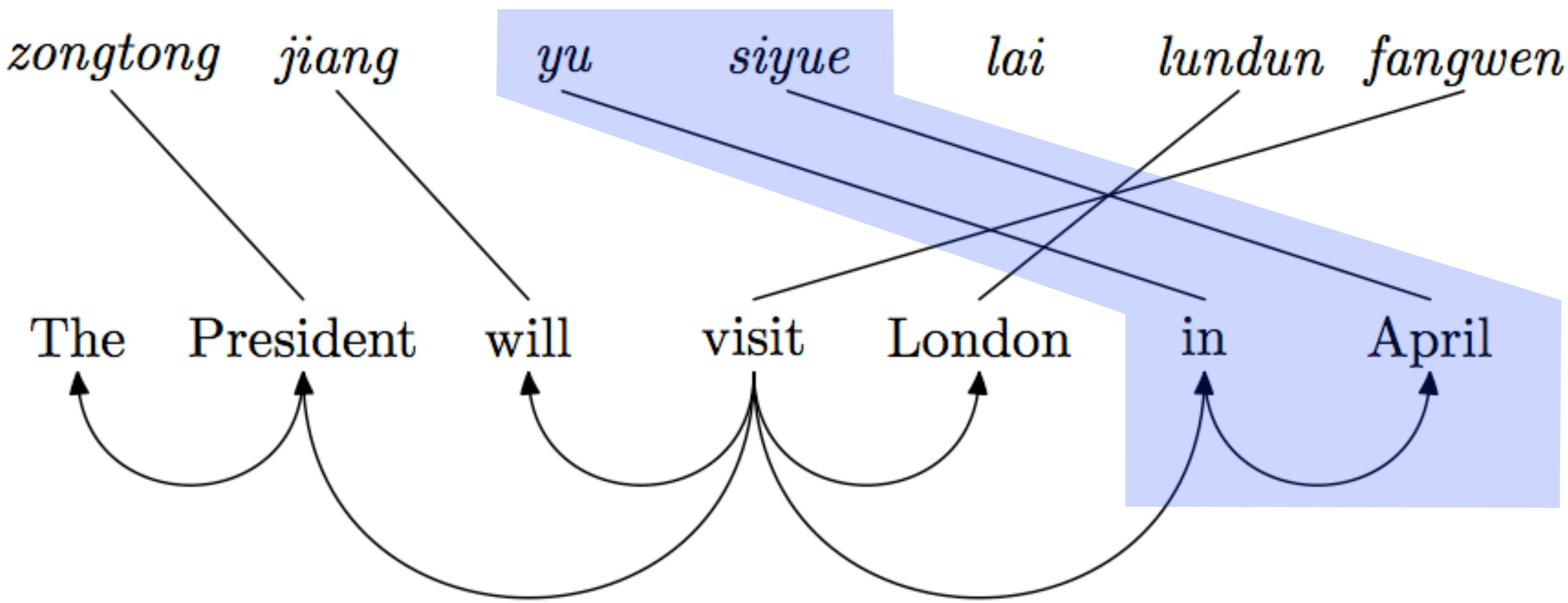
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	{}	fixed

(Shen et al., 2008)

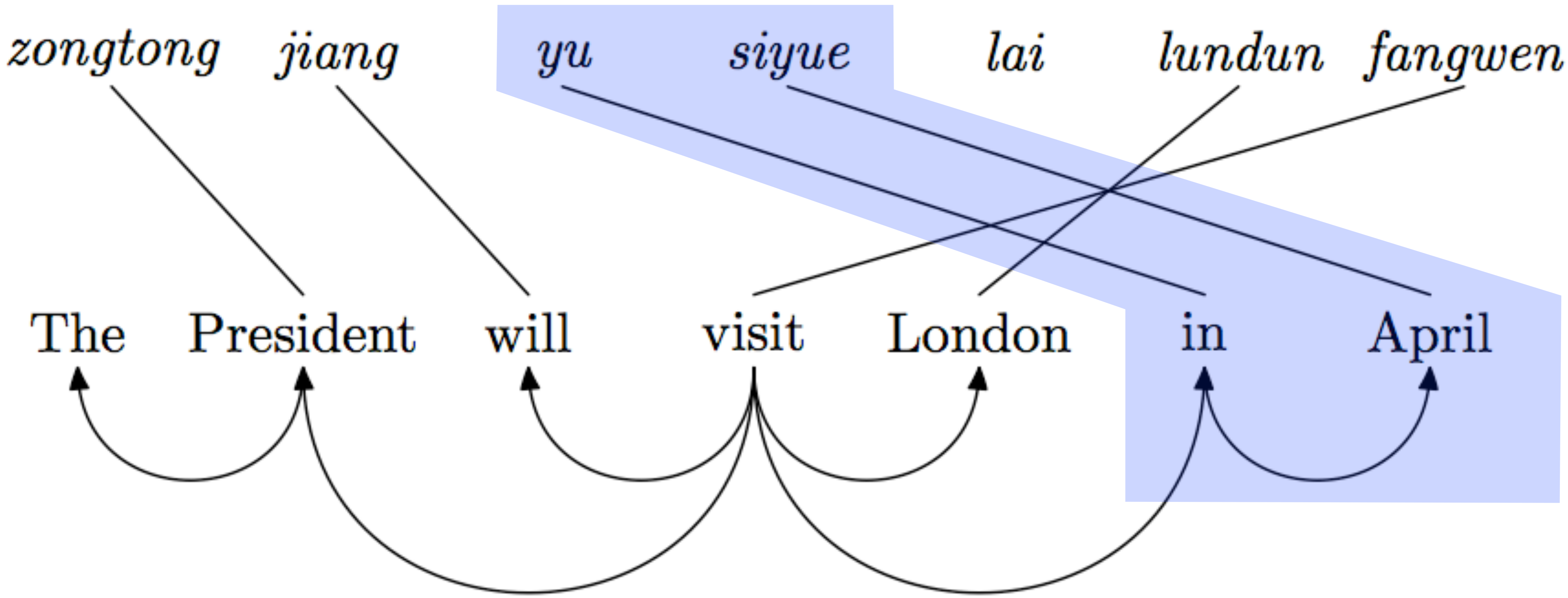
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	{}	fixed

(Shen et al., 2008)

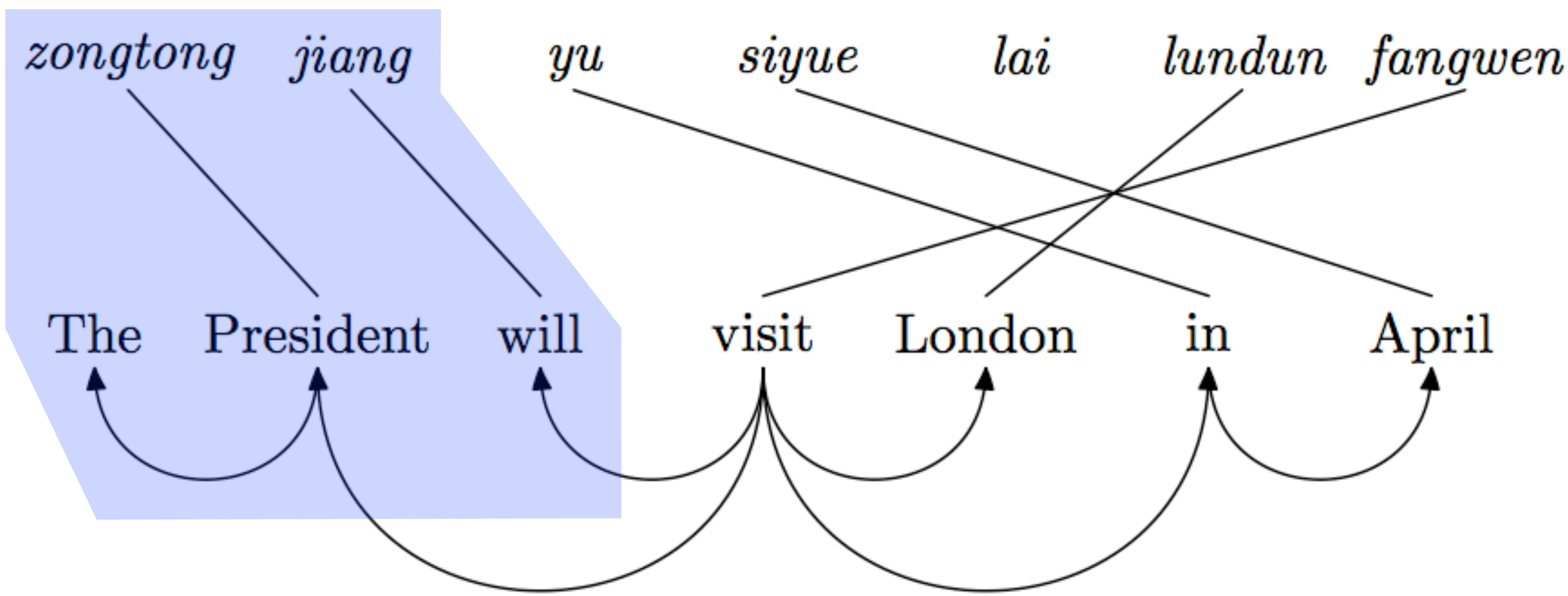
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	$\{\}$	fixed
r_2	*yu siyue*	in April	$\{1 \rightarrow 2\}$	fixed

(Shen et al., 2008)

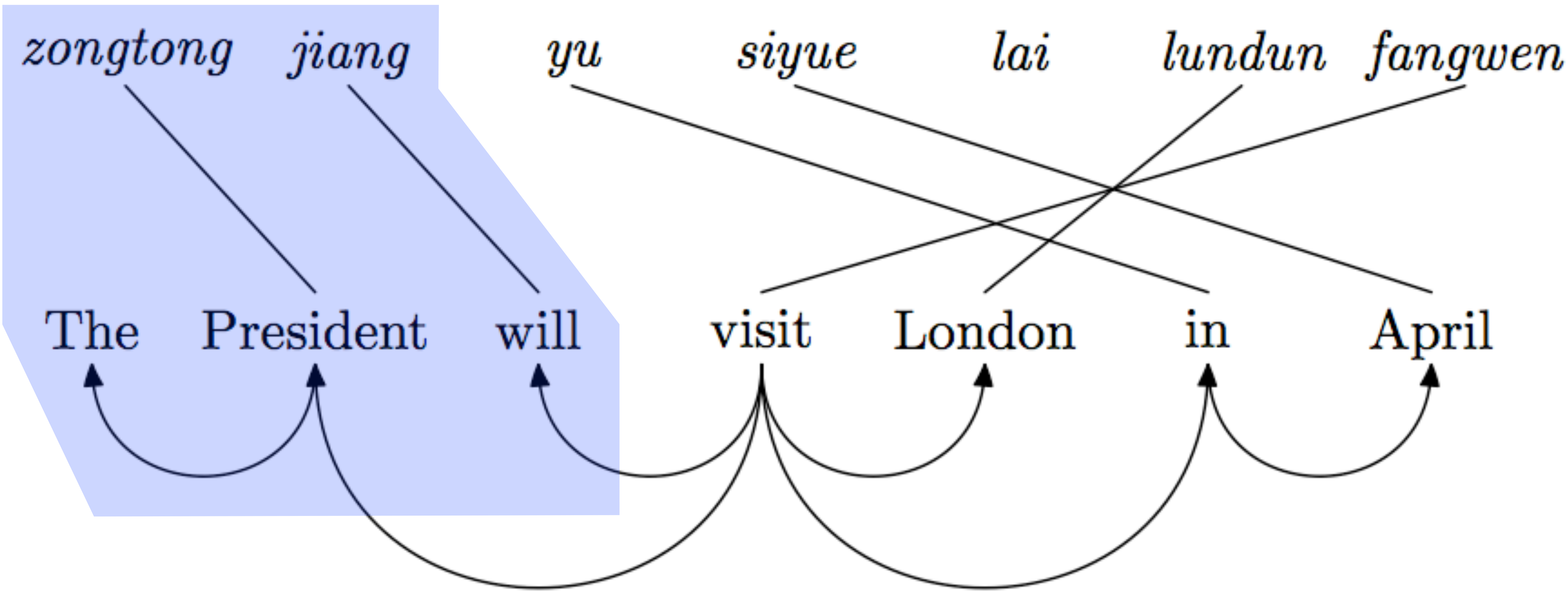
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	{}	fixed
r_2	*yu siyue*	in April	$\{1 \rightarrow 2\}$	fixed

(Shen et al., 2008)

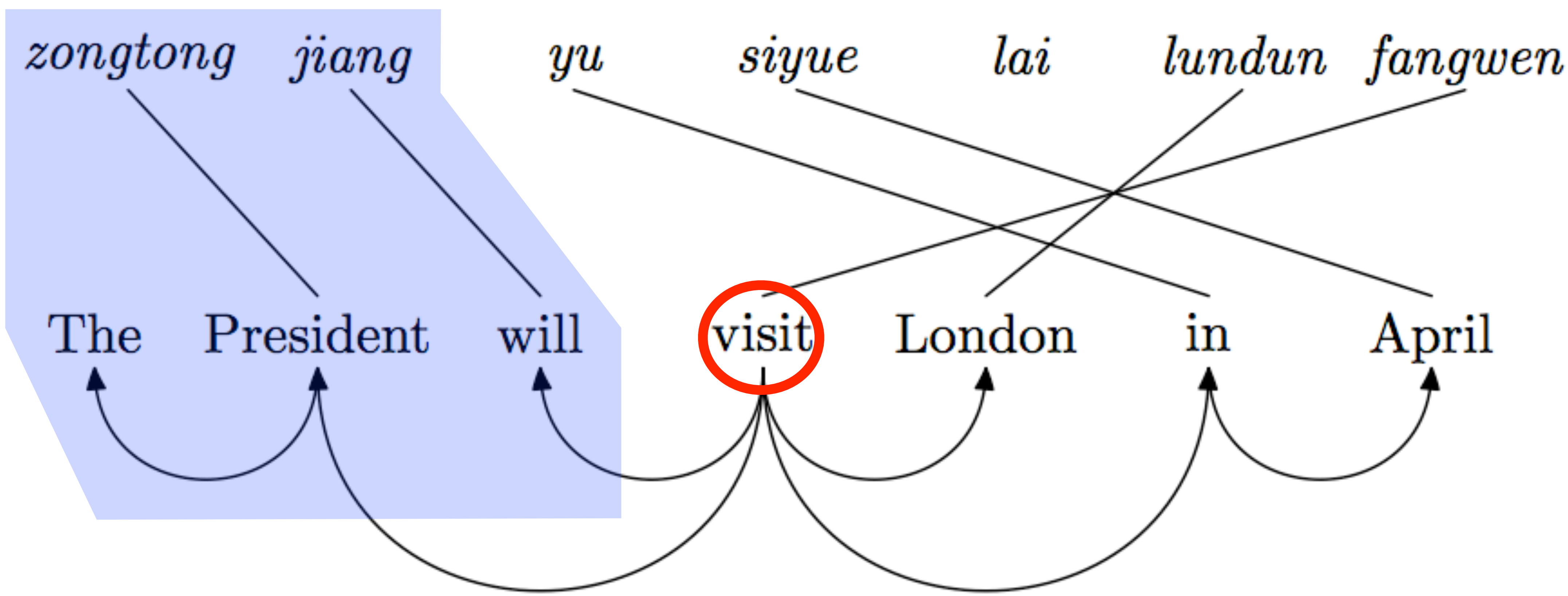
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	$\{\}$	fixed
r_2	*yu siyue*	in April	$\{1 \rightarrow 2\}$	fixed
r_3	*zongtong jiang*	The President will	$\{2 \rightarrow 1\}$	floating left

(Shen et al., 2008)

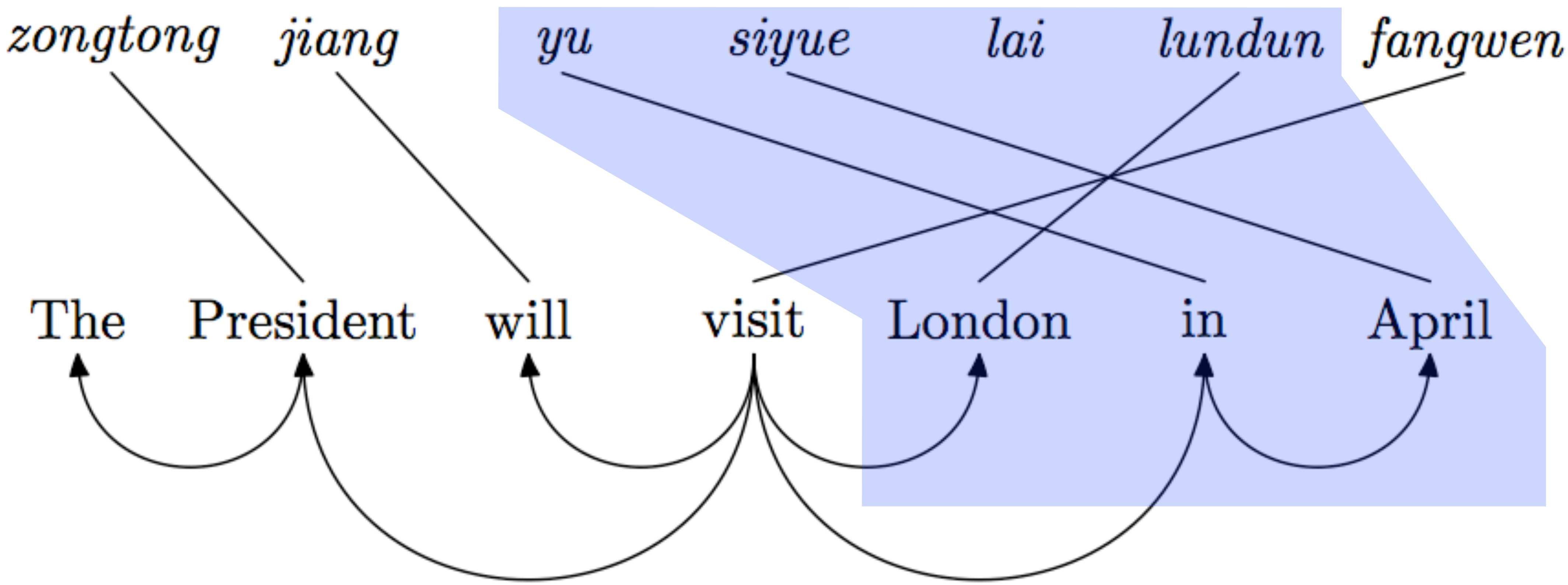
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	$\{\}$	fixed
r_2	*yu siyue*	in April	$\{1 \rightarrow 2\}$	fixed
r_3	*zongtong jiang*	The President will	$\{2 \rightarrow 1\}$	floating left

(Shen et al., 2008)

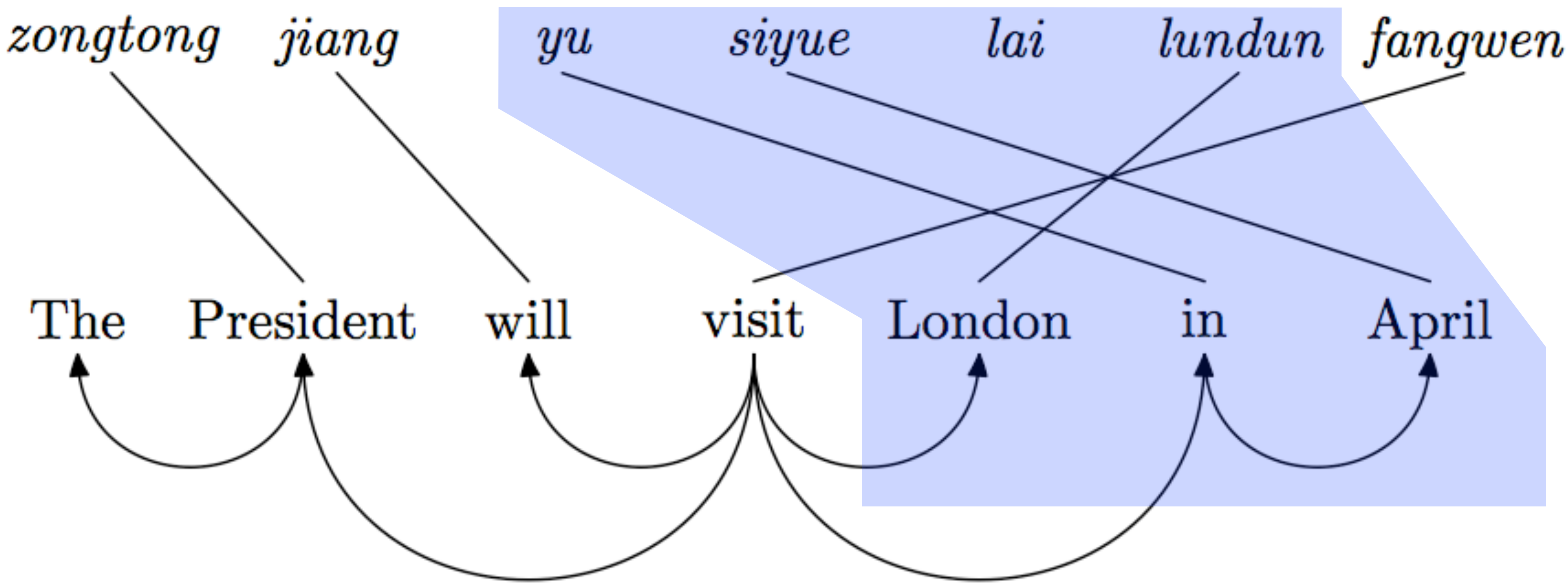
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	$\{\}$	fixed
r_2	*yu siyue*	in April	$\{1 \rightarrow 2\}$	fixed
r_3	*zongtong jiang*	The President will	$\{2 \rightarrow 1\}$	floating left

(Shen et al., 2008)

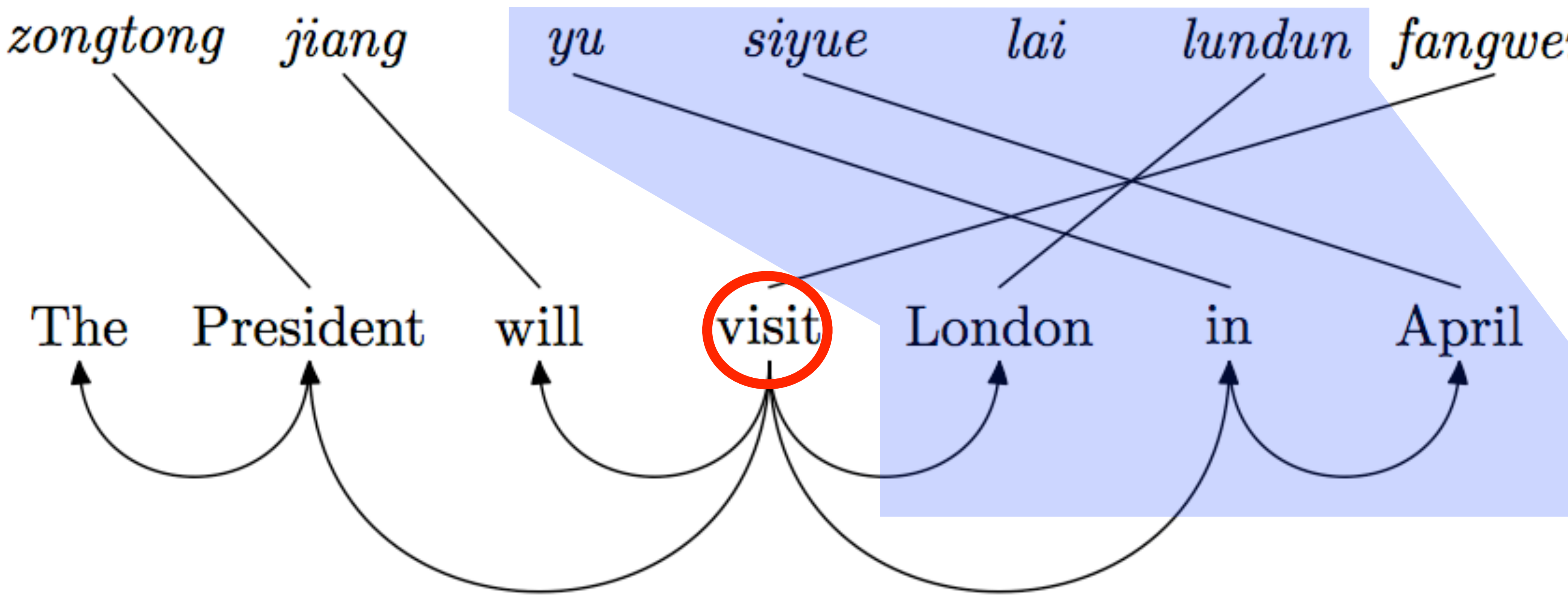
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	{}	fixed
r_2	*yu siyue*	in April	$\{1 \rightarrow 2\}$	fixed
r_3	*zongtong jiang*	The President will	$\{2 \rightarrow 1\}$	floating left
r_4	*yu siyue lai lundun*	London in April	$\{2 \rightarrow 3\}$	floating right

(Shen et al., 2008)

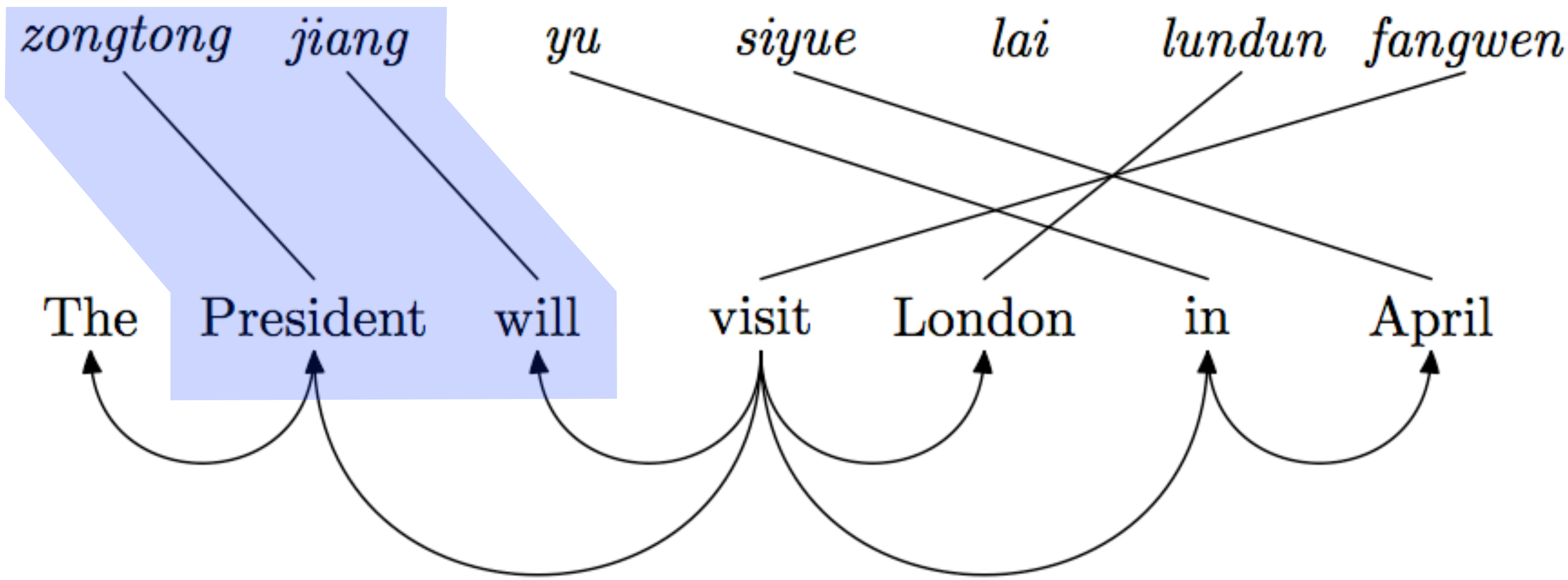
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	$\{\}$	fixed
r_2	*yu siyue*	in April	$\{1 \rightarrow 2\}$	fixed
r_3	*zongtong jiang*	The President will	$\{2 \rightarrow 1\}$	floating left
r_4	*yu siyue lai lundun*	London in April	$\{2 \rightarrow 3\}$	floating right

(Shen et al., 2008)

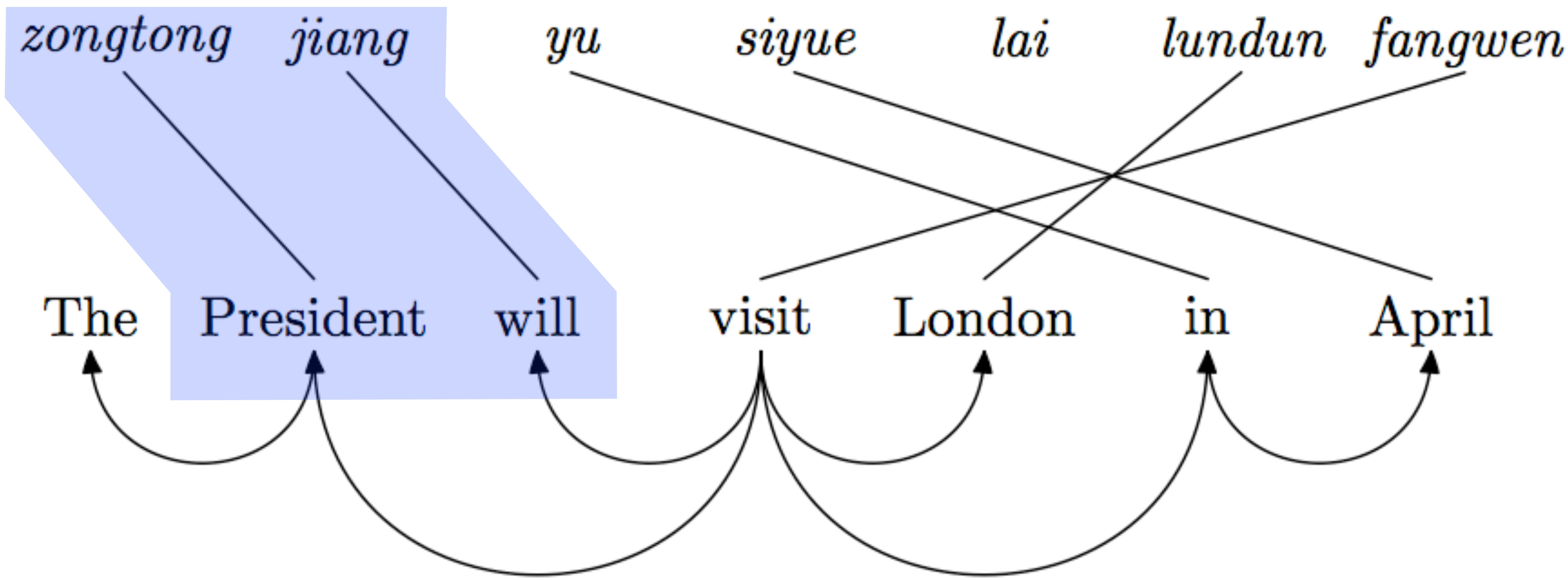
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	{}	fixed
r_2	*yu siyue*	in April	$\{1 \to 2\}$	fixed
r_3	*zongtong jiang*	The President will	$\{2 \to 1\}$	floating left
r_4	*yu siyue lai lundun*	London in April	$\{2 \to 3\}$	floating right

(Shen et al., 2008)

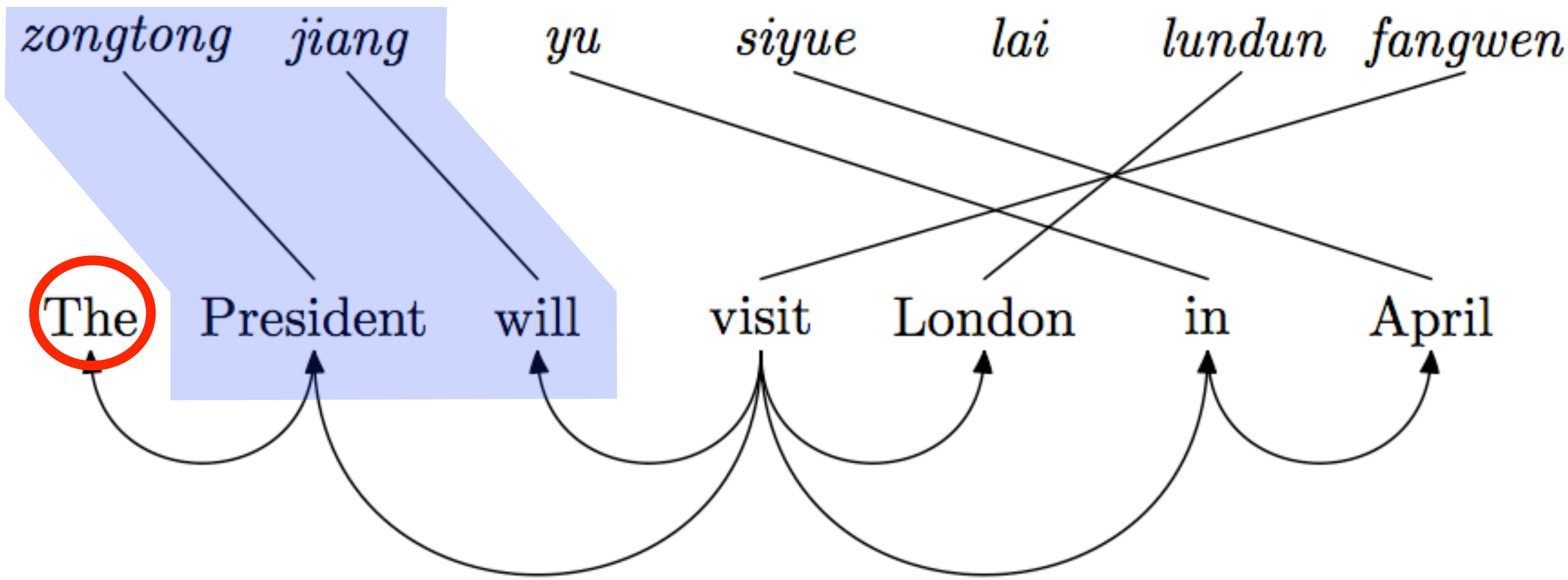
Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	$\{\}$	fixed
r_2	*yu siyue*	in April	$\{1 \rightarrow 2\}$	fixed
r_3	*zongtong jiang*	The President will	$\{2 \rightarrow 1\}$	floating left
r_4	*yu siyue lai lundun*	London in April	$\{2 \rightarrow 3\}$	floating right
r_5	*zongtong jiang*	President will	$\{\}$	ill-formed

(Shen et al., 2008)

Rule Extraction



	source phrase	target phrase	dependency	category
r_1	*fangwen*	visit	{}	fixed
r_2	*yu siyue*	in April	$\{1 \rightarrow 2\}$	fixed
r_3	*zongtong jiang*	The President will	$\{2 \rightarrow 1\}$	floating left
r_4	*yu siyue lai lundun*	London in April	$\{2 \rightarrow 3\}$	floating right
r_5	*zongtong jiang*	President will	{}	ill-formed

(Shen et al., 2008)

Shift-Reduce Parsing

- A state (i.e., parser configuration) consists of
  - a stack of items
  - coverage vector
- Each item is a well-formed structure
- Three actions (Huang et al., 2009)
  - shift: move a target dependency structure onto the stack
  - reduce left: combine the two items on the stack with the root of the the first item as the head
  - reduce right: combine the two items on the stack with the root of the second item as the head

Shift-Reduce Parsing

step	action	rule	stack	coverage
0				○ ○ ○ ○ ○ ○ ○

Shift-Reduce Parsing

step	action	rule	stack	coverage
0				○○○○○○○
1	S	r_3	[The President will]	●●○○○○○

Shift-Reduce Parsing

step	action	rule	stack	coverage
0				○○○○○○○
1	S	r_3	[The President will]	●●○○○○○
2	S	r_1	[The President will] [visit]	●●○○○○●

Shift-Reduce Parsing

step	action	rule	stack	coverage
0				○ ○ ○ ○ ○ ○ ○
1	S	r_3	[The President will]	● ● ○ ○ ○ ○ ○
2	S	r_1	[The President will] [visit]	● ● ○ ○ ○ ○ ●
3	R_l		[The President will visit]	● ● ○ ○ ○ ○ ●

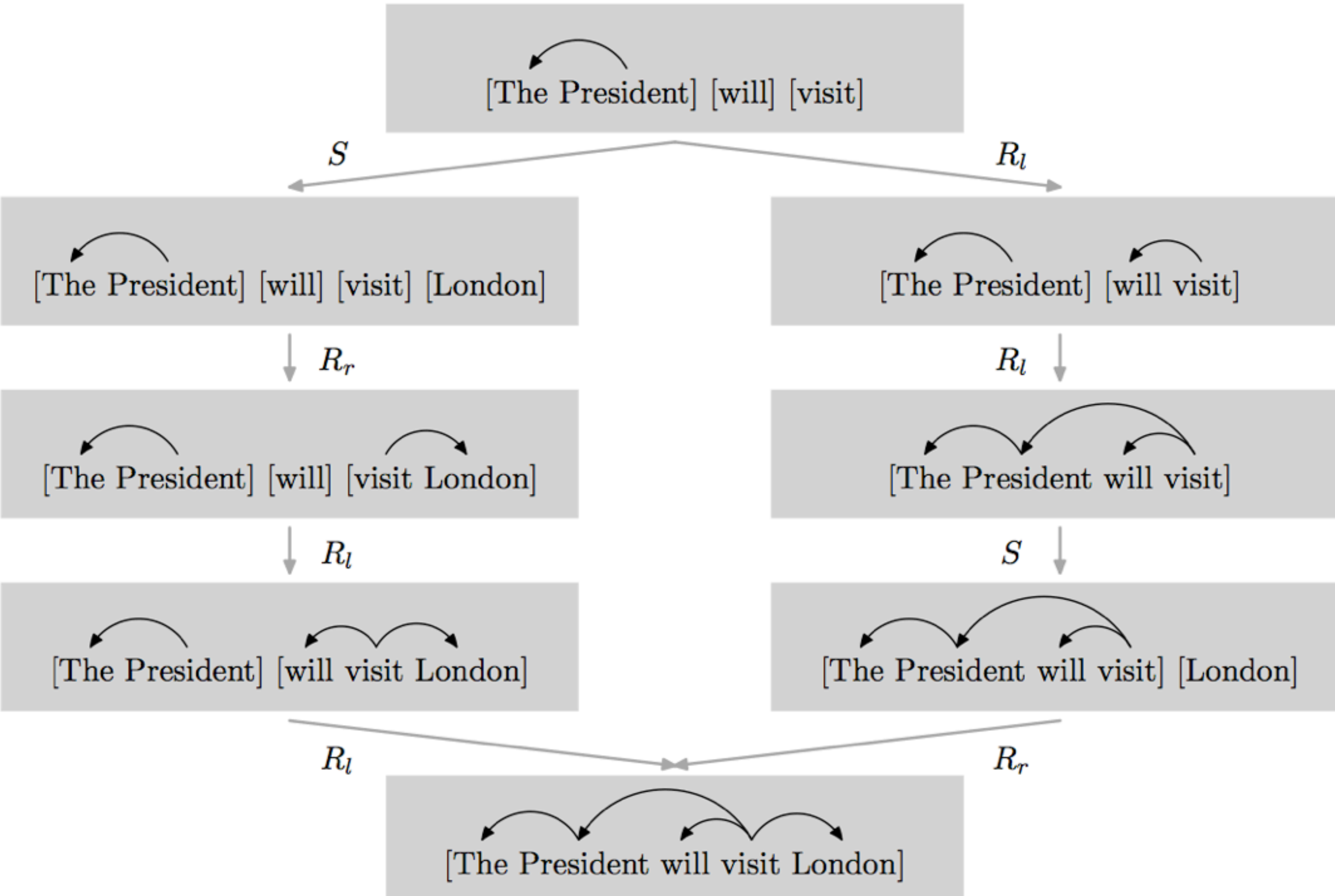
Shift-Reduce Parsing

step	action	rule	stack	coverage
0				○○○○○○○
1	S	r_3	[The President will]	●●○○○○○
2	S	r_1	[The President will] [visit]	●●○○○○●
3	R_l		[The President will visit]	●●○○○○●
4	S	r_4	[The President will visit] [London in April]	●●●●●●●

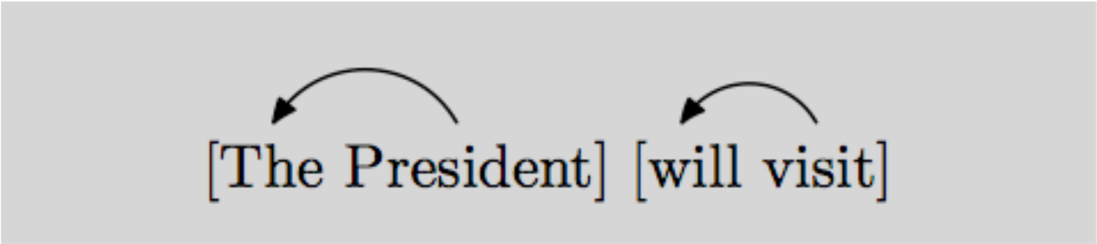
Shift-Reduce Parsing

step	action	rule	stack	coverage
0				○○○○○○○
1	S	r_3	[The President will]	●●○○○○○
2	S	r_1	[The President will] [visit]	●●○○○○●
3	R_l		[The President will visit]	●●○○○○●
4	S	r_4	[The President will visit] [London in April]	●●●●●●●
5	R_r		[The President will visit London in April]	●●●●●●●

Ambiguity

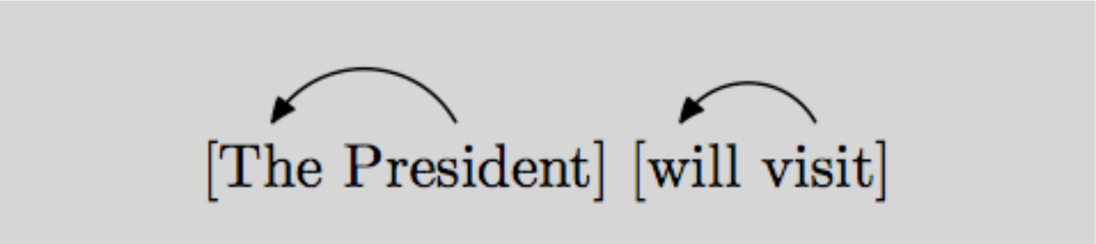

[The President] [will] [visit]
S
R_l
[The President] [will] [visit] [London]
[The President] [will visit]
R_r
R_l
[The President] [will] [visit London]
[The President will visit]
R_l
S
[The President] [will visit London]
[The President will visit] [London]
R_l
R_r
[The President will visit London]

Shift-Reduce Conflicts



- The shift-reduce parser faces four types of conflicts
  - shift vs. shift
  - shift vs. reduce left
  - shift vs. reduce right
  - reduce left vs. reduce right

Shift-Reduce Conflicts



all the three actions are applicable!

- The shift-reduce parser faces four types of conflicts
  - shift vs. shift
  - shift vs. reduce left
  - shift vs. reduce right
  - reduce left vs. reduce right

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will]

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will]

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will]

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will]

S

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will]

S

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will]

S

[The President will] [visit]

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

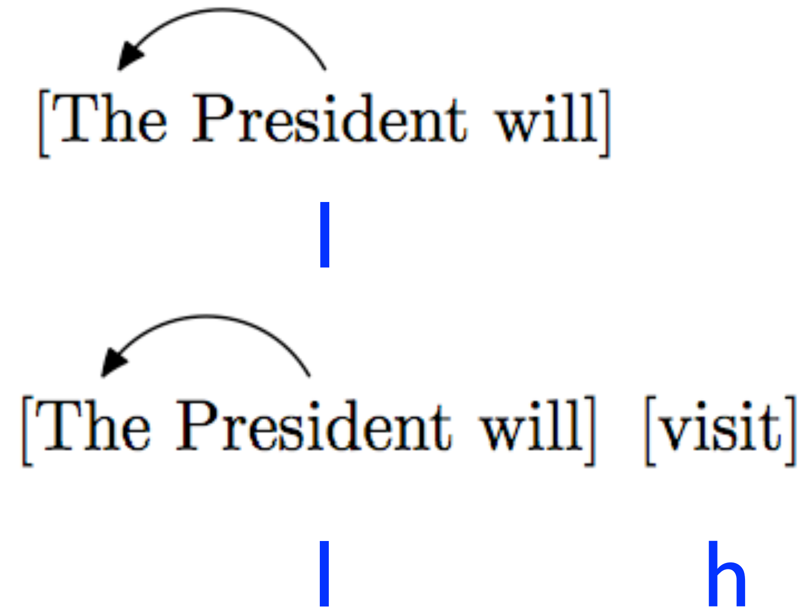
[The President will]

S

[The President will] [visit]

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



S

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will]

l

[The President will] [visit]

l h

S

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will] S

l

[The President will] [visit] R_l

l h

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will] S

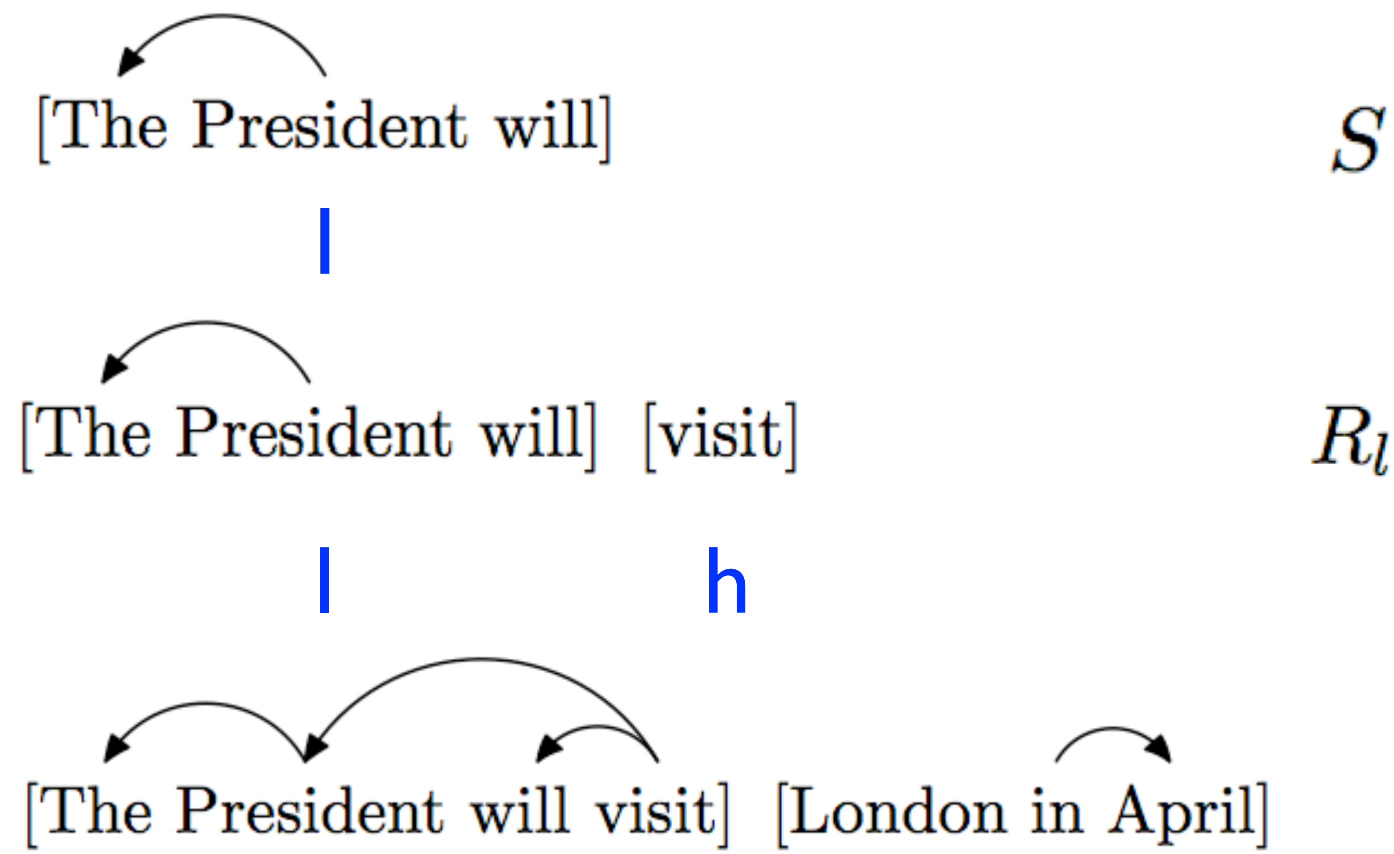
l

[The President will] [visit] R_l

l h

Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will] S

l

[The President will] [visit] R_l

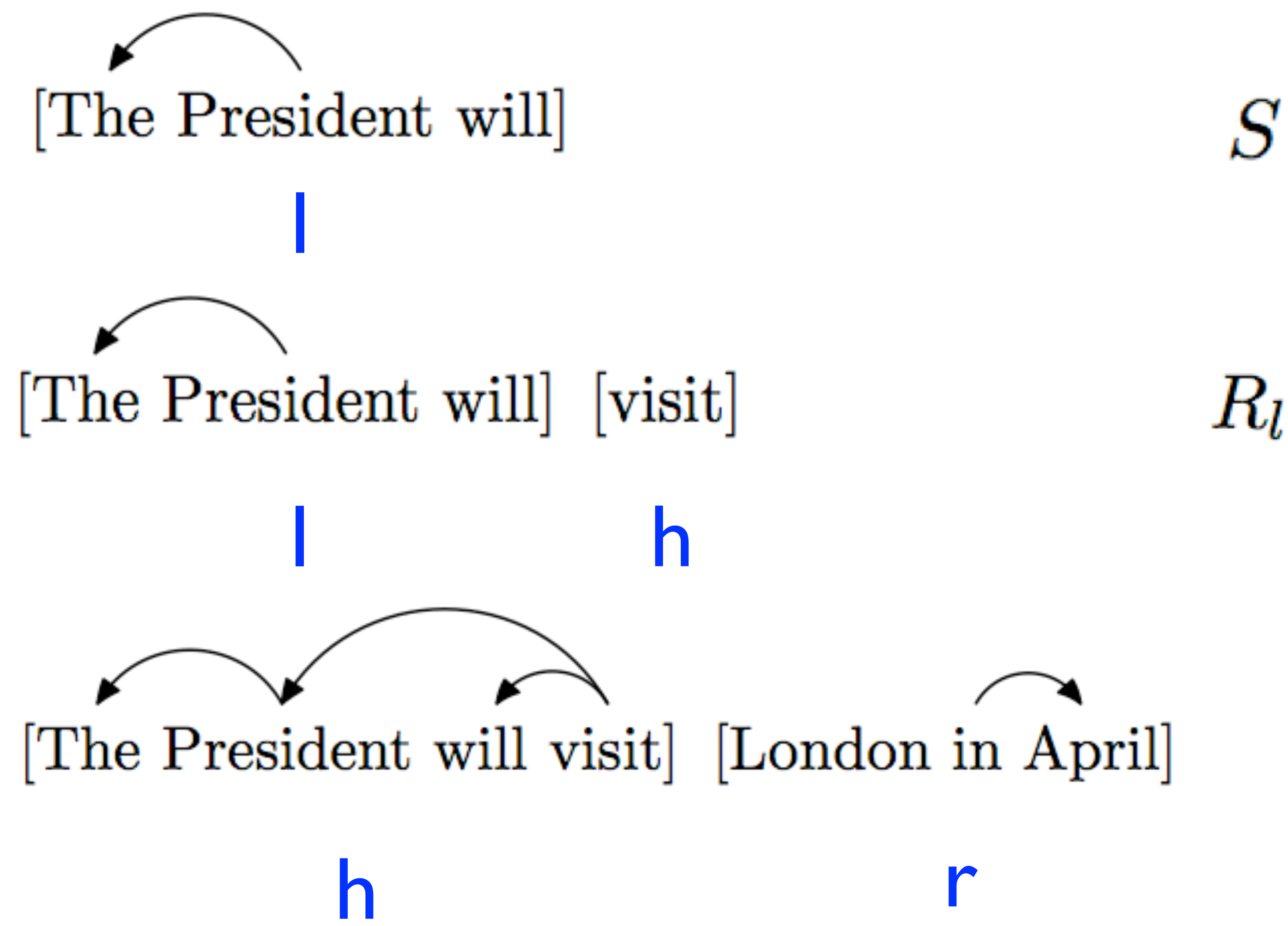
l h

[The President will visit] [London in April]

h

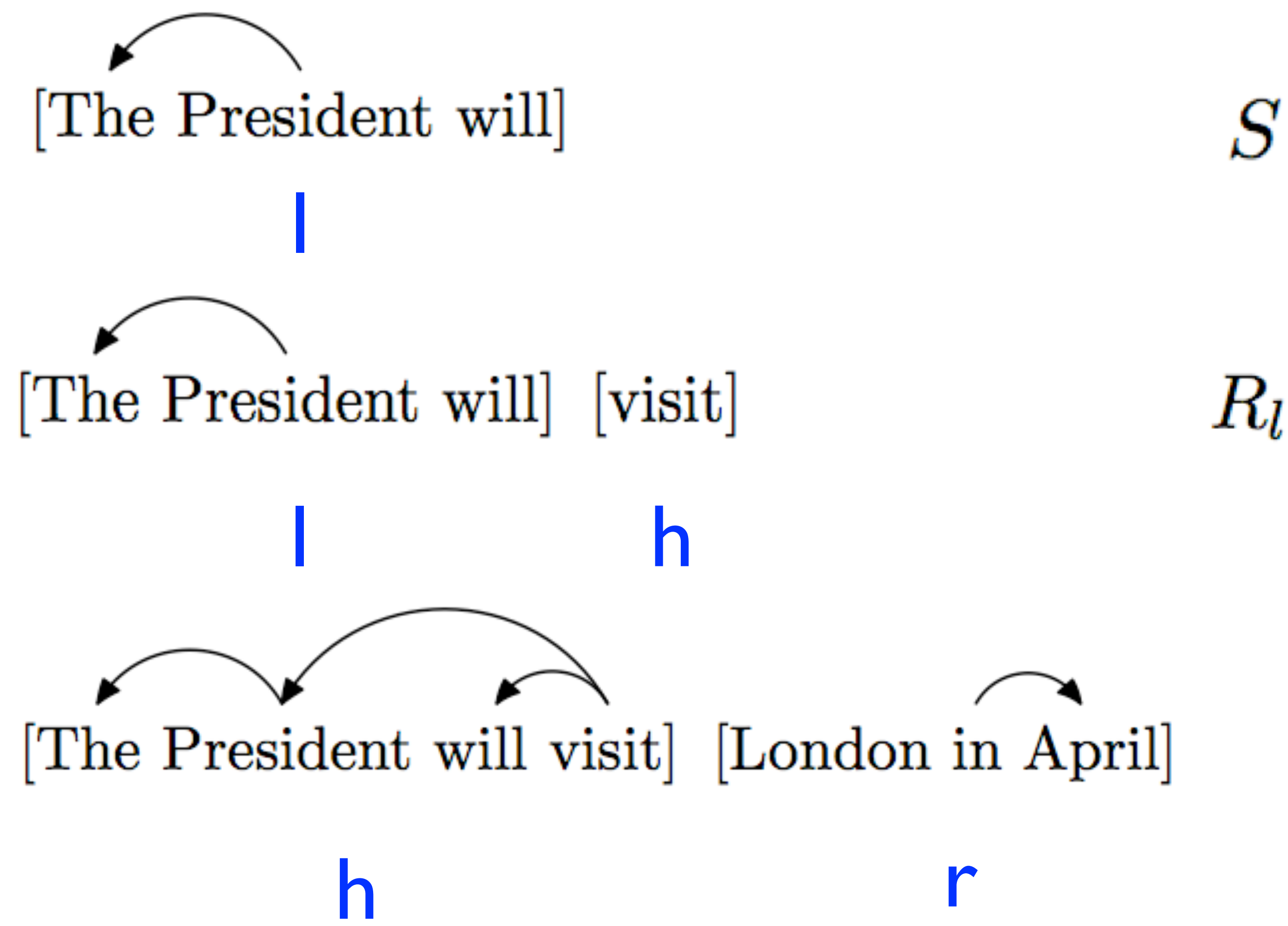
Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



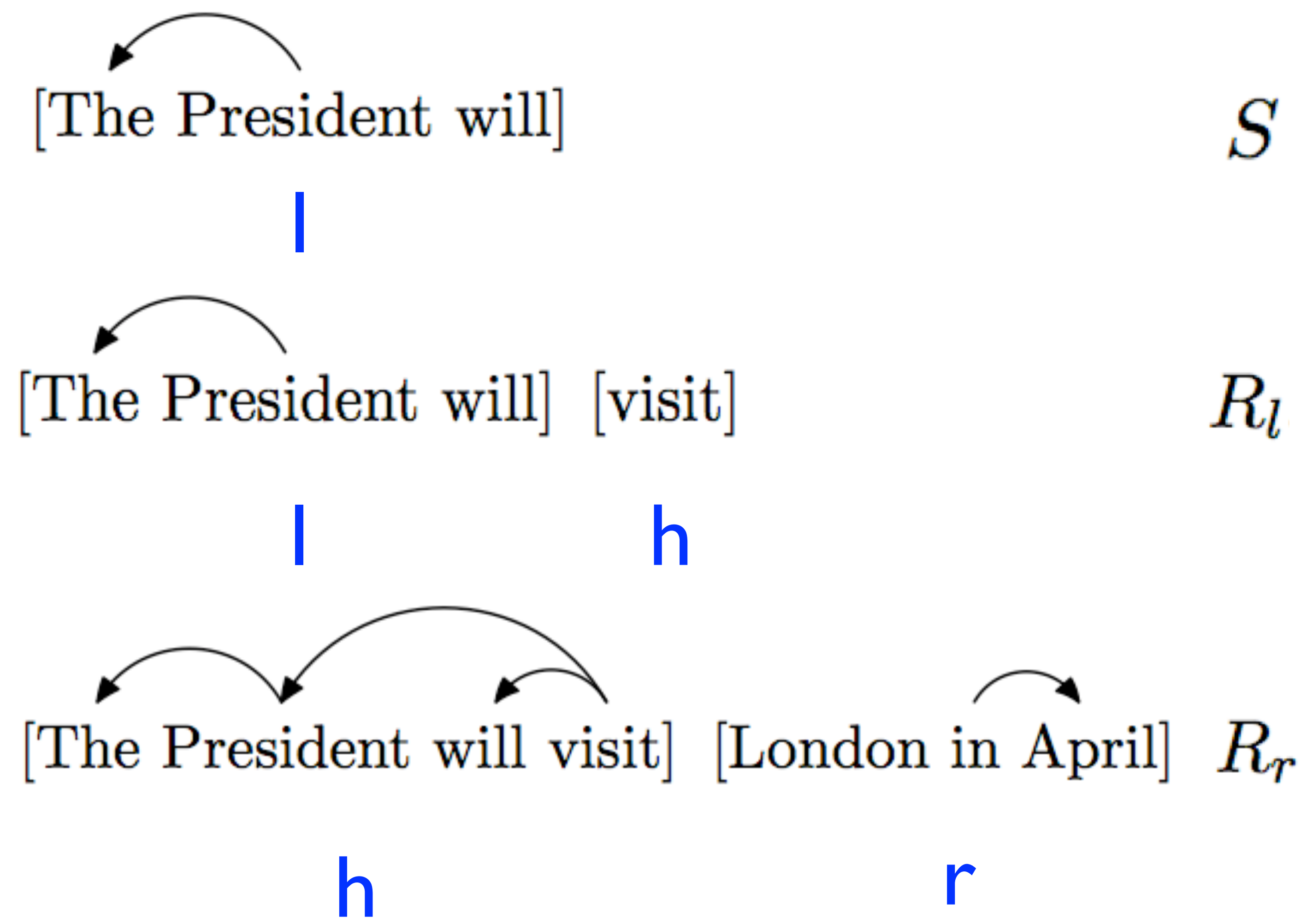
Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



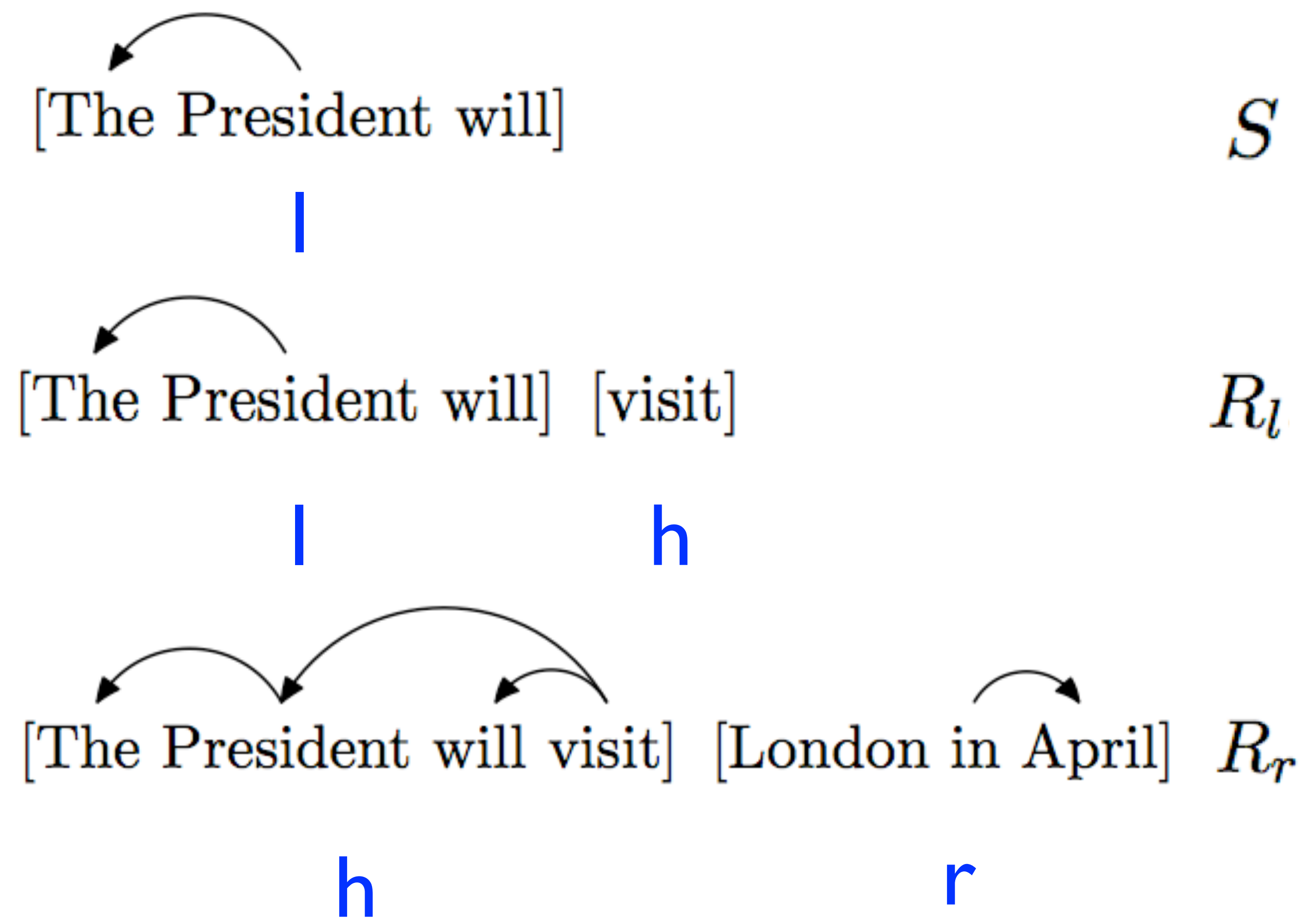
Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



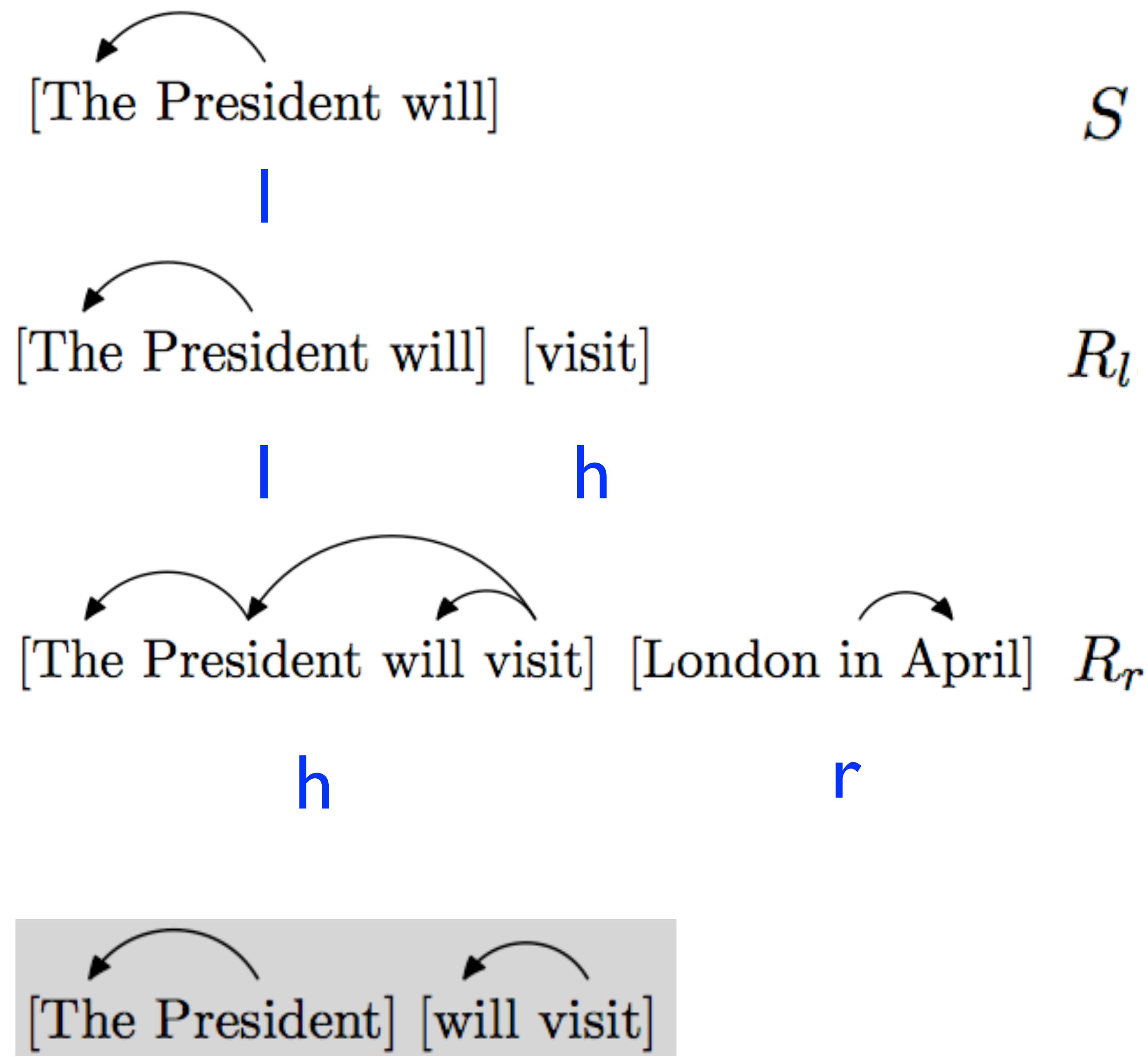
Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



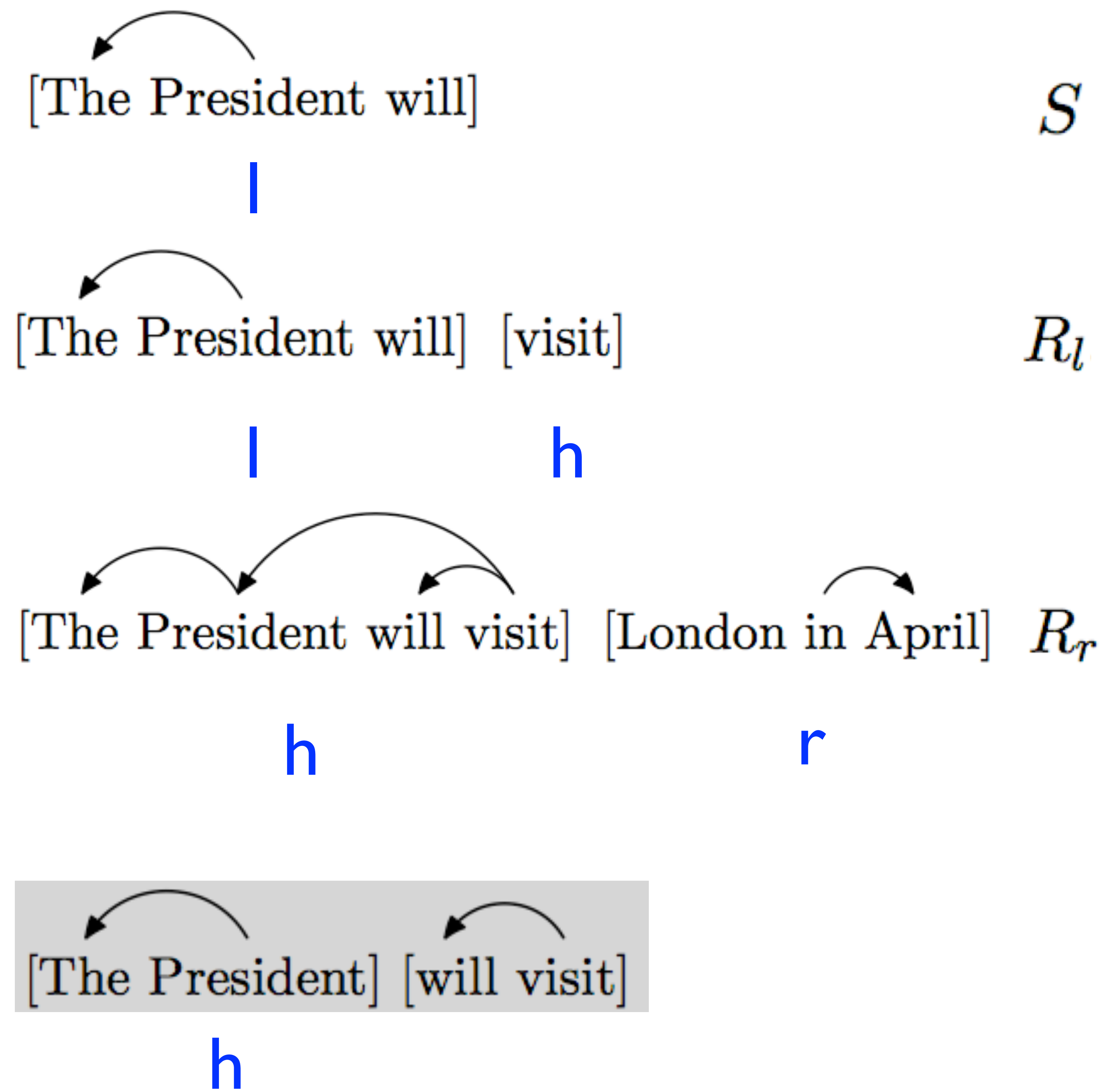
Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



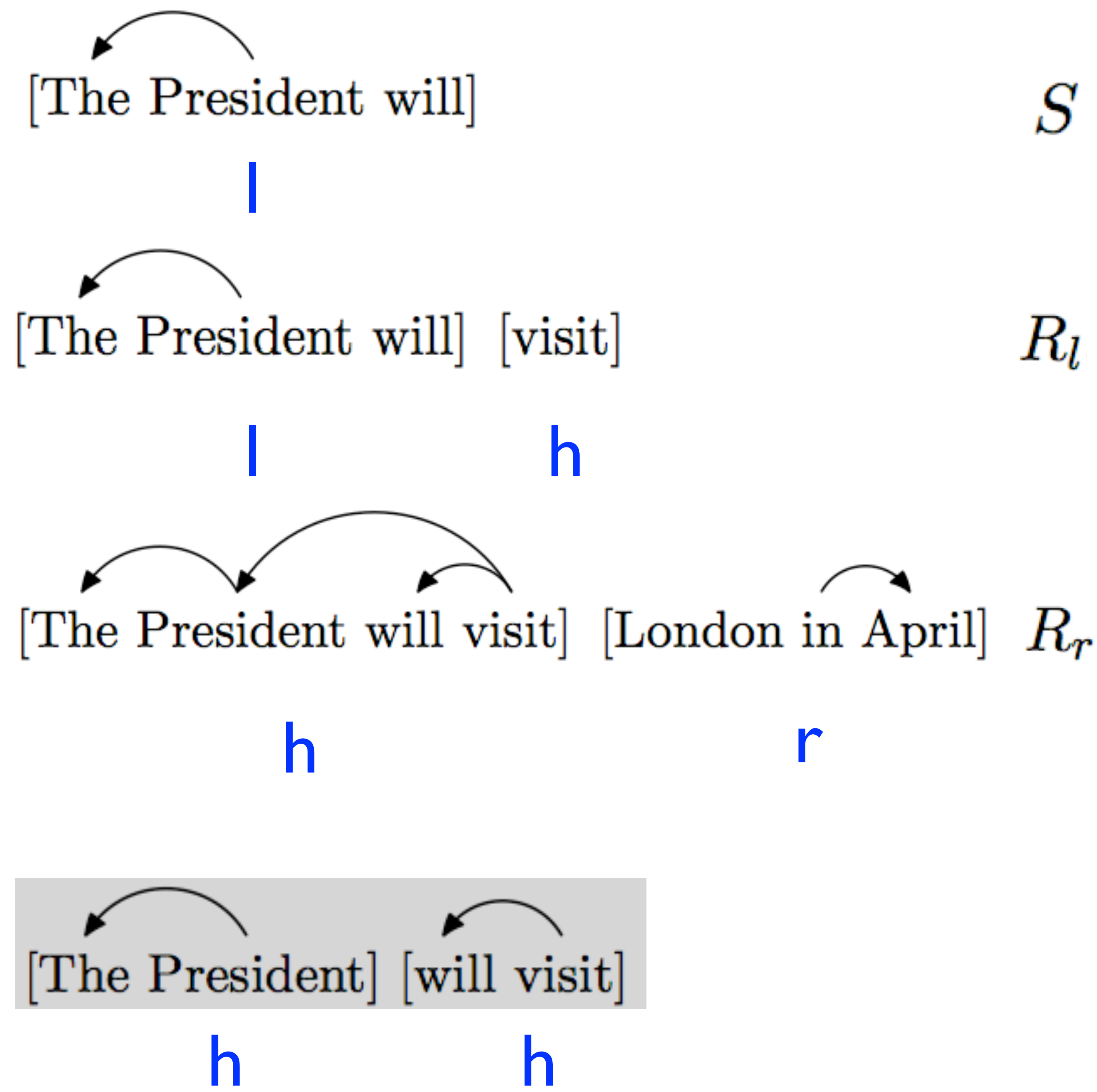
Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	

[The President will] S

l

[The President will] [visit] R_l

l h

[The President will visit] [London in April] R_r

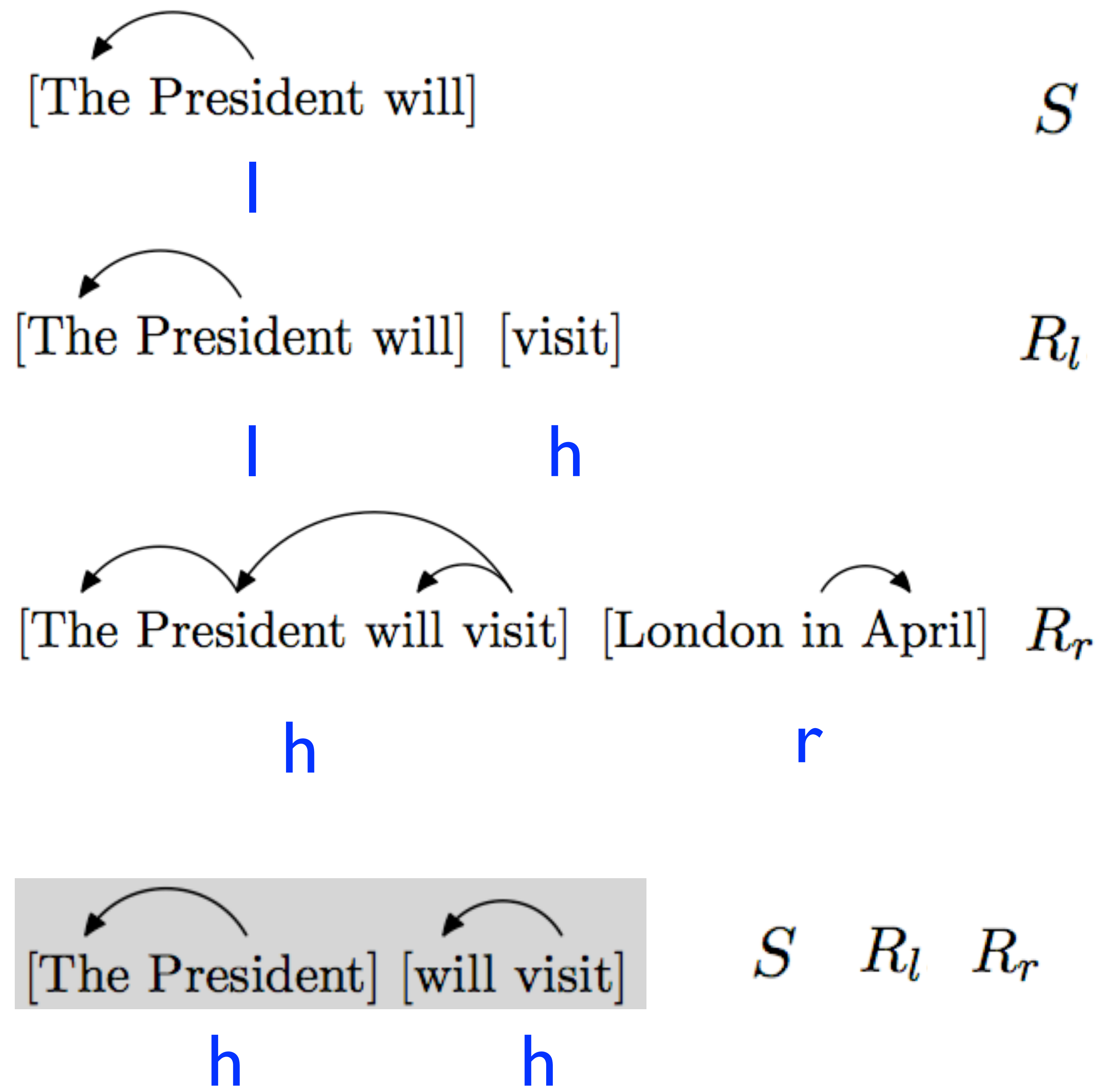
h r

[The President] [will visit]

h h

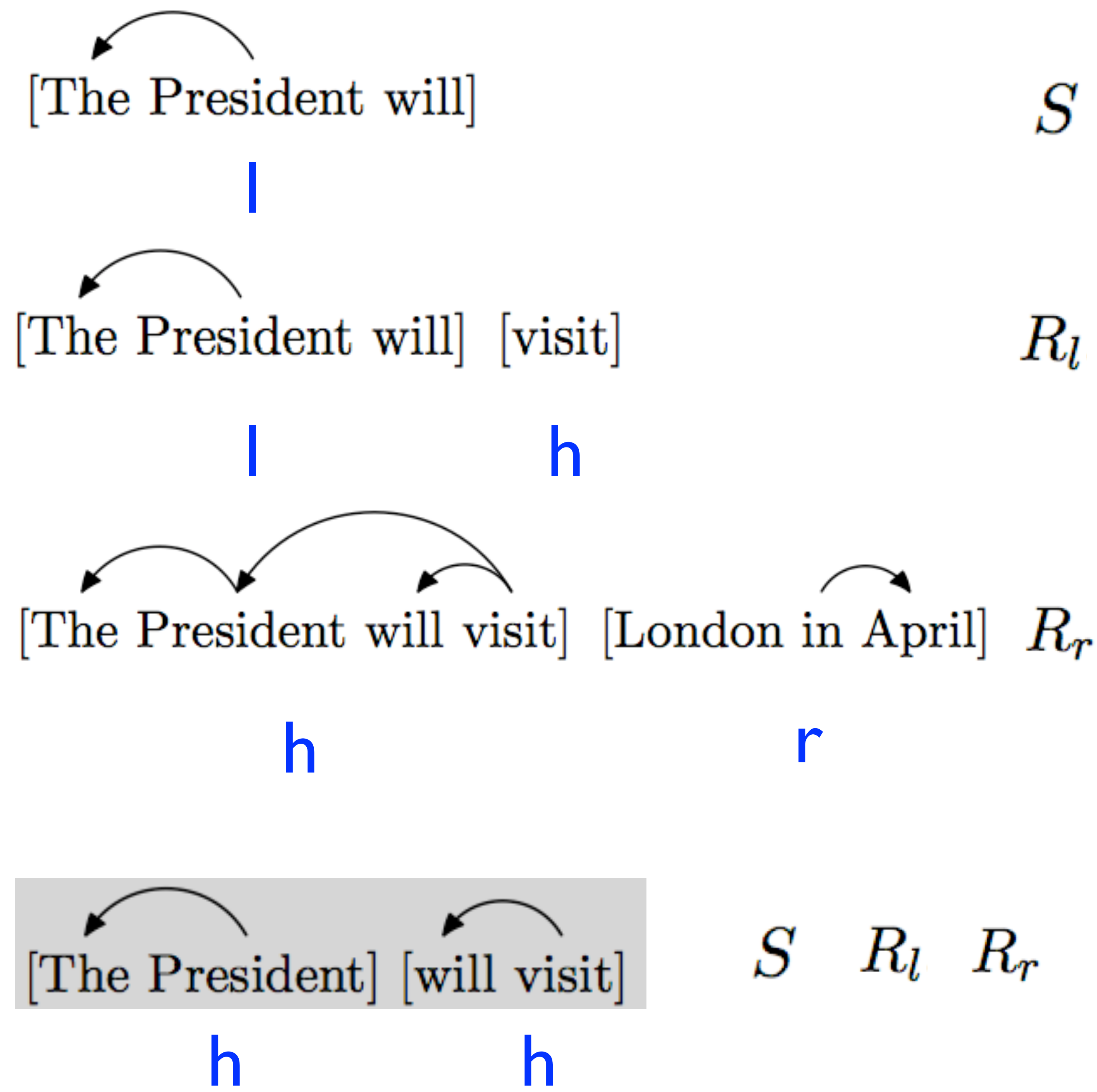
Resolving Shift-Reduce Conflicts

s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



Resolving Shift-Reduce Conflicts

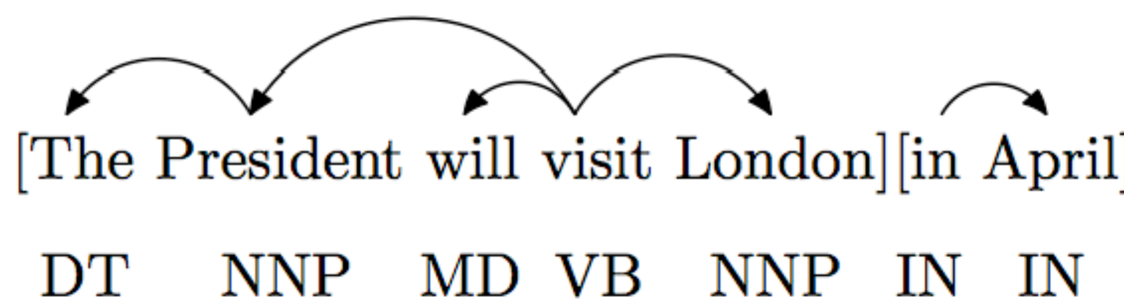
s_{t-1}	s_t	legal	action(s)
		yes	S
	h	yes	S
	l	yes	S
	r	no	
h	h	yes	S, R_l, R_r
h	l	yes	S
h	r	yes	R_r
l	h	yes	R_l
l	l	yes	S
l	r	no	
r	h	no	
r	l	no	
r	r	no	



only “h+h” is ambiguous!

The MaxEnt Classifier

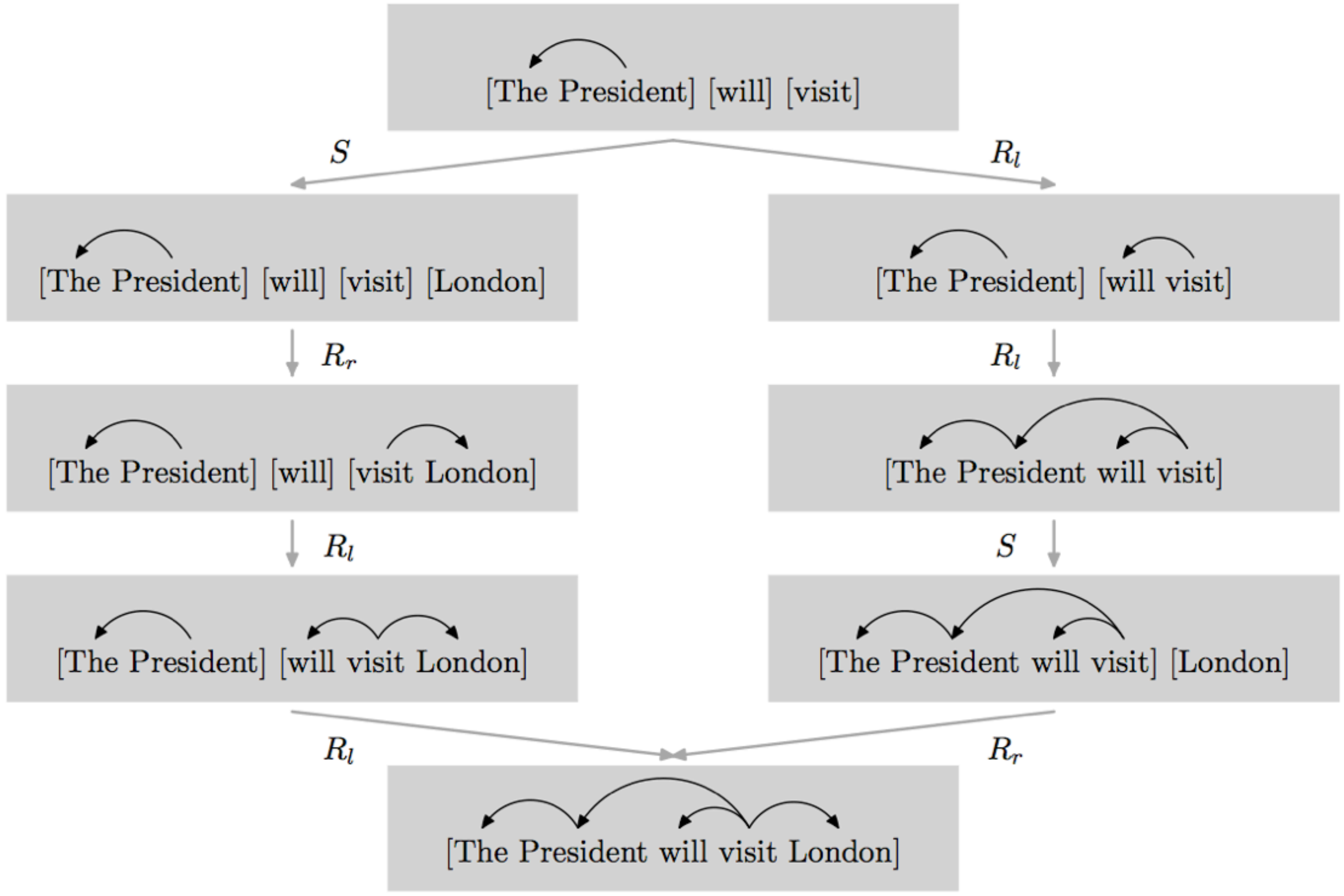
- We use a maximum entropy classifier to resolve conflicts for the “h+h” case



type	feature templates		
Unigram	c	$W_h(s_t)$	$W_h(s_{t-1})$
	$W_{lc}(s_t)$	$W_{rc}(s_{t-1})$	$T_h(s_t)$
	$T_h(s_{t-1})$	$T_{lc}(s_t)$	$T_{rc}(s_{t-1})$
Bigram	$W_h(s_t) \circ W_h(s_{t-1})$	$T_h(S_t) \circ T_h(s_{t-1})$	$W_h(s_t) \circ T_h(s_t)$
	$W_h(s_{t-1}) \circ T_h(s_{t-1})$	$W_h(s_t) \circ W_{rc}(s_{t-1})$	$W_h(s_{t-1}) \circ W_{lc}(s_t)$
Trigram	$c \circ W_h(s_t) \circ W(s_{t-1})$	$c \circ T_h(s_t) \circ T_h(s_{t-1})$	$W_h(s_t) \circ W_h(s_{t-1}) \circ T_{lc}(s_t)$
	$W_h(s_t) \circ W_h(s_{t-1}) \circ T_{rc}(s_{t-1})$	$T_h(s_t) \circ T_h(s_{t-1}) \circ T_{lc}(s_t)$	$T_h(s_t) \circ T_h(s_{t-1}) \circ T_{rc}(s_{t-1})$

Training Examples for the MaxEnt Classifier

- We build a derivation graph to compactly represent all derivations



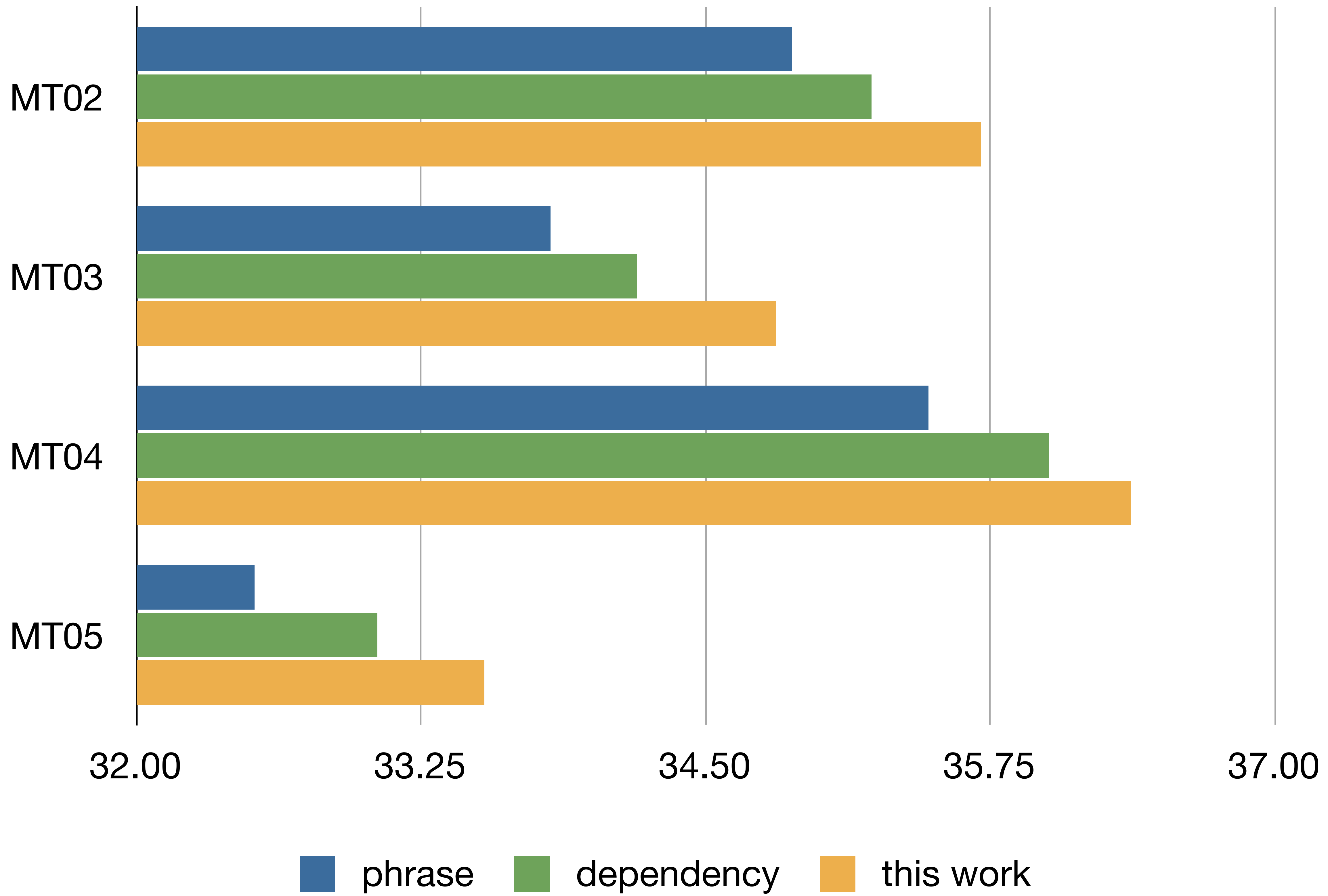
Beam Search Shift-Reduce Parsing

- We divide features into two categories
  - *standard:* rule probs, n-gram LM, reordering, etc.
  - *dependency*: depLM, ill-formed penalty, MaxEnt
- Beam search shift-reduce parsing (Zhang and Clark, 2008)
- Hypergraph reranking (Huang, 2008)
  - 1st pass: produce a hypergraph with std features
  - 2nd pass: hypergraph reranking with dep features

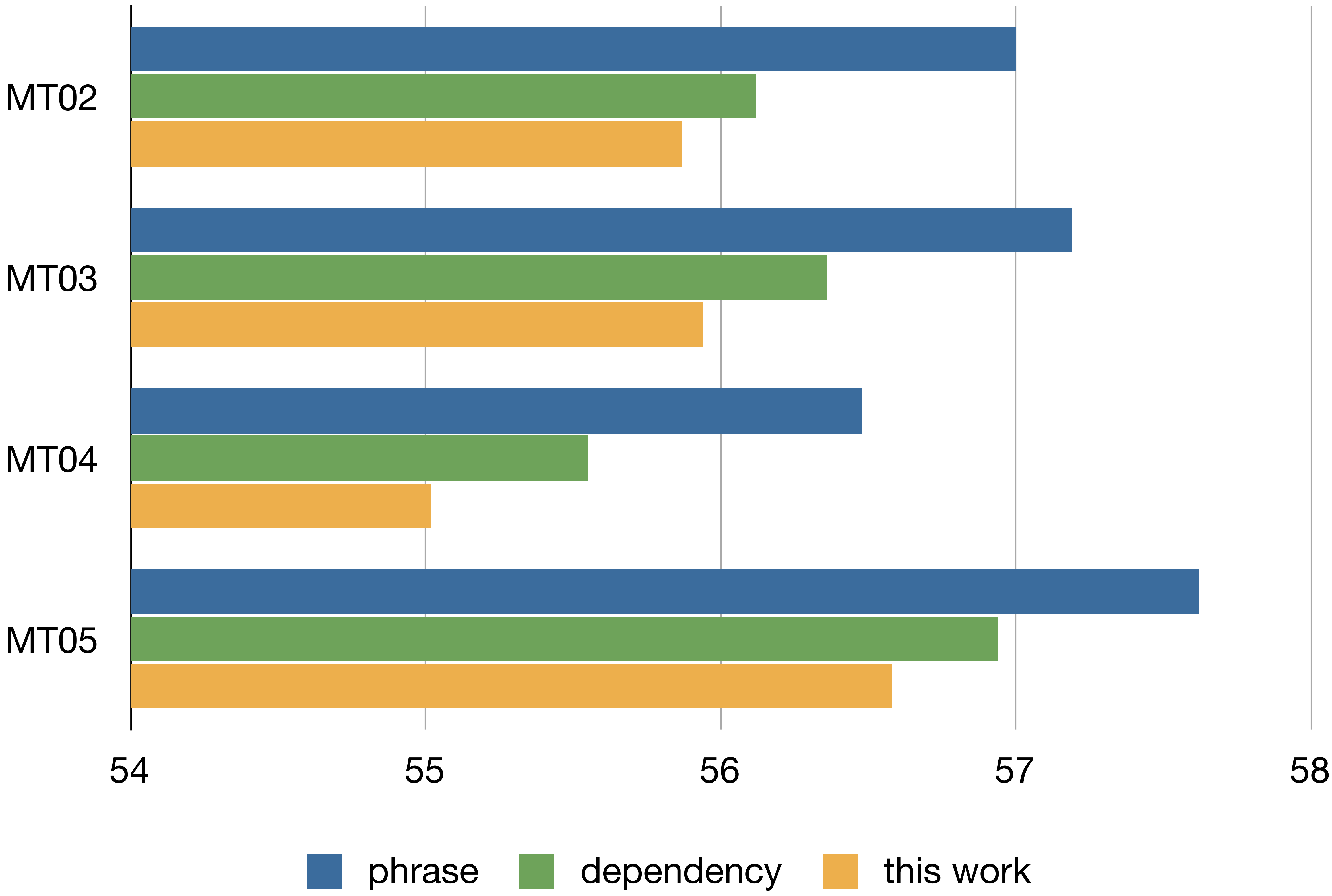
Experiments

- Baseline systems
  - Moses (Koehn et al., 2007)
  - string-to-dependency (Shen et al., 2008)
- Datasets
  - training: 2.9M Chinese-English sentence pairs
  - development: NIST 2002
  - test: NIST 2003, 2004, 2005
- Evaluation Metric: uncased BLEU and TER

BLEU


MT02
MT03
MT04
MT05
32.00
33.25
34.50
35.75
37.00
phrase
dependency
this work

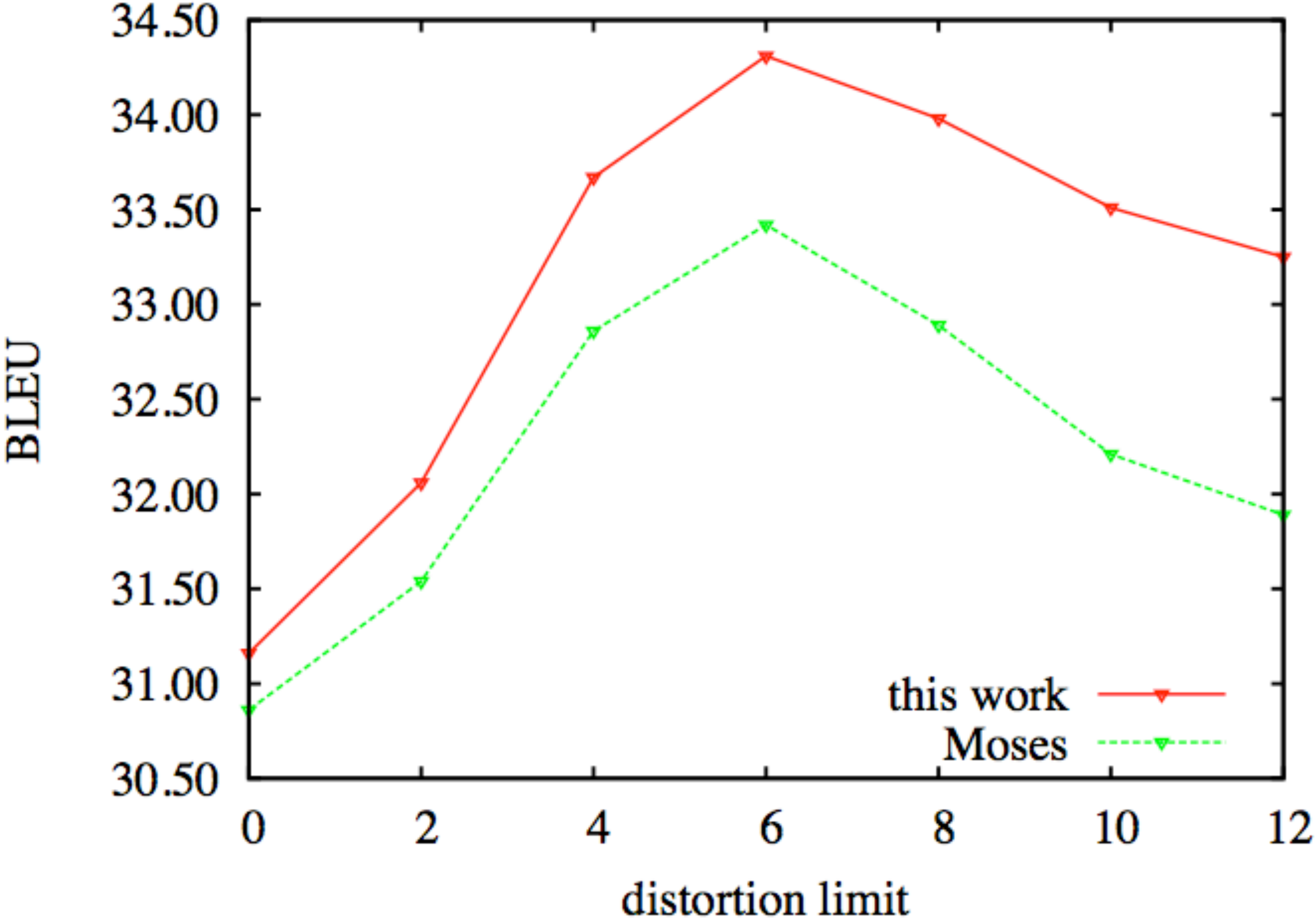
TER


MT02
MT03
MT04
MT05
54
55
56
57
58
phrase
dependency
this work

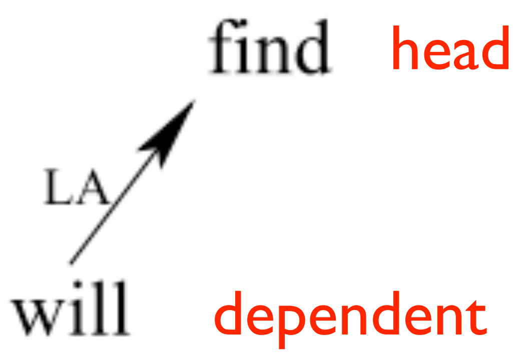
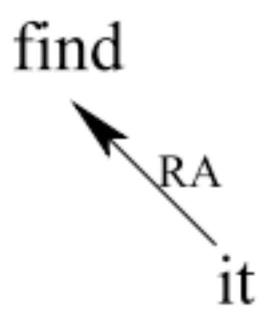
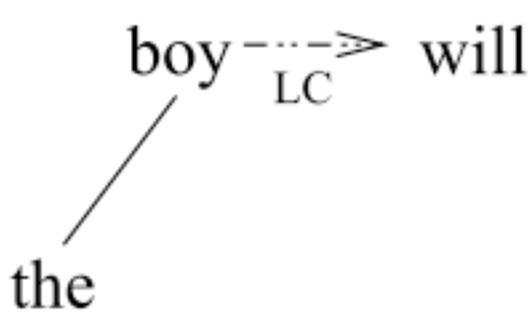
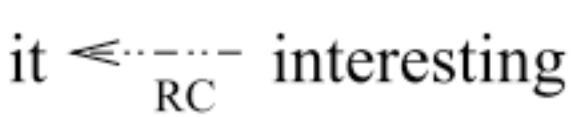
depLM vs. MaxEnt

features	BLEU	TER
standard	34.79	56.93
+depLM	35.29	56.17
+MaxEnt	35.40	56.09
+depLM & MaxEnt	35.71	55.87

BLEU with Varying Dist. Limits



Expressiveness and Complexity

	Shen et al. (2008)	this work
	left adjoining	reduce left
	right adjoining	reduce right
	left concatenation	N/A
	right concatenation	N/A

Conclusion

- We have presented a shift-reduce decoding algorithm for string-to-dependency translation
  - only uses phrases
  - resolves conflicts using a maxent classifier
- Future work
  - introducing more actions
  - comparison with using undirected floating structures

Thanks

We are grateful to Collin Cherry for offering constructive comments and suggestions.