

Strongly Incremental Constituency Parsing with Graph Neural Networks

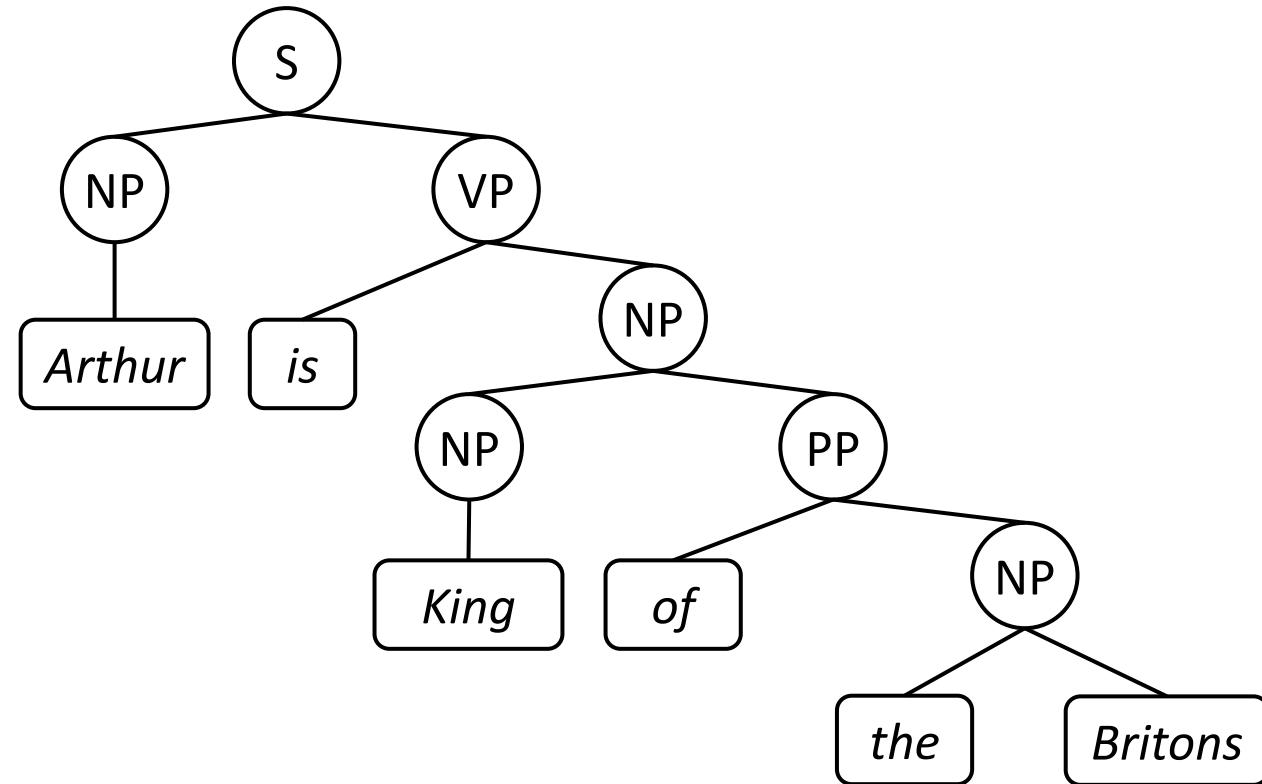
Kaiyu Yang and Jia Deng



PRINCETON
UNIVERSITY

Constituency Parsing

Arthur is King of the Britons



Shift-Reduce Parsers

stack



Arthur is King of the Britons



Shift-Reduce Parsers

stack



Arthur is King of the Britons



- SHIFT: Move the next word into the stack

shift

Shift-Reduce Parsers

stack



is King of the Britons



- SHIFT: Move the next word into the stack

shift, unary_reduce(NP)

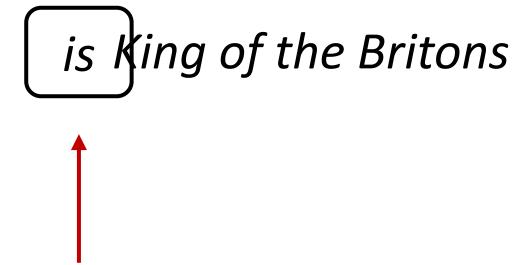
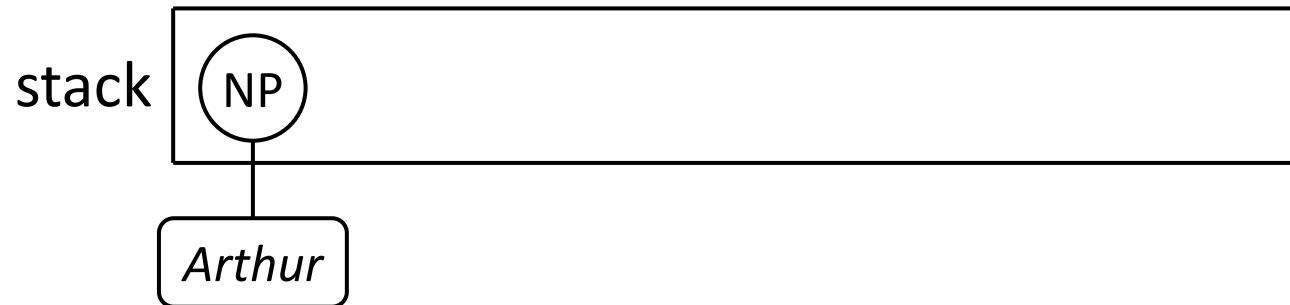
Shift-Reduce Parsers



- SHIFT: Move the next word into the stack

shift, unary_reduce(NP)

Shift-Reduce Parsers



- SHIFT: Move the next word into the stack

shift, unary_reduce(NP), shift

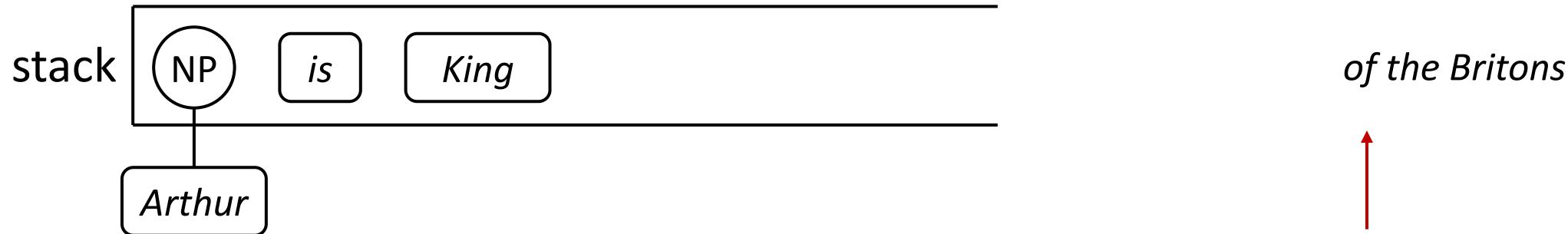
Shift-Reduce Parsers



- SHIFT: Move the next word into the stack

shift, unary_reduce(NP), shift, **shift**

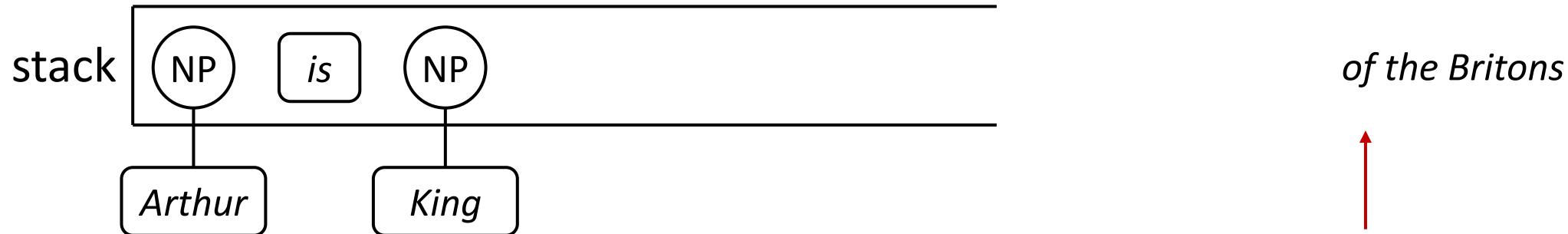
Shift-Reduce Parsers



- SHIFT: Move the next word into the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP)

Shift-Reduce Parsers



- SHIFT: Move the next word into the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP)

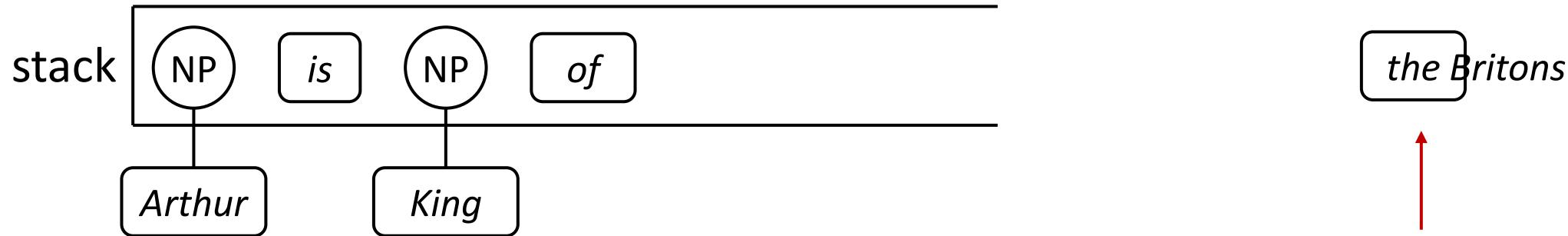
Shift-Reduce Parsers



- SHIFT: Move the next word into the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), **shift**

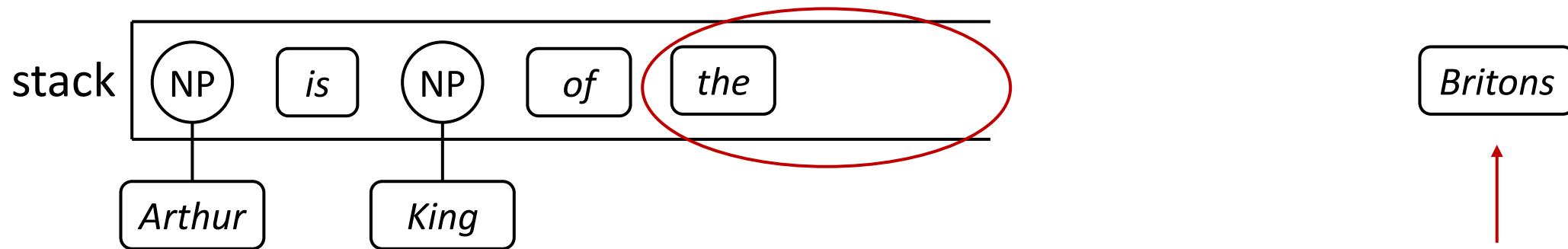
Shift-Reduce Parsers



- SHIFT: Move the next word into the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, **shift**

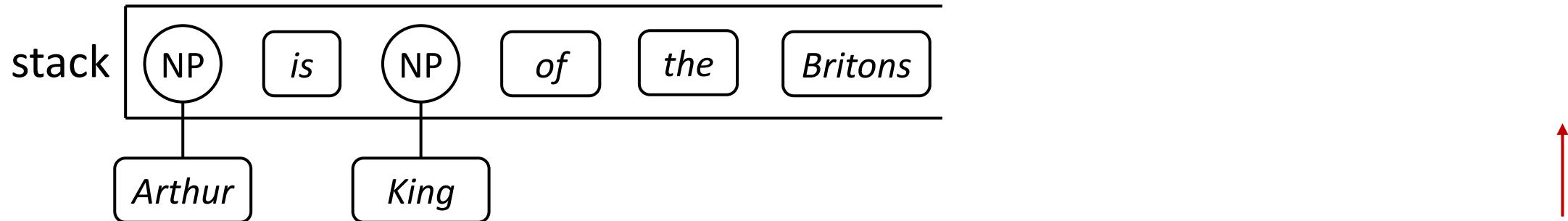
Shift-Reduce Parsers



- SHIFT: Move the next word into the stack
- REDUCE: Combine the top two elements in the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, **shift**

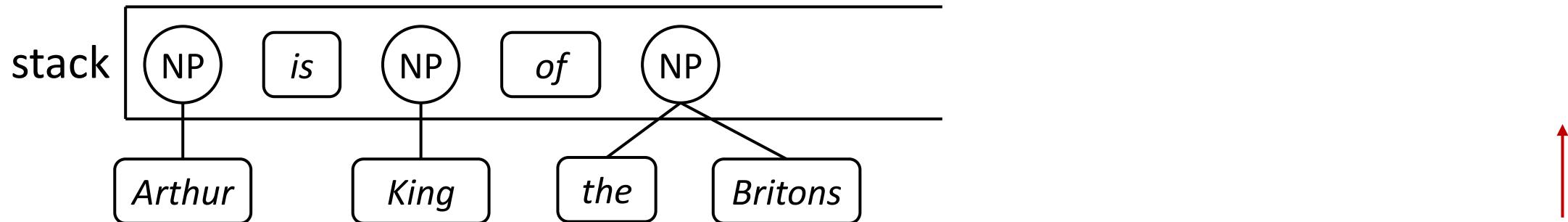
Shift-Reduce Parsers



- SHIFT: Move the next word into the stack
- REDUCE: Combine the top two elements in the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, **binary_reduce(NP)**

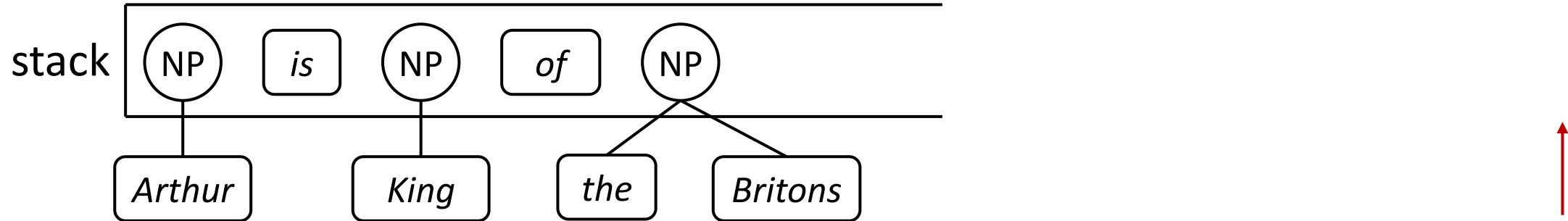
Shift-Reduce Parsers



- SHIFT: Move the next word into the stack
- REDUCE: Combine the top two elements in the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, **binary_reduce(NP)**

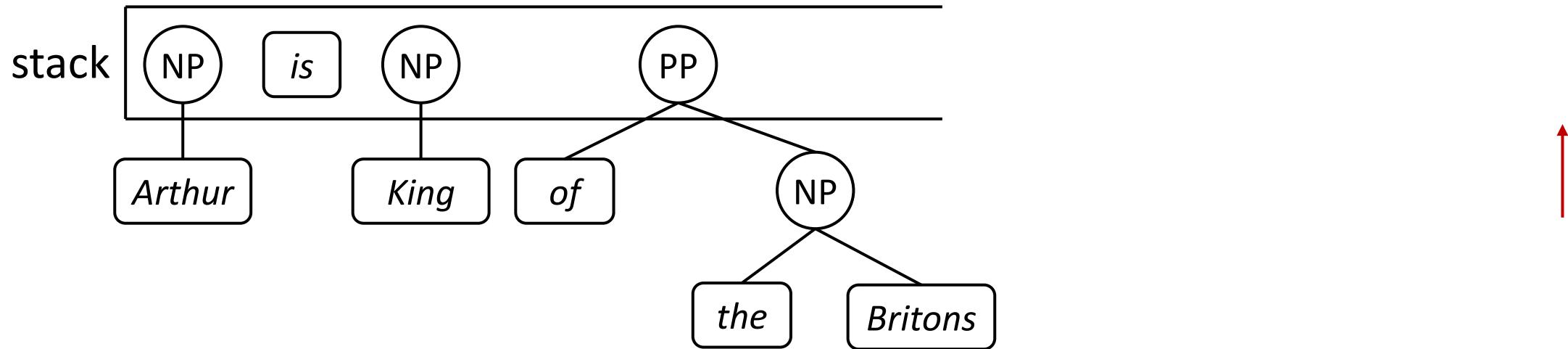
Shift-Reduce Parsers



- SHIFT: Move the next word into the stack
 - REDUCE: Combine the top two elements in the stack

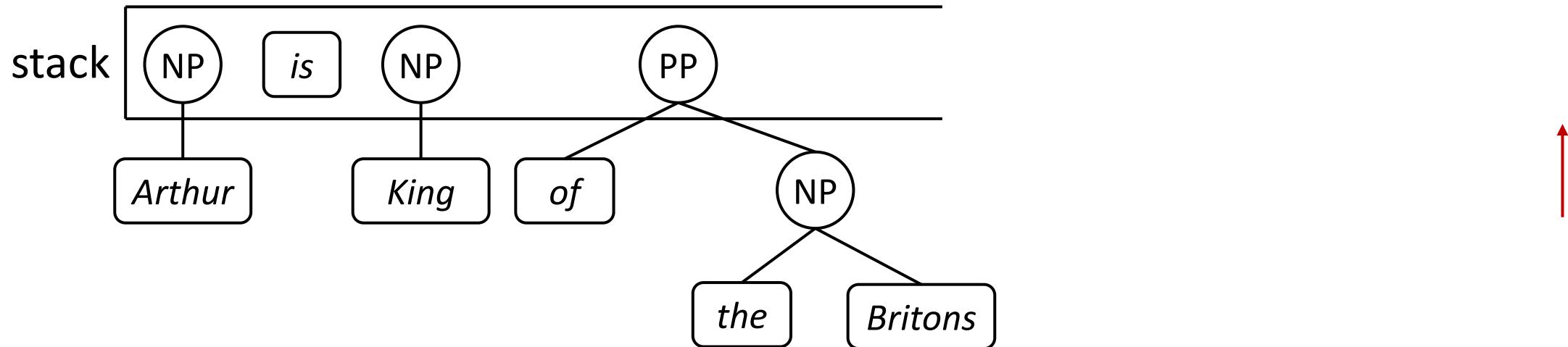
shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary reduce(PP)

Shift-Reduce Parsers



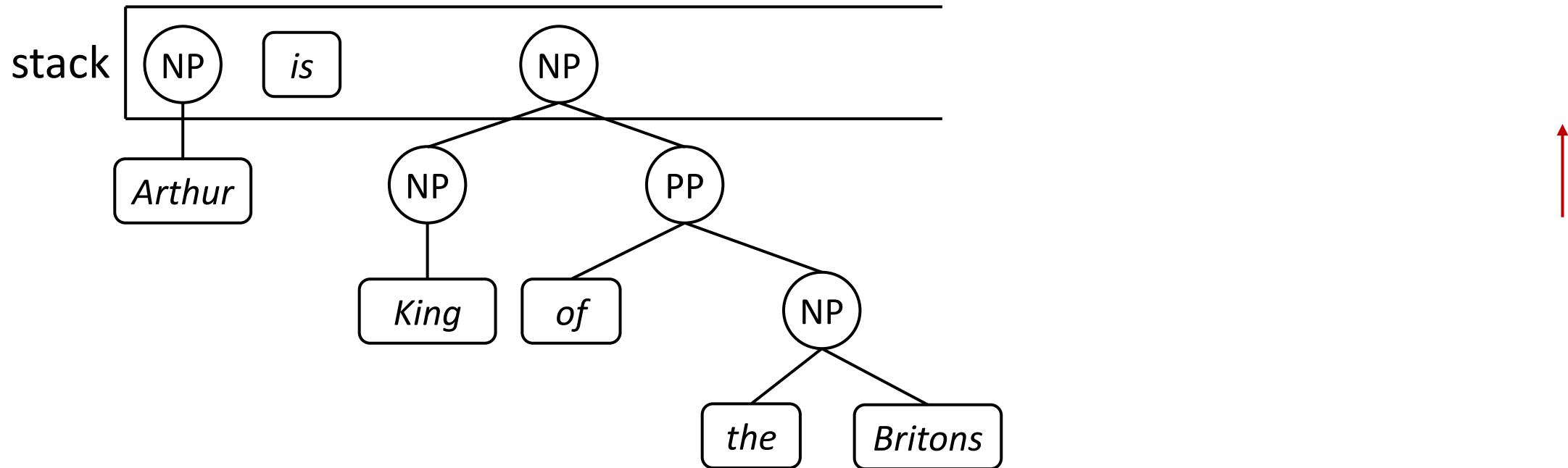
shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary_reduce(PP)

Shift-Reduce Parsers



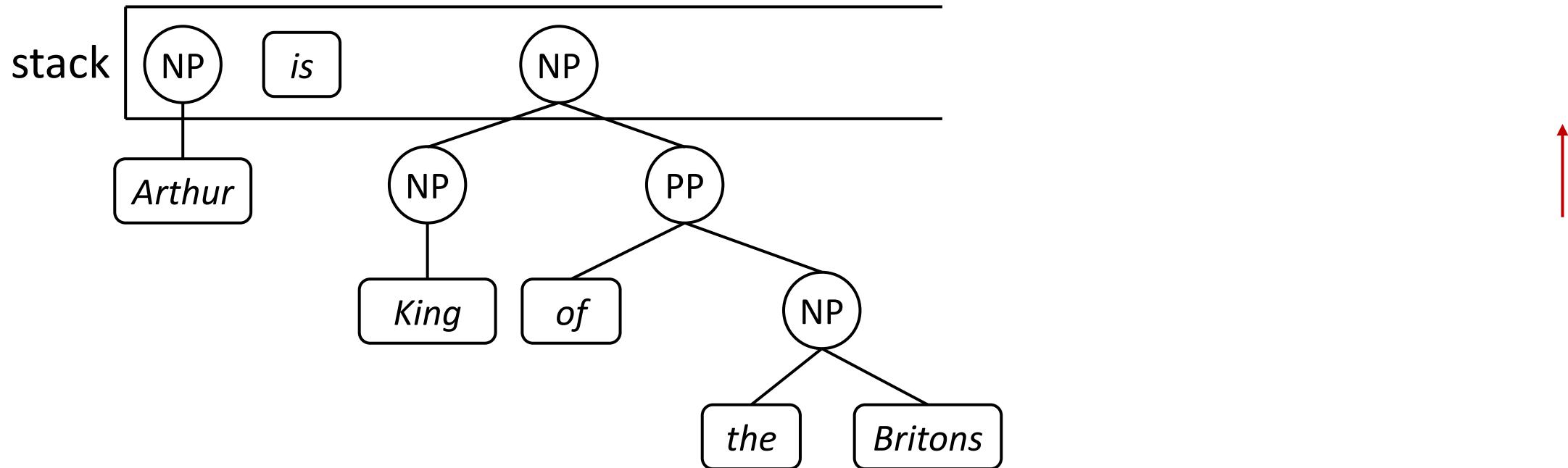
shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary_reduce(PP), **binary_reduce(NP)**

Shift-Reduce Parsers



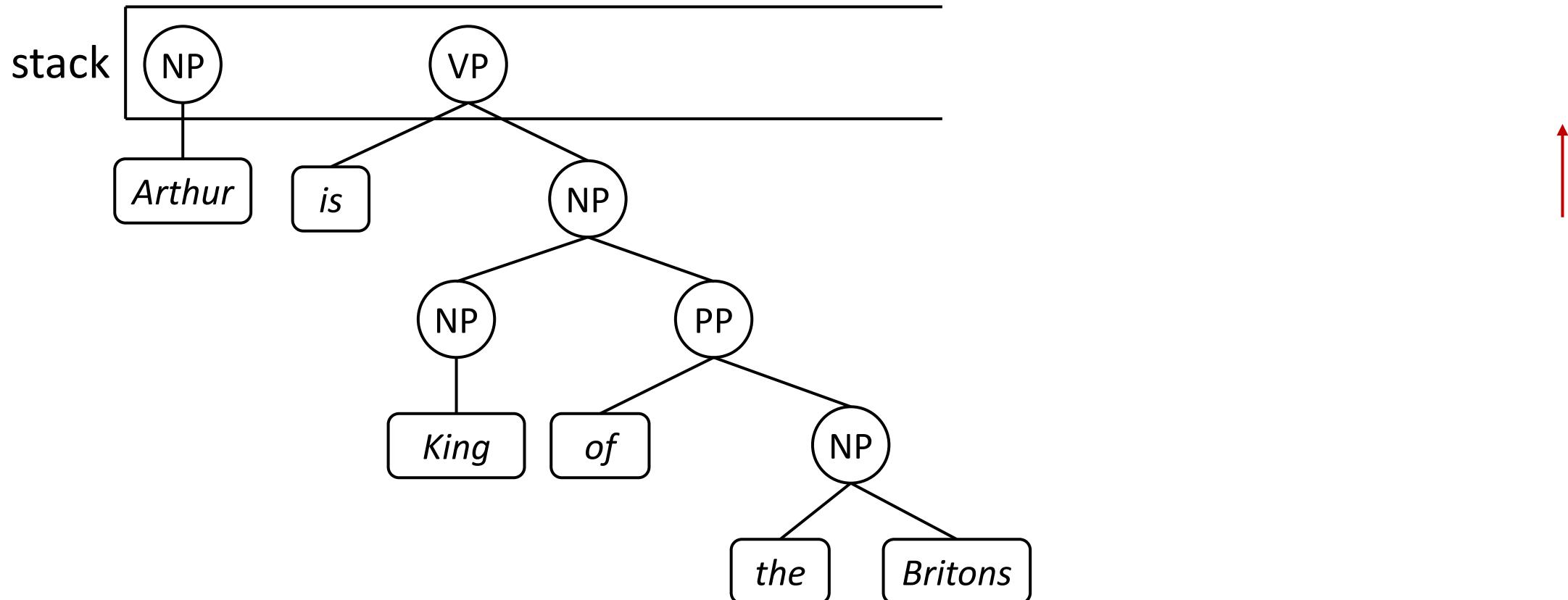
shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary_reduce(PP), **binary_reduce(NP)**

Shift-Reduce Parsers



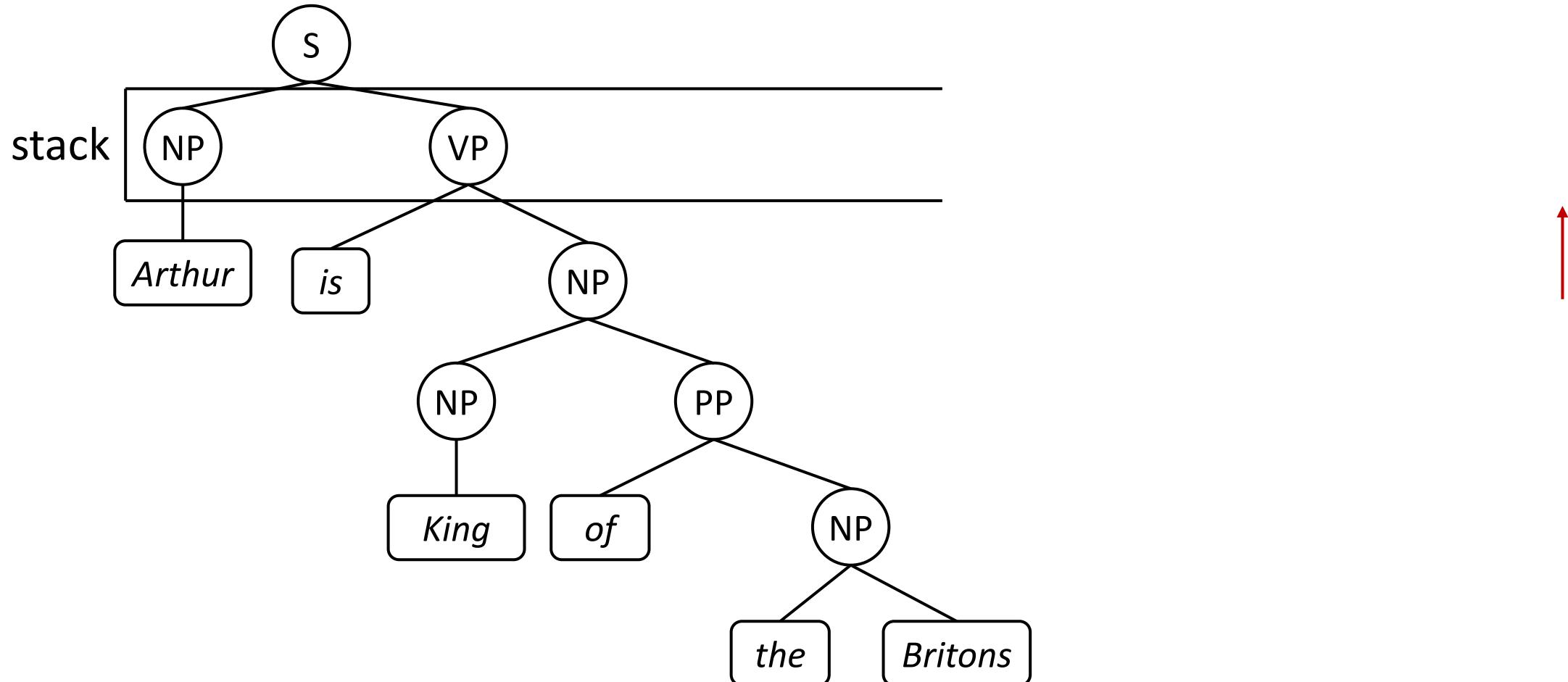
shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary_reduce(PP), binary_reduce(NP), **binary_reduce(VP)**

Shift-Reduce Parsers



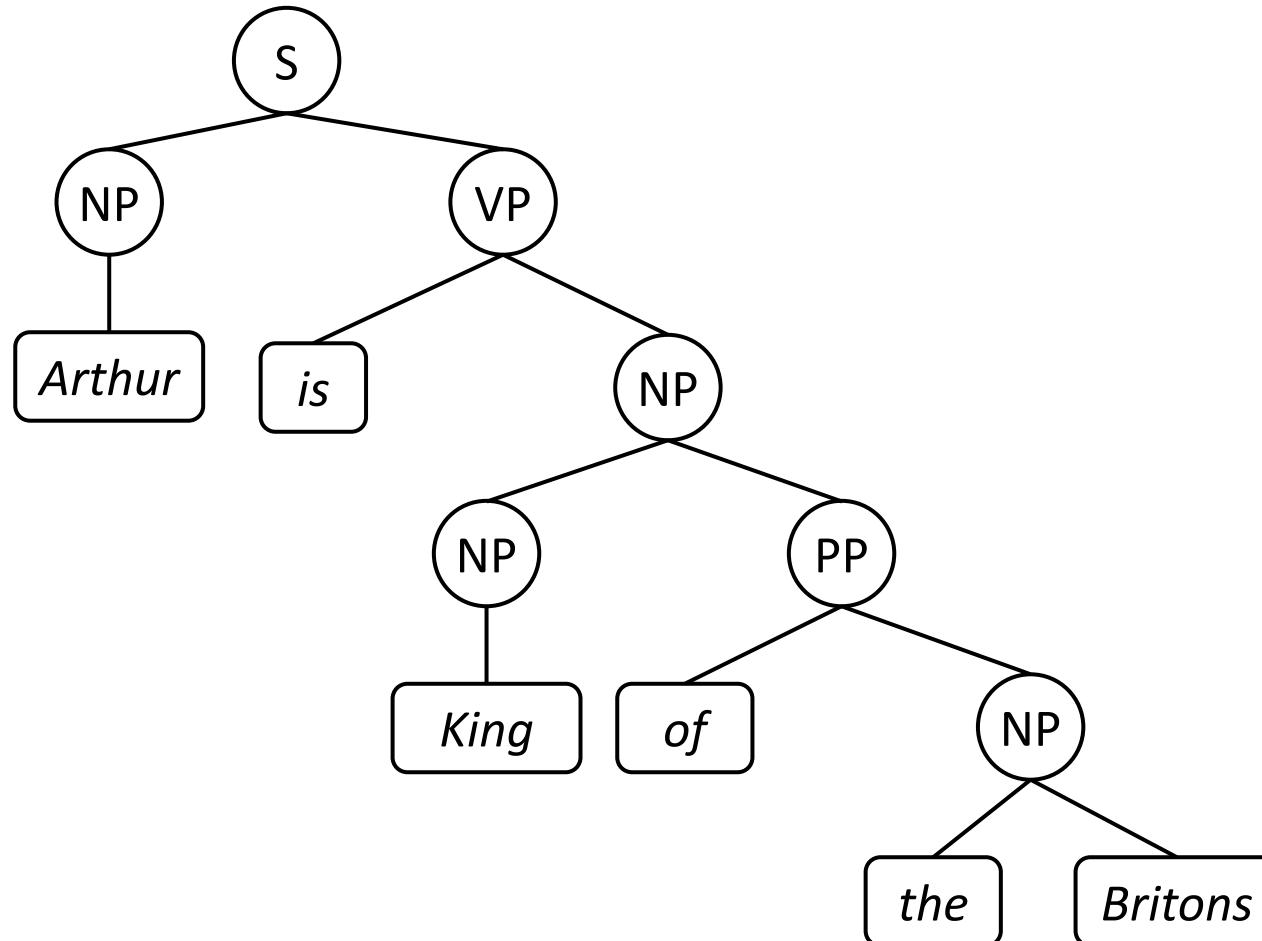
shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary_reduce(PP), binary_reduce(NP), **binary_reduce(VP)**

Shift-Reduce Parsers



shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary_reduce(PP), binary_reduce(NP), binary_reduce(VP), **binary_reduce(S)**

Shift-Reduce Parsers



shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary_reduce(PP), binary_reduce(NP), binary_reduce(VP), binary_reduce(S)

Strongly Incremental Parsing

- Shift-reduce parsers differ from human parsing
- Human parsing appears to be *strongly incremental*

[Marslen-Wilson, 1973]

[Sturt and Lombardo, 2005]

[Stabler, 2015]

Strongly Incremental Parsing

Arthur is King of the Britons



- Human parsing appears to be *strongly incremental*:
 - One word per step: no more, no less

Strongly Incremental Parsing



- Human parsing appears to be *strongly incremental*:
 - One word per step: no more, no less

Strongly Incremental Parsing



- Human parsing appears to be *strongly incremental*:
 - One word per step: no more, no less

Strongly Incremental Parsing



- Human parsing appears to be *strongly incremental*:
 - One word per step: no more, no less

Strongly Incremental Parsing



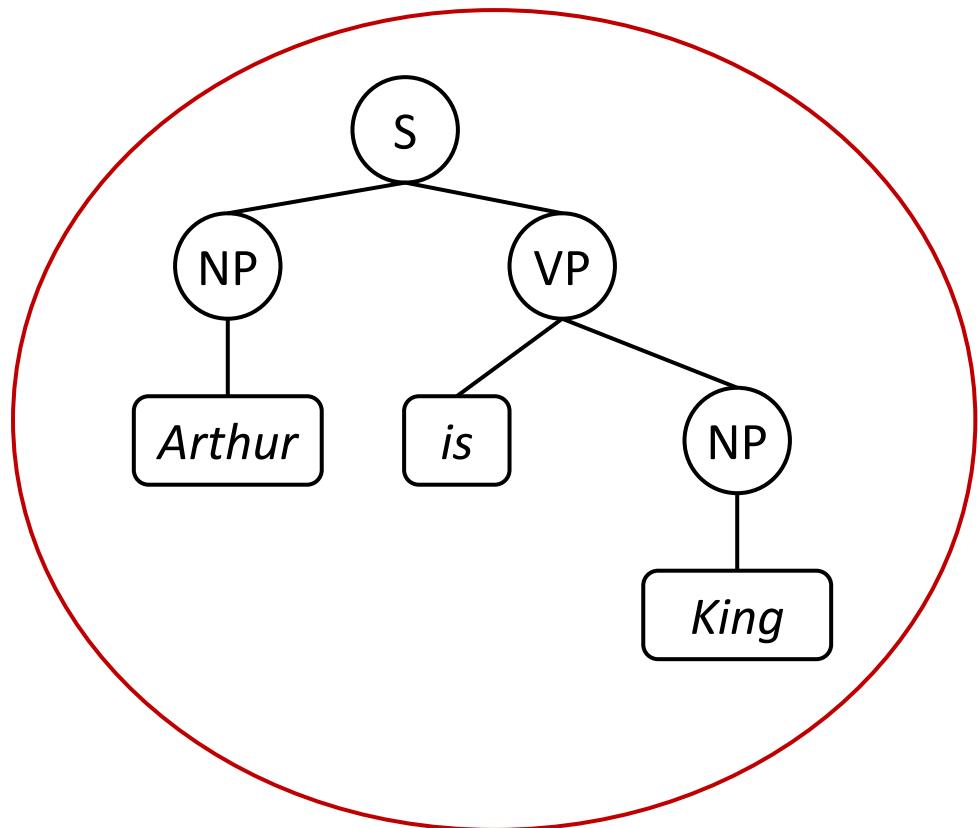
- Human parsing appears to be *strongly incremental*:
 - One word per step: no more, no less

Strongly Incremental Parsing



- Human parsing appears to be *strongly incremental*:
 - One word per step: no more, no less

Strongly Incremental Parsing



of the Britons



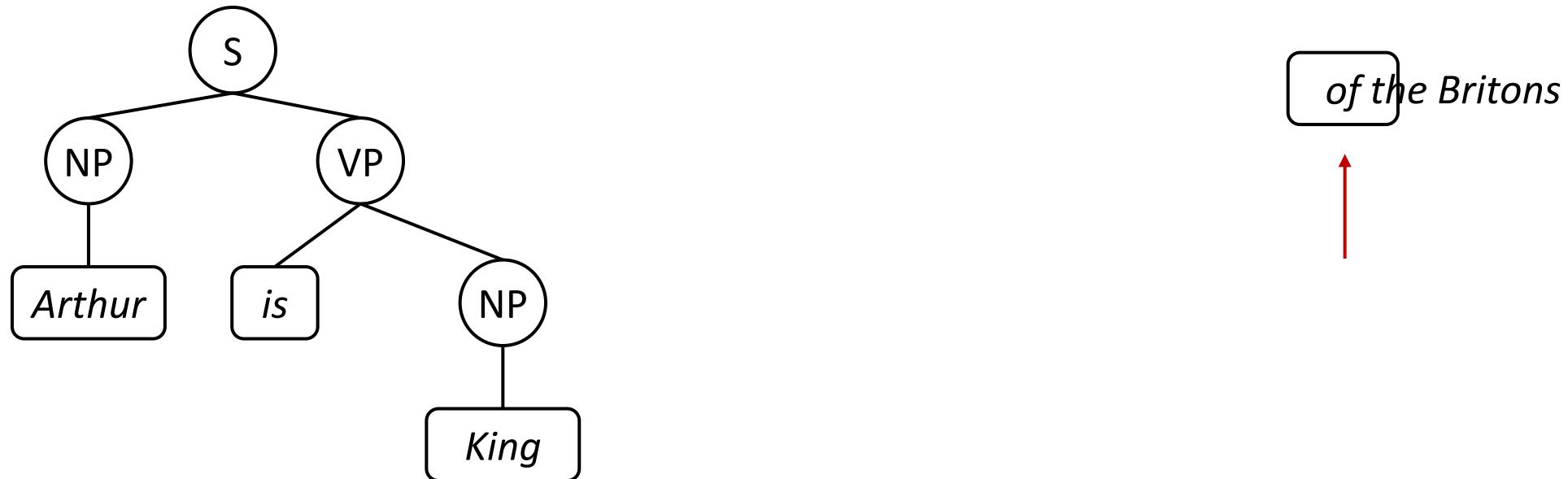
- Human parsing appears to be *strongly incremental*:
 - One word per step: no more, no less
 - A single connected partial parse tree

Attach-Juxtapose Transition System

- We propose a strongly incremental transition system named *attach-juxtapose*

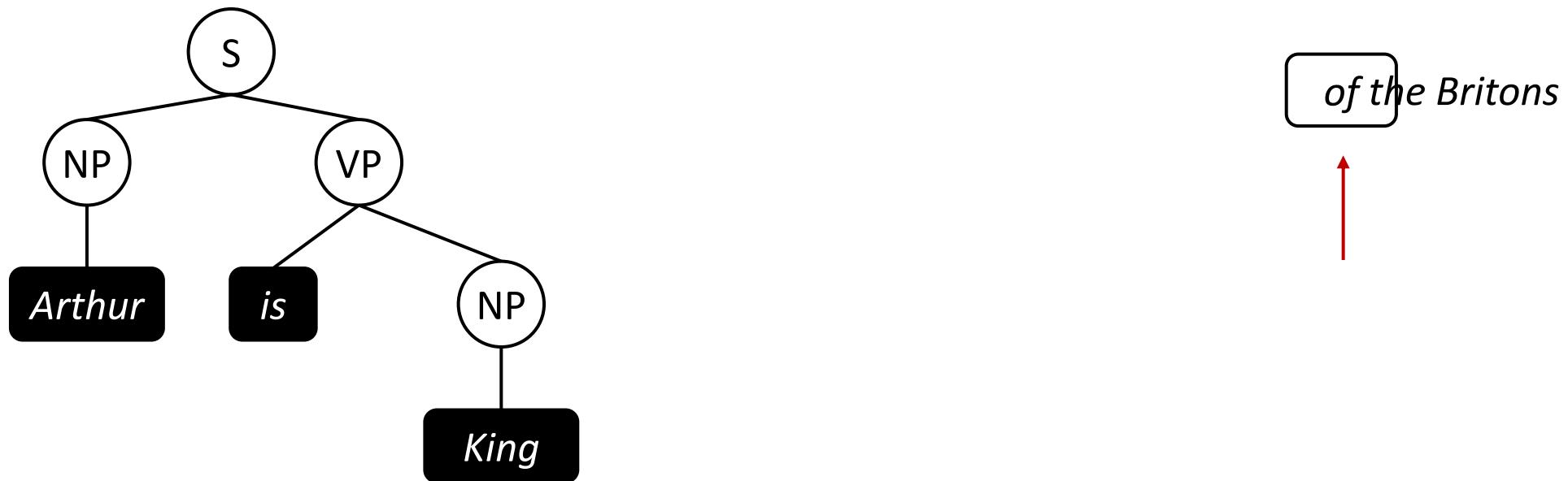
Attach-Juxtapose Transition System

- We propose a strongly incremental transition system named *attach-juxtapose*
- The state is a partial tree and the next word
- Actions determine *where* and *how* to integrate the next word



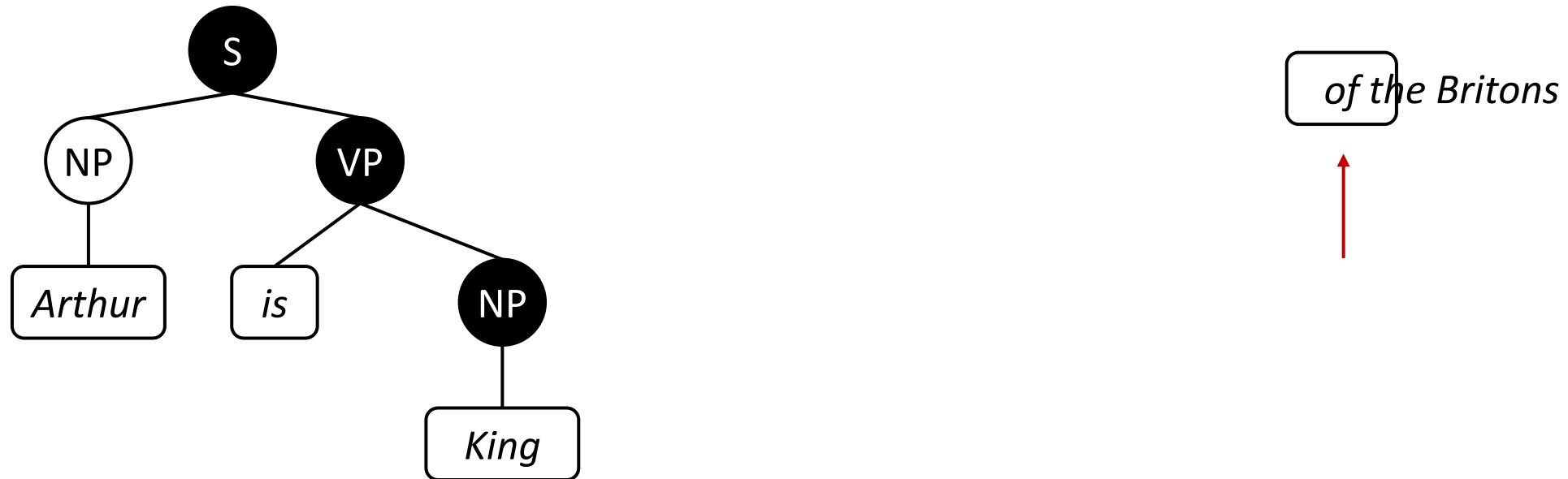
Where to Add the New Word?

- The new word is to the right of existing words, so it must appear on the *rightmost chain*

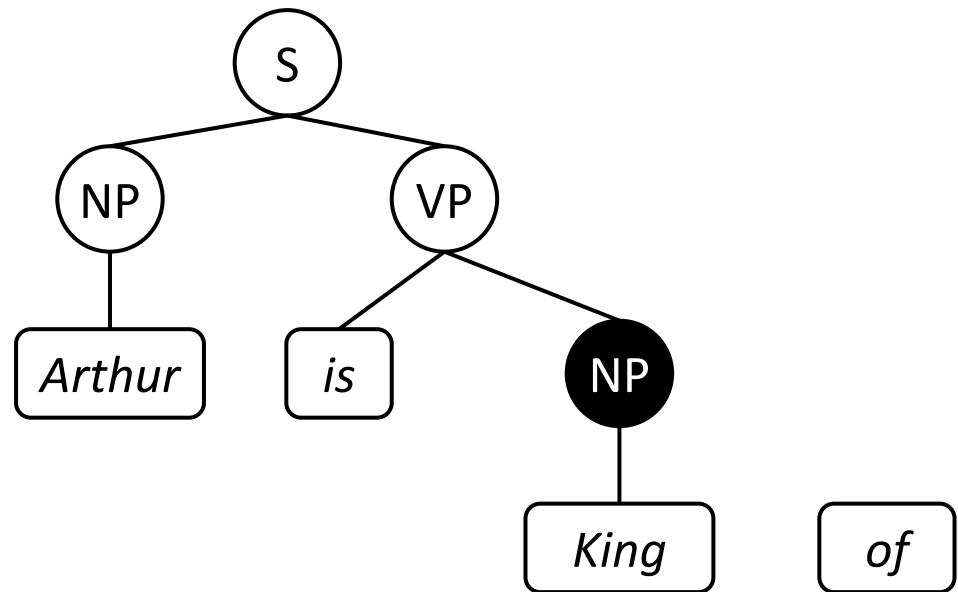


Where to Add the New Word?

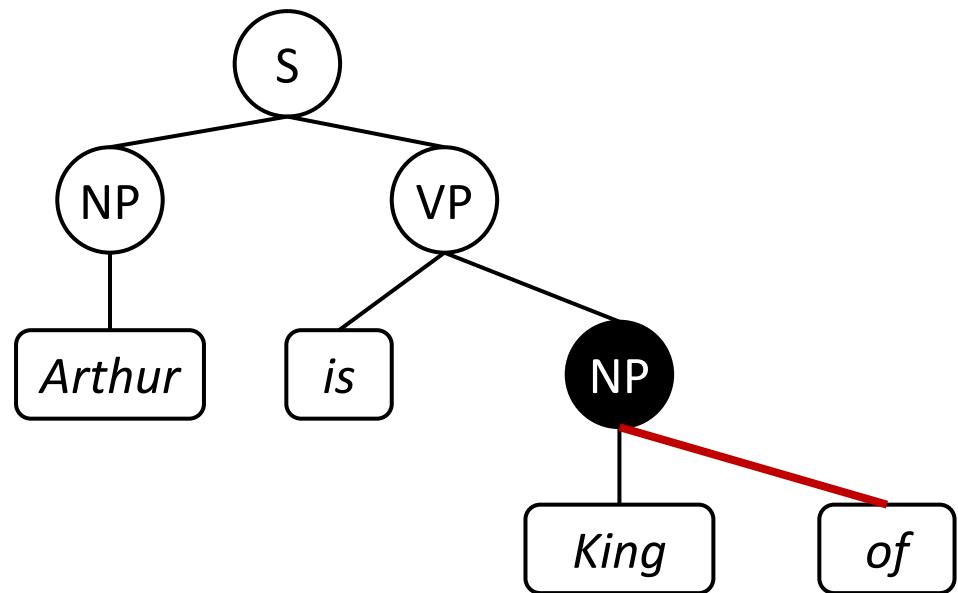
- The new word is to the right of existing words, so it must appear on the *rightmost chain*



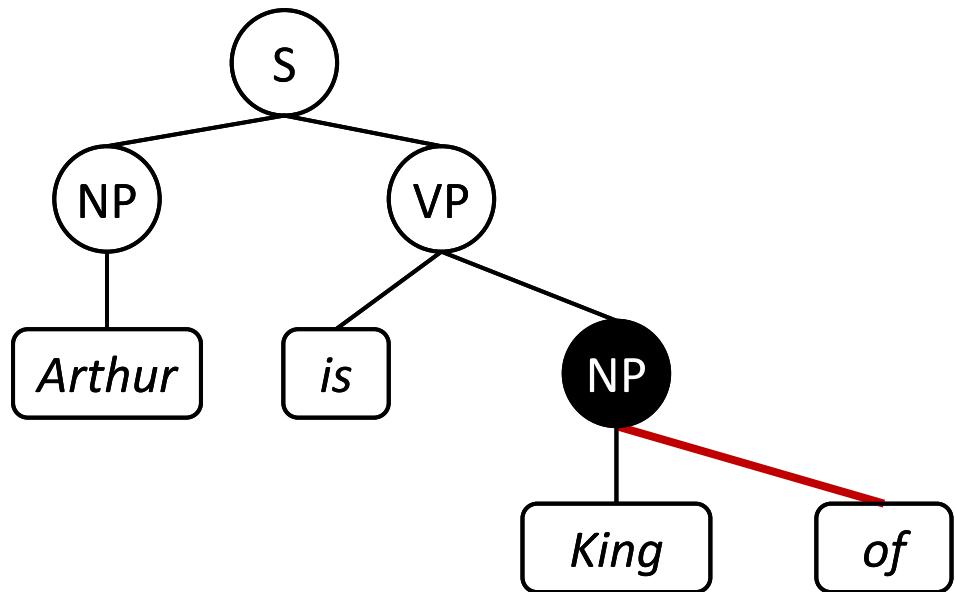
How to Add the New Word?



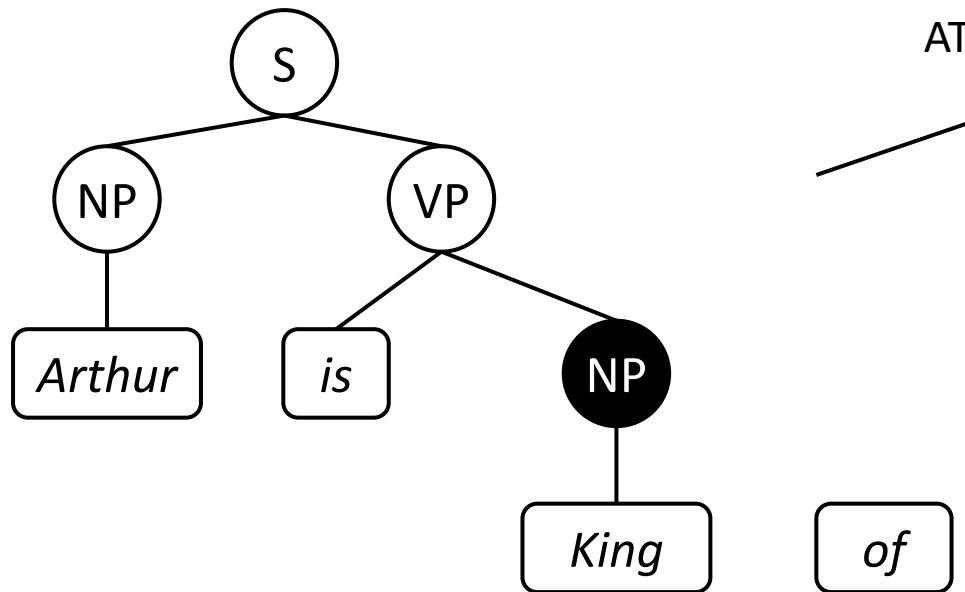
How to Add the New Word?



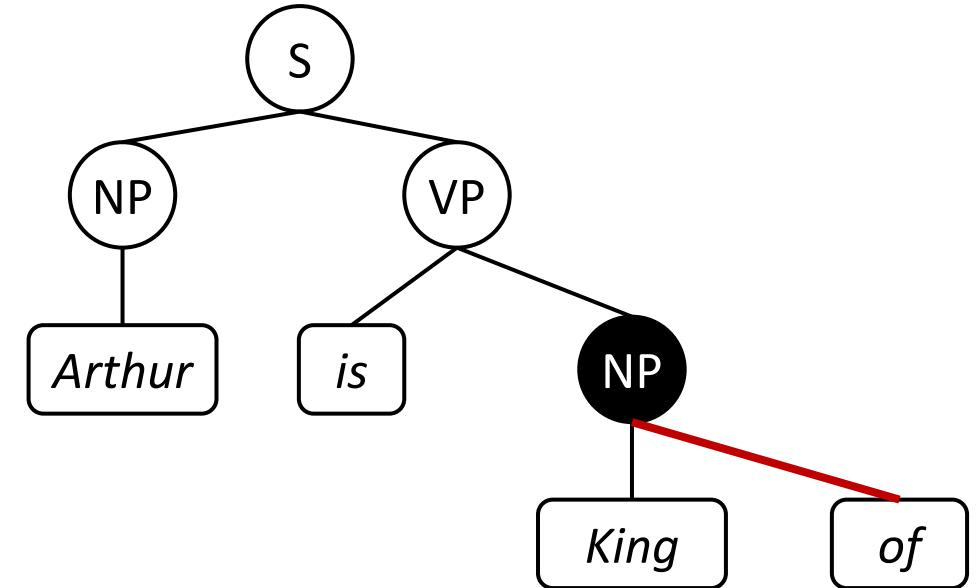
How to Add the New Word?



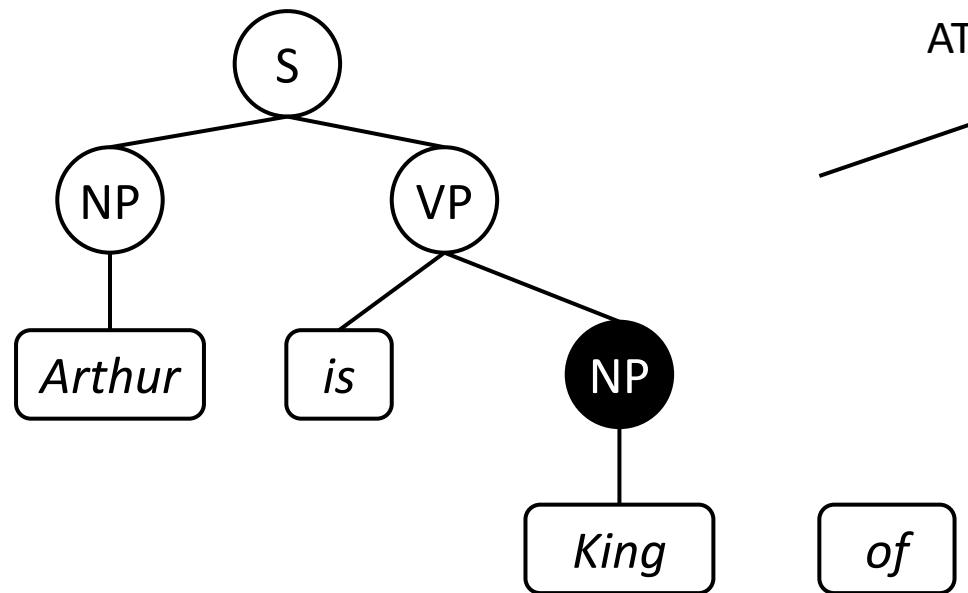
How to Add the New Word?



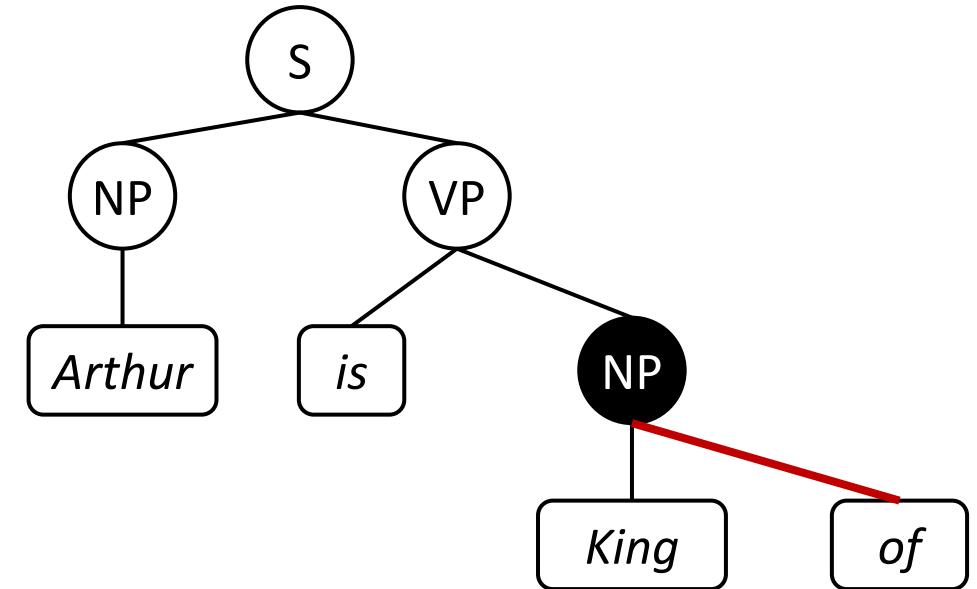
ATTACH



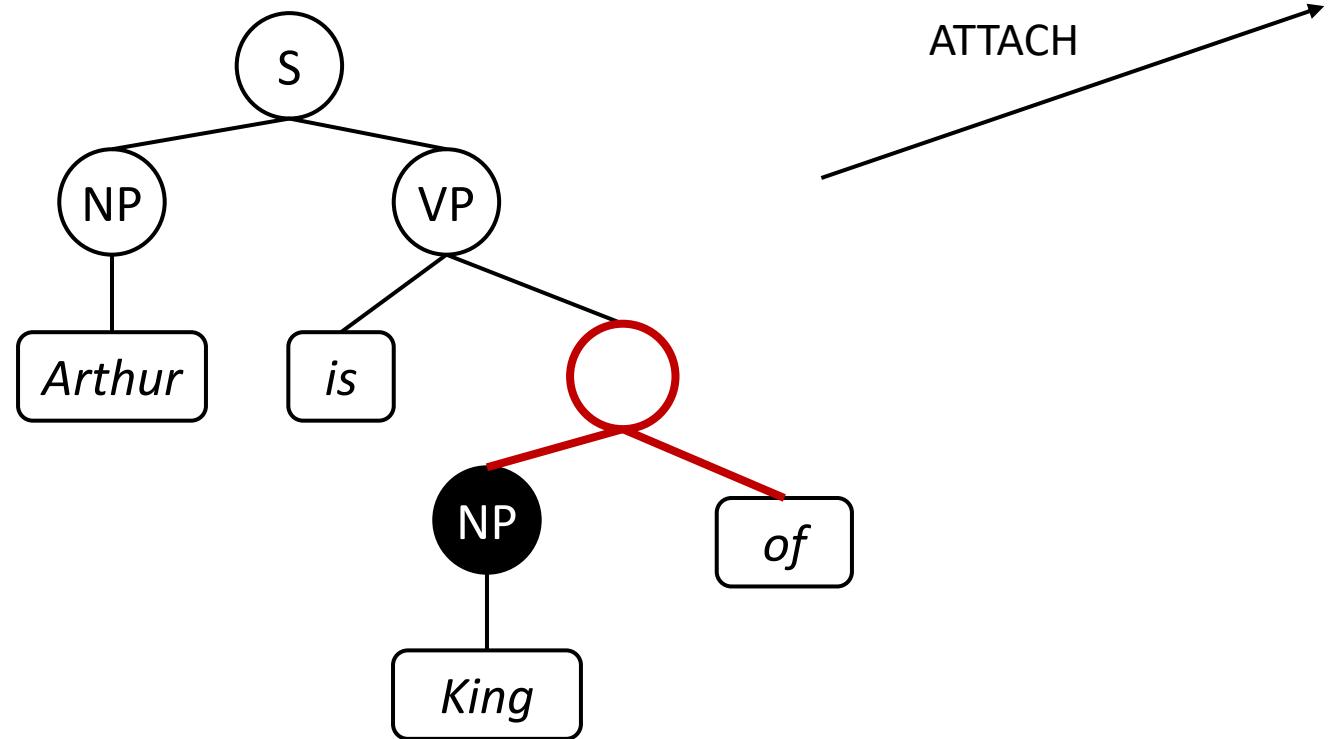
How to Add the New Word?



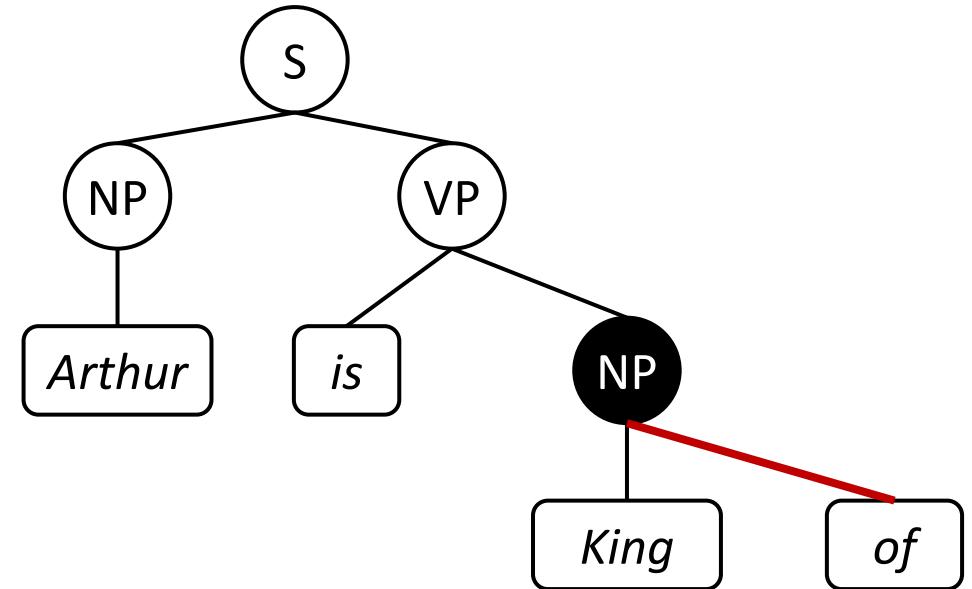
ATTACH



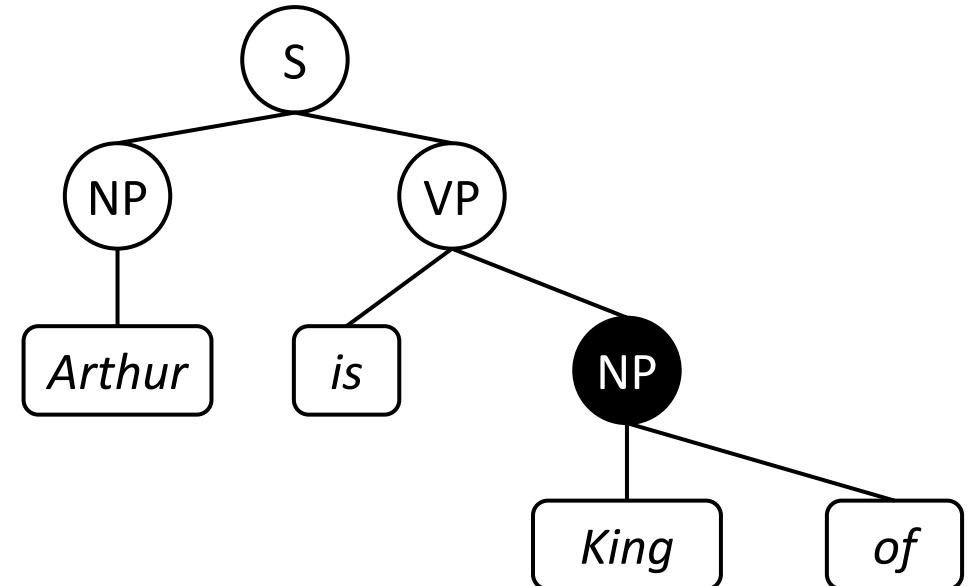
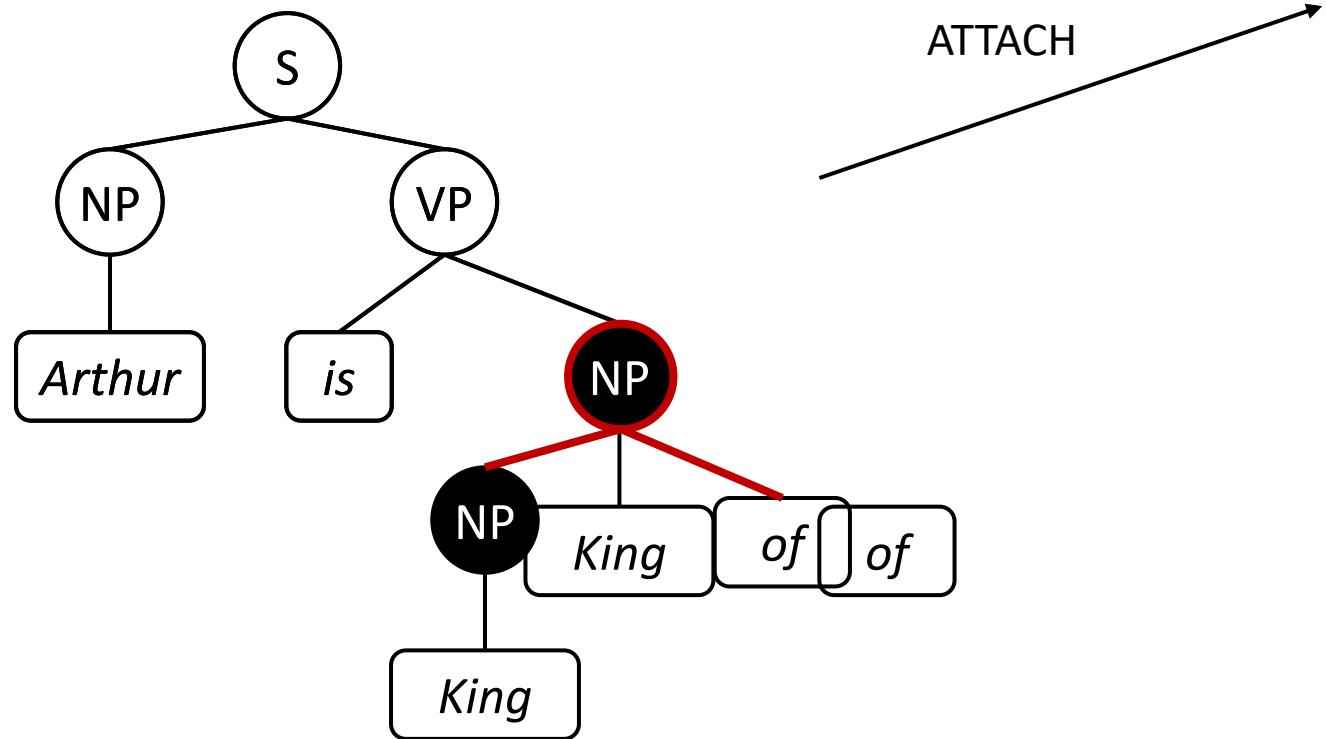
How to Add the New Word?



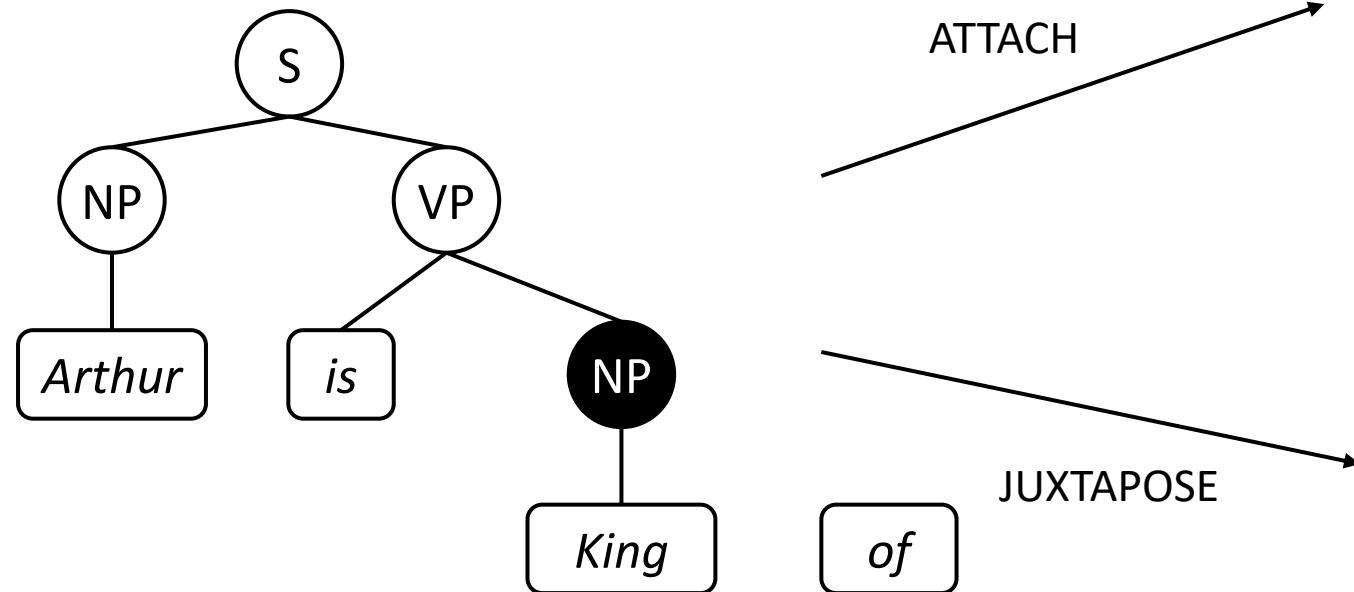
ATTACH



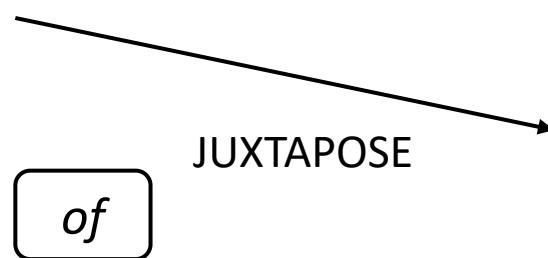
How to Add the New Word?



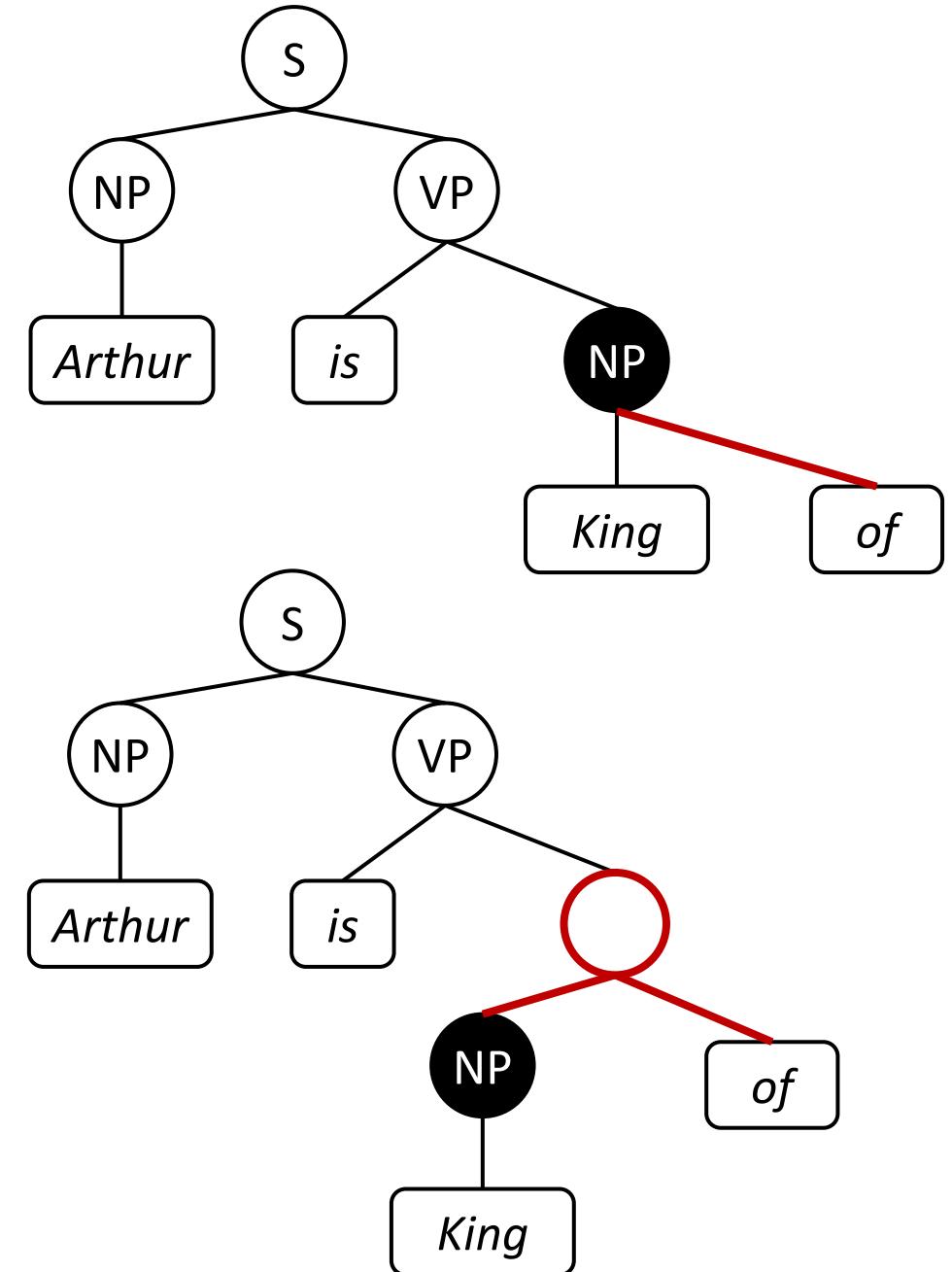
How to Add the New Word?



ATTACH

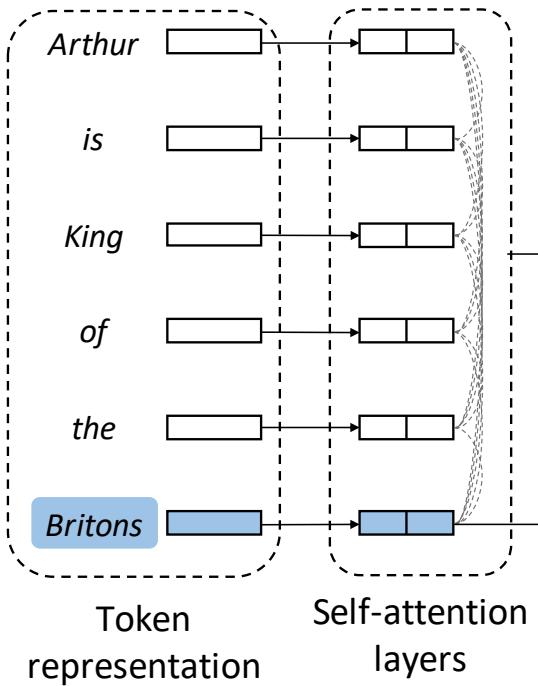


JUXTAPOSE



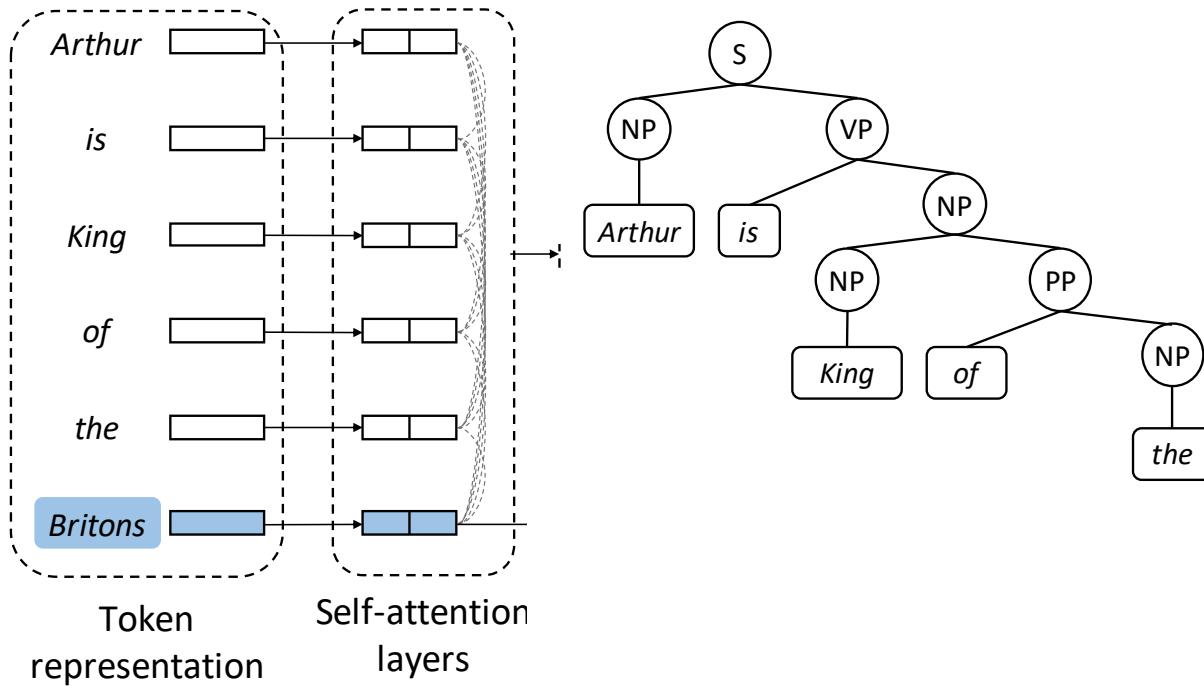
Action Generation with Graph Neural Networks

- Encoder: BERT/XLNet + additional self-attention layers



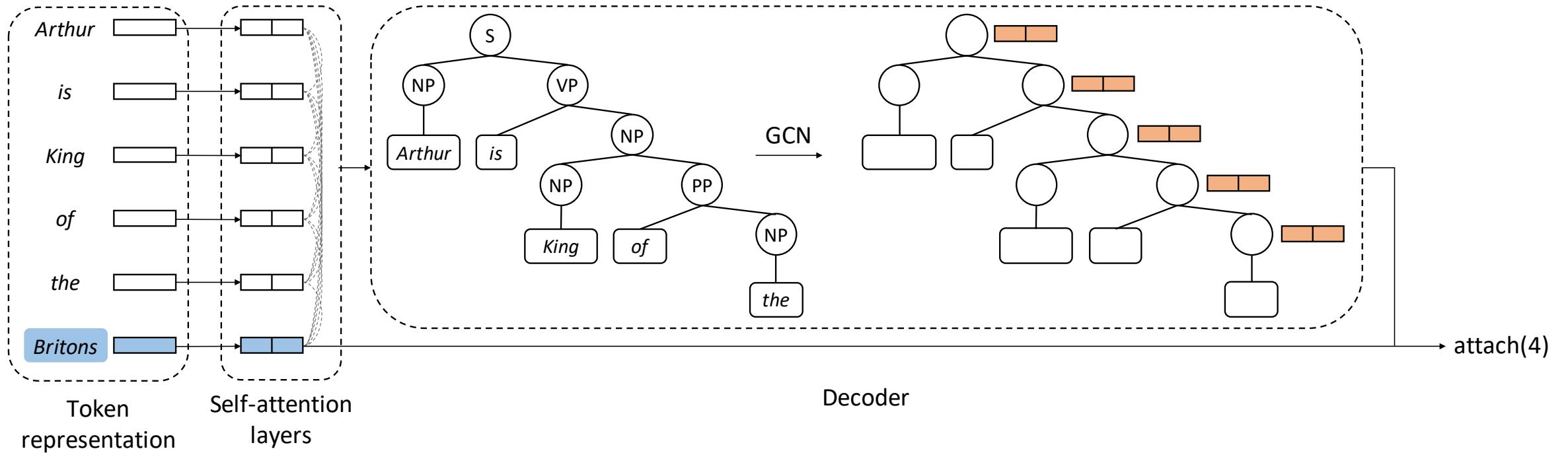
Action Generation with Graph Neural Networks

- Encoder: BERT/XLNet + additional self-attention layers
- Decoder: Generate attach-juxtapose actions by applying GNNs on the partial tree



Action Generation with Graph Neural Networks

- Encoder: BERT/XLNet + additional self-attention layers
- Decoder: Generate attach-juxtapose actions by applying GNNs on the partial tree



Experimental Results

- Competitive with state of the art on Penn Treebank

Model	EM	F1	LP	LR	#Params
Liu and Zhang [22]	-	91.8	-	-	-
Liu and Zhang [22] (BERT) [†]	57.05	95.71	-	-	-
Kitaev and Klein [21]	47.31	93.55	93.90	93.20	26M
Kitaev and Klein [21] (ELMo)	53.06	95.13	95.40	94.85	107M
Kitaev et al. [20] (BERT)	-	95.59	95.46	95.73	342M
Zhou and Zhao [49] (GloVe) *	47.72	93.78	93.92	93.64	51M
Zhou and Zhao [49] (BERT) *	55.84	95.84	95.98	95.70	349M
Zhou and Zhao [49] (XLNet) *	<u>58.73</u>	96.33	96.46	<u>96.21</u>	374M
Mrini et al. [27] (XLNet) *	58.65	96.38	<u>96.53</u>	96.24	459M
Ours (BERT)	57.29 ± 0.57	95.79 ± 0.05	96.04 ± 0.05	95.55 ± 0.06	377M
Ours (XLNet)	59.17 ± 0.33	<u>96.34 ± 0.03</u>	96.55 ± 0.02	96.13 ± 0.04	391M

Experimental Results

- Competitive with state of the art on Penn Treebank
- Improves upon state of the art on Chinese Treebank

Model	EM	F1	LP	LR
Kitaev et al. [20]	-	91.75	91.96	91.55
Kitaev et al. [20] (BERT) [†]	44.42	92.14	-	-
Zhou and Zhao [49] *	-	92.18	92.33	<u>92.03</u>
Mrini et al. [27] (BERT) *	-	<u>92.64</u>	<u>93.45</u>	91.85
Liu and Zhang [22]	-	86.1	-	-
Liu and Zhang [22] (BERT) [†]	<u>44.94</u>	91.81	-	-
Ours (BERT)	49.72 ± 0.83	93.59 ± 0.26	93.80 ± 0.26	93.40 ± 0.28

Strongly Incremental Constituency Parsing with Graph Neural Networks

Kaiyu Yang and Jia Deng



<https://github.com/princeton-vl/attach-juxtapose-parser>

