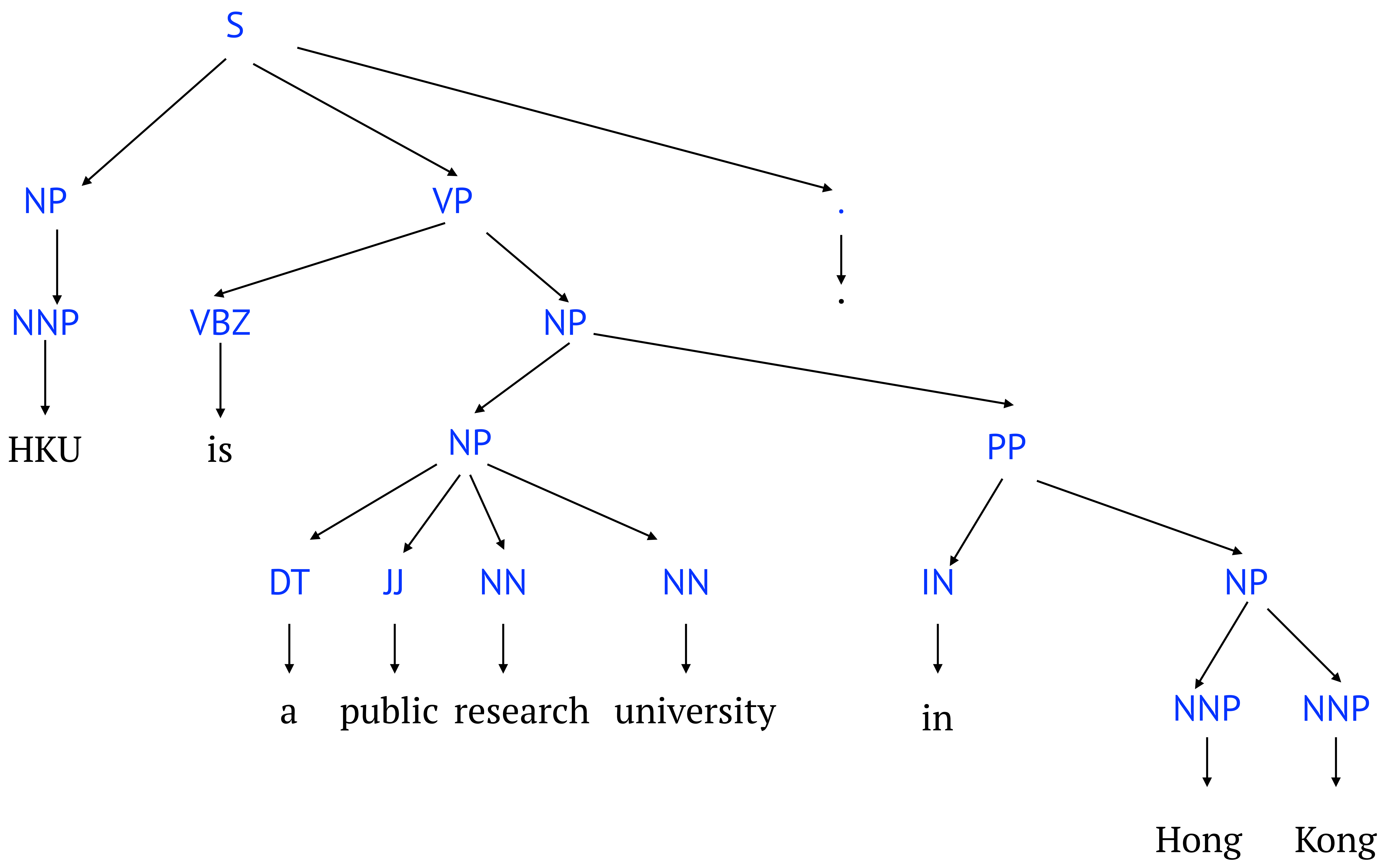
Recursive Neural Networks, Shift-reduce Parsing and Stack-LSTMs

COMP3361 — Week 7

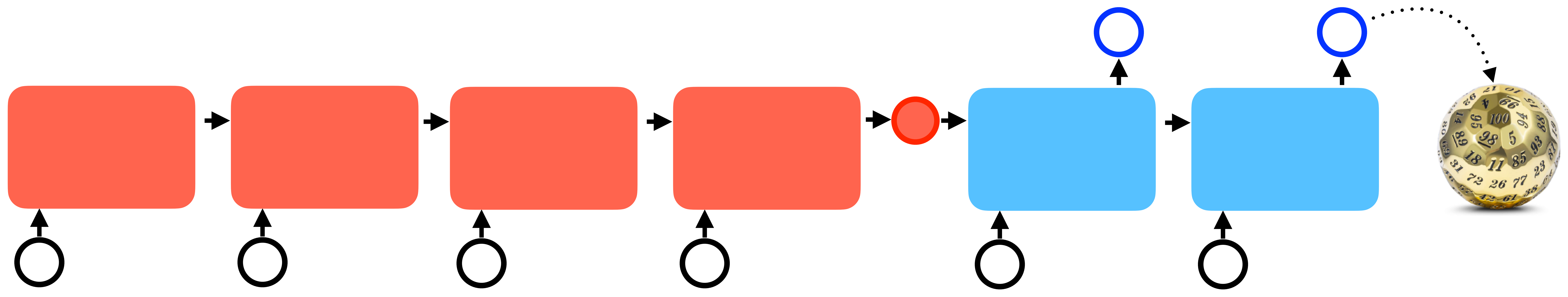
Lingpeng Kong

Department of Computer Science, The University of Hong Kong

Parse Trees


S
NP
VP
.
NNP
VBZ
NP
.
HKU
is
NP
PP
DT
JJ
NN
NN
IN
NP
a
public
research
university
in
NNP
NNP
Hong
Kong

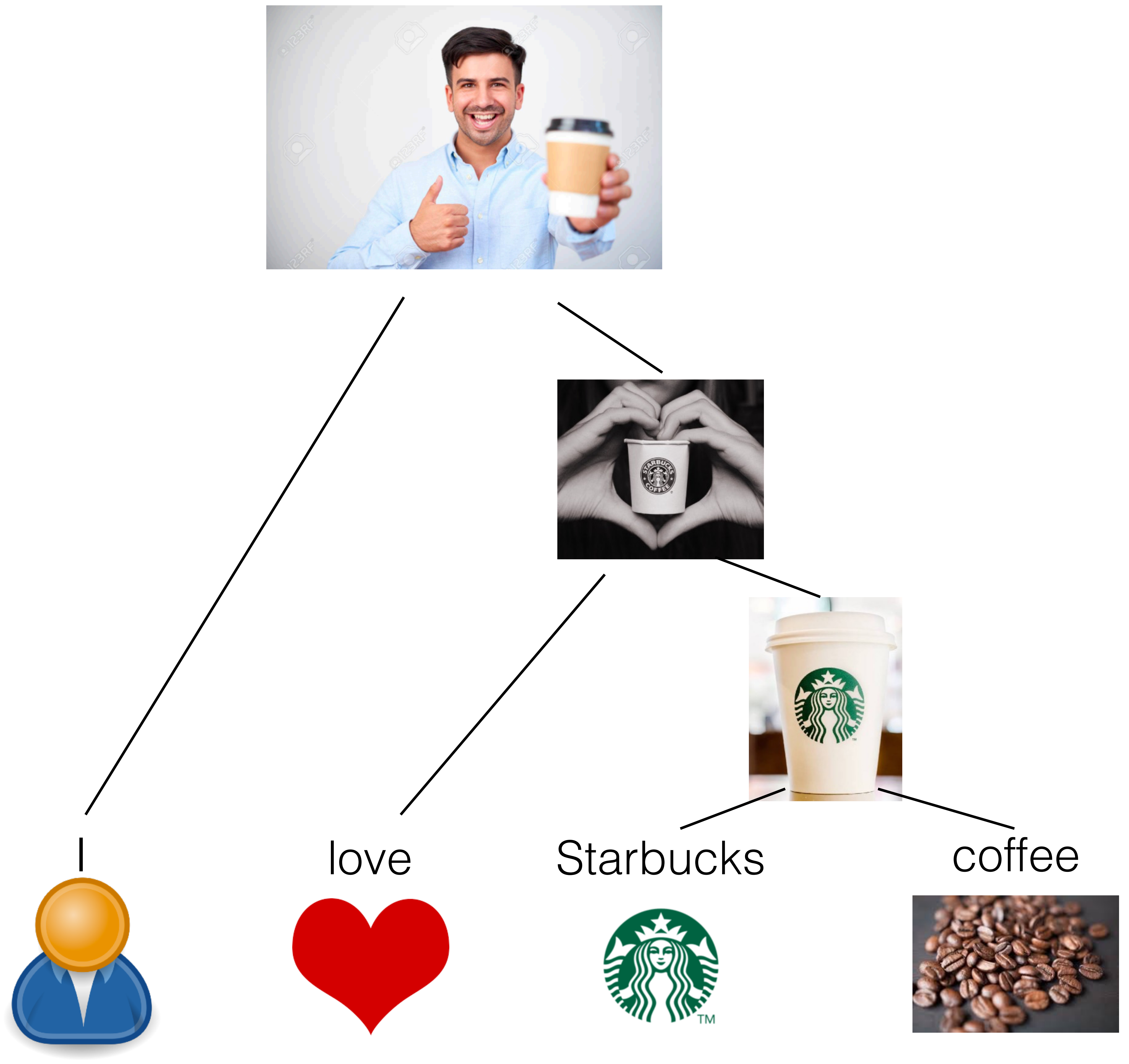
Sequence to Sequence Model



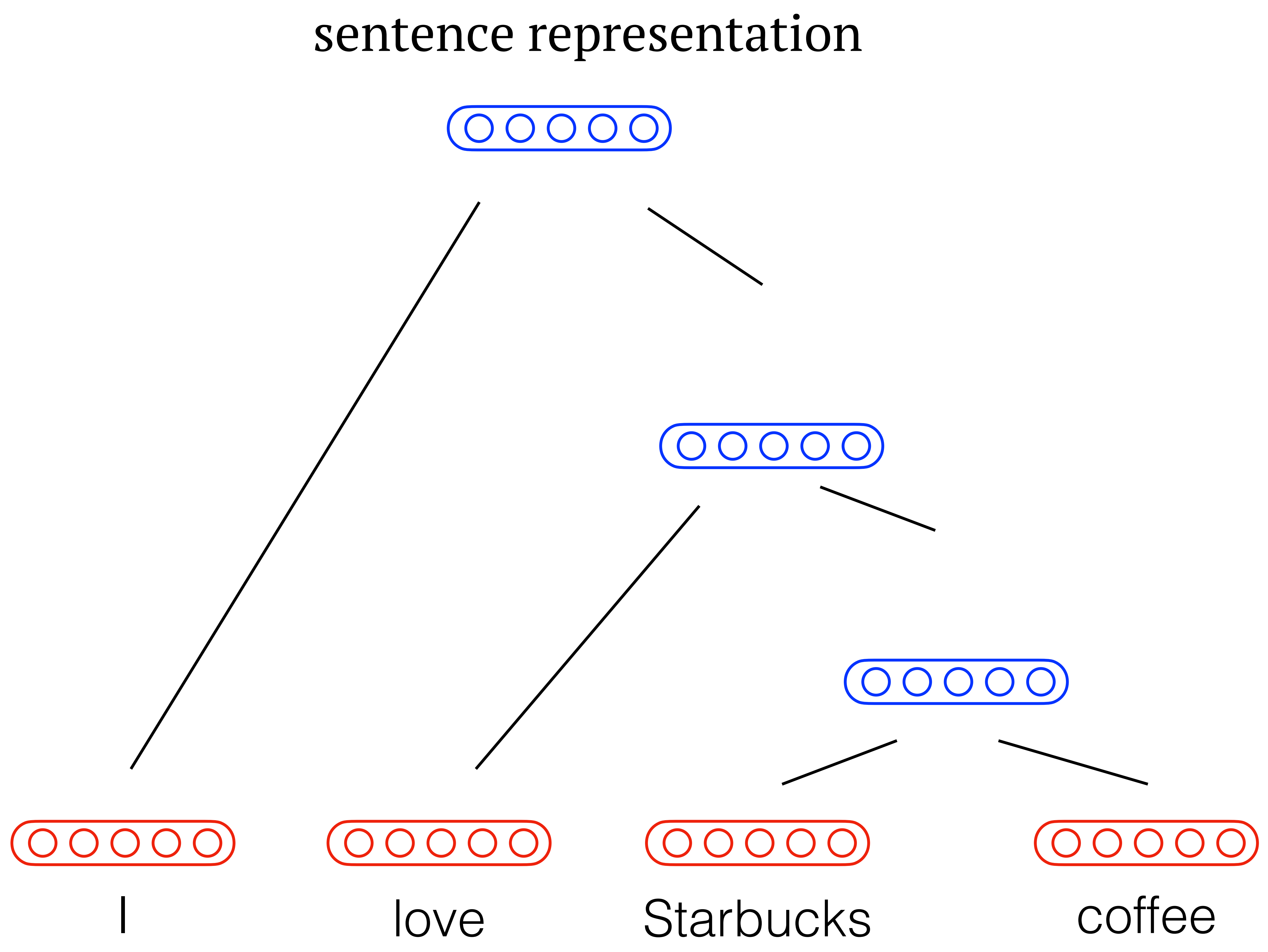
Encoder

Decoder

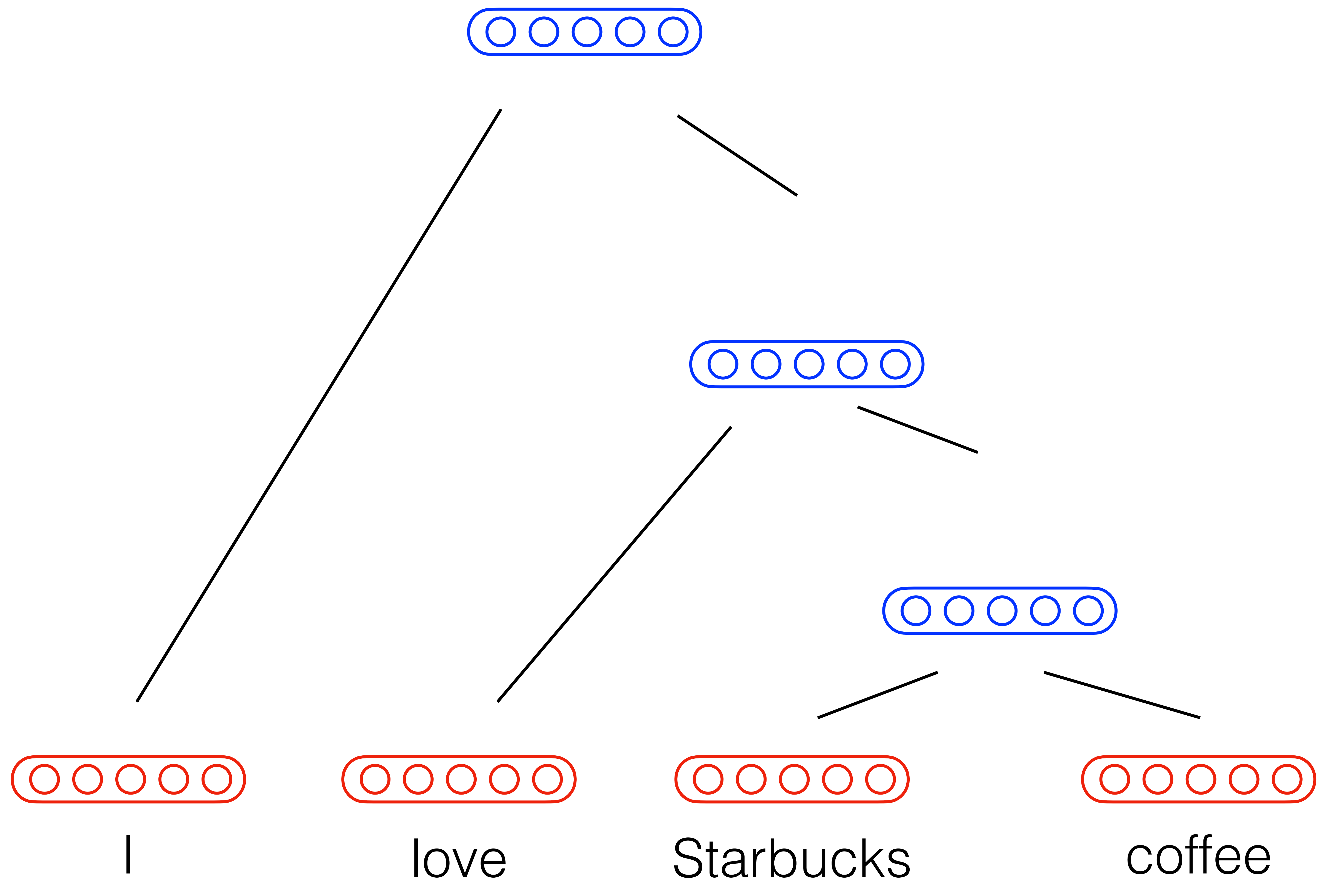
Recursive Neural Networks as Encoder



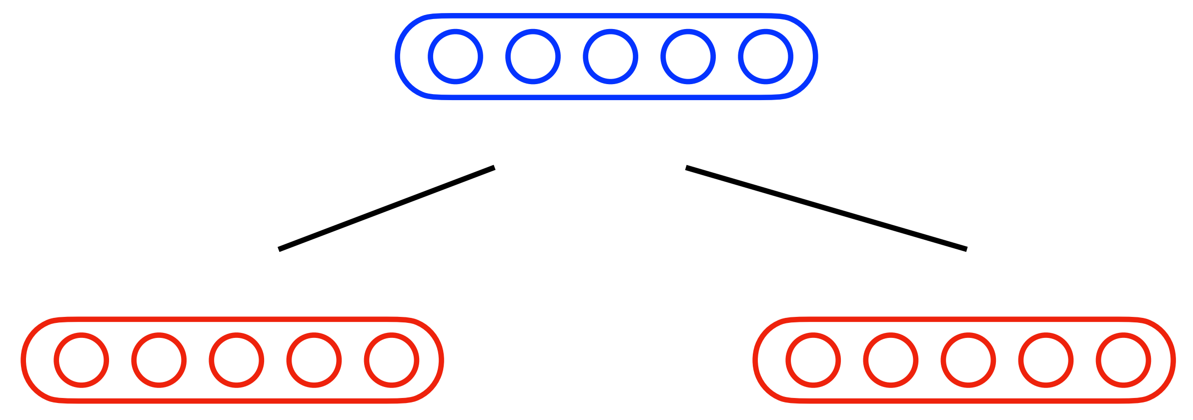
Recursive Neural Networks as Encoder

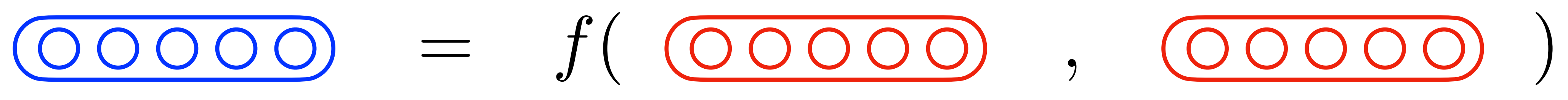


Recursive Neural Networks as Encoder

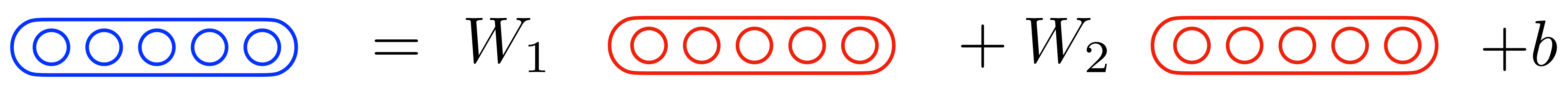


compositional function:



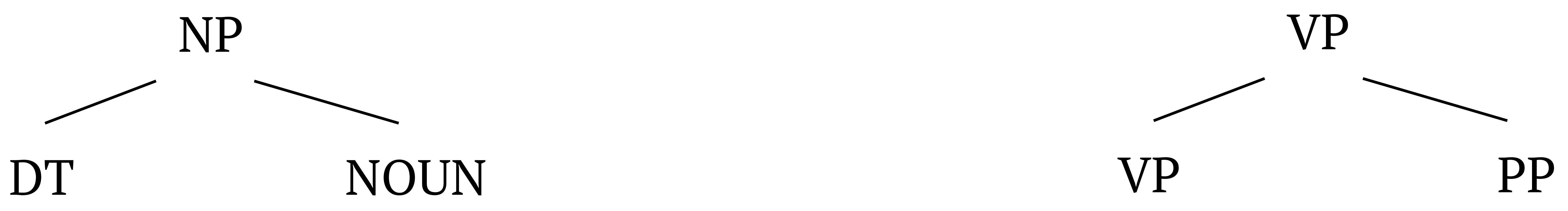


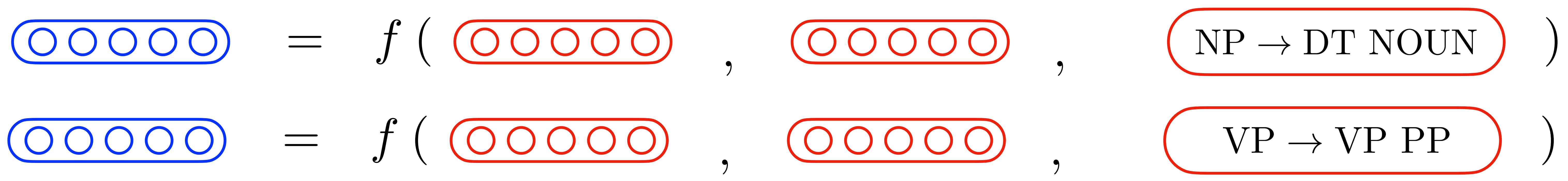
for example:



Recursive Neural Networks as Encoder

compositional function:

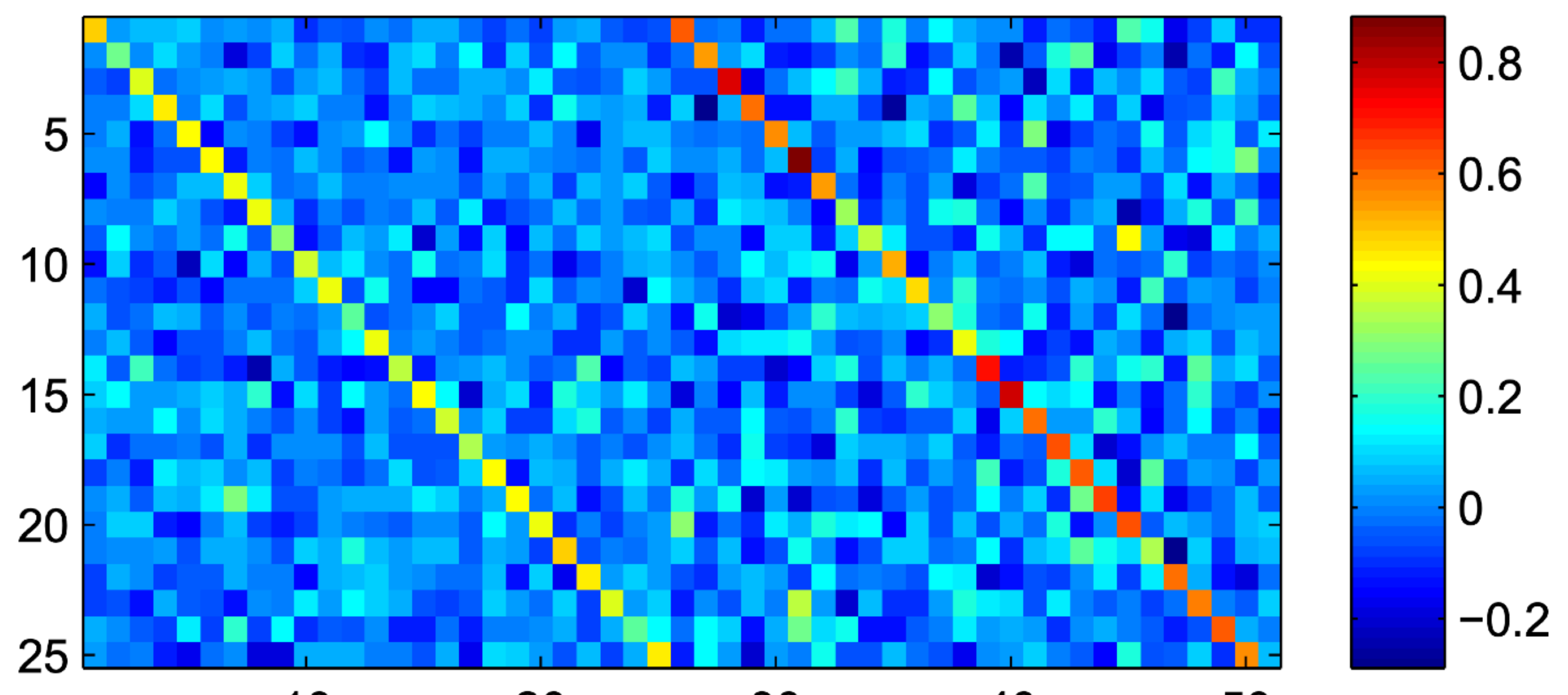




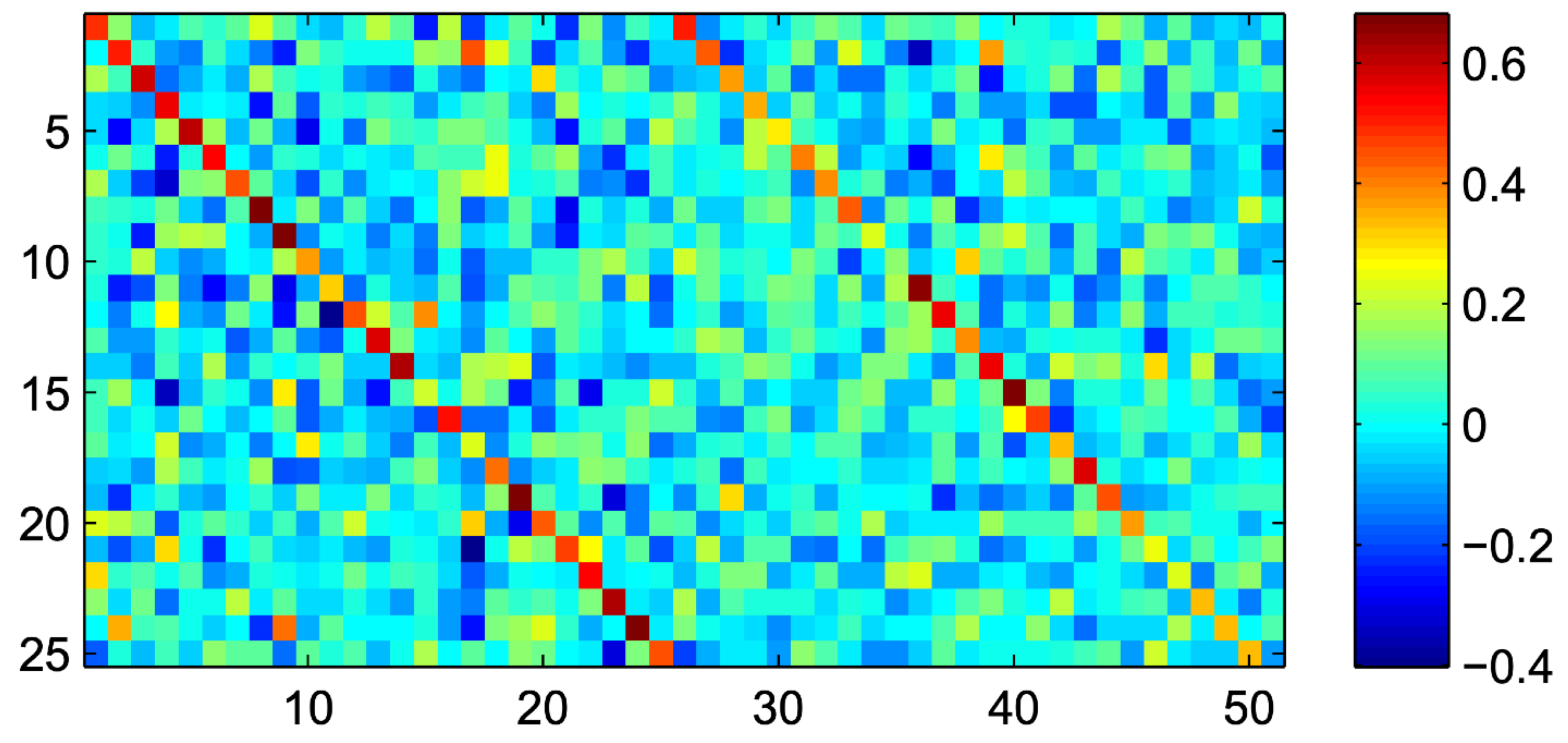
what's good about it?

Recursive Neural Networks as Encoder

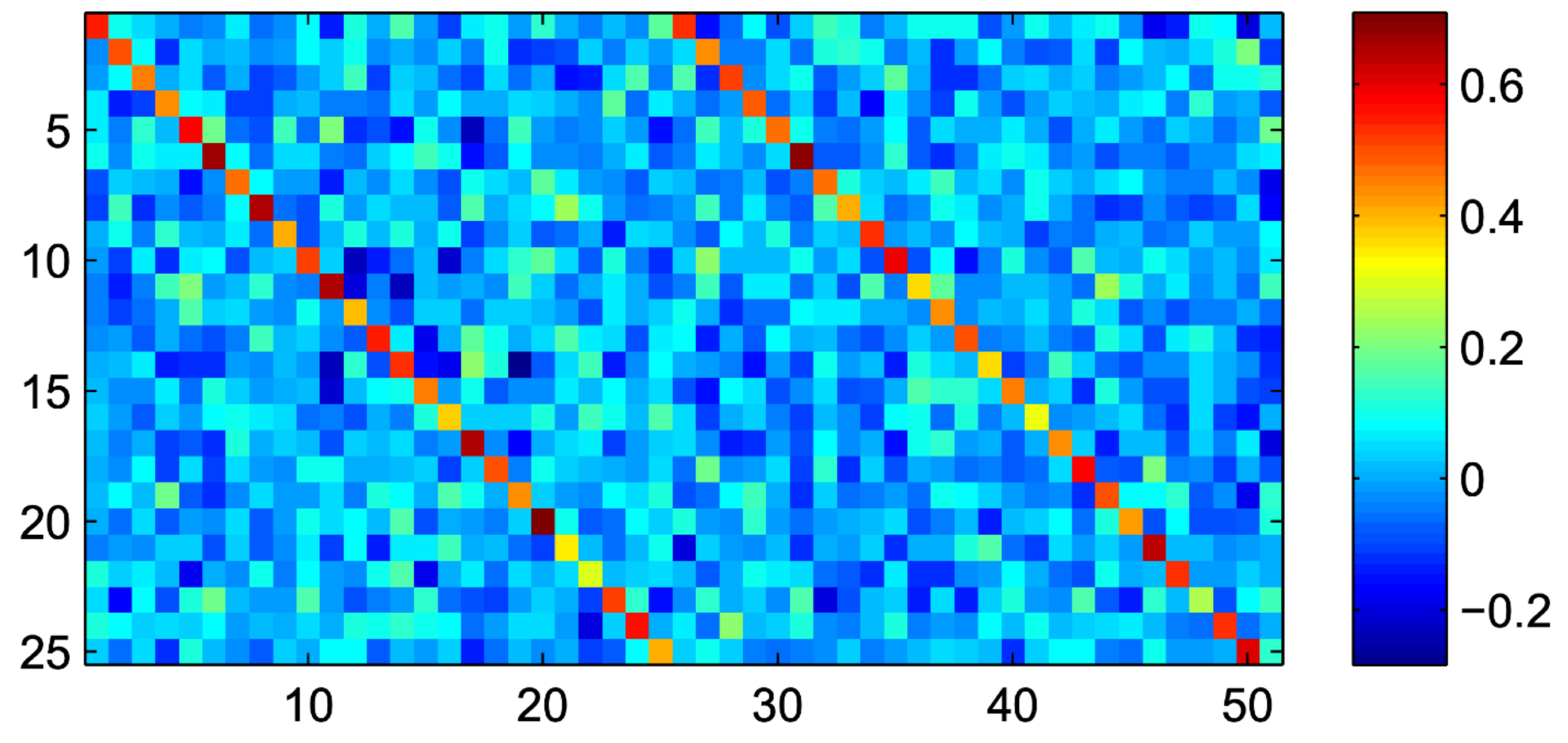
compositional function:



DT-NP



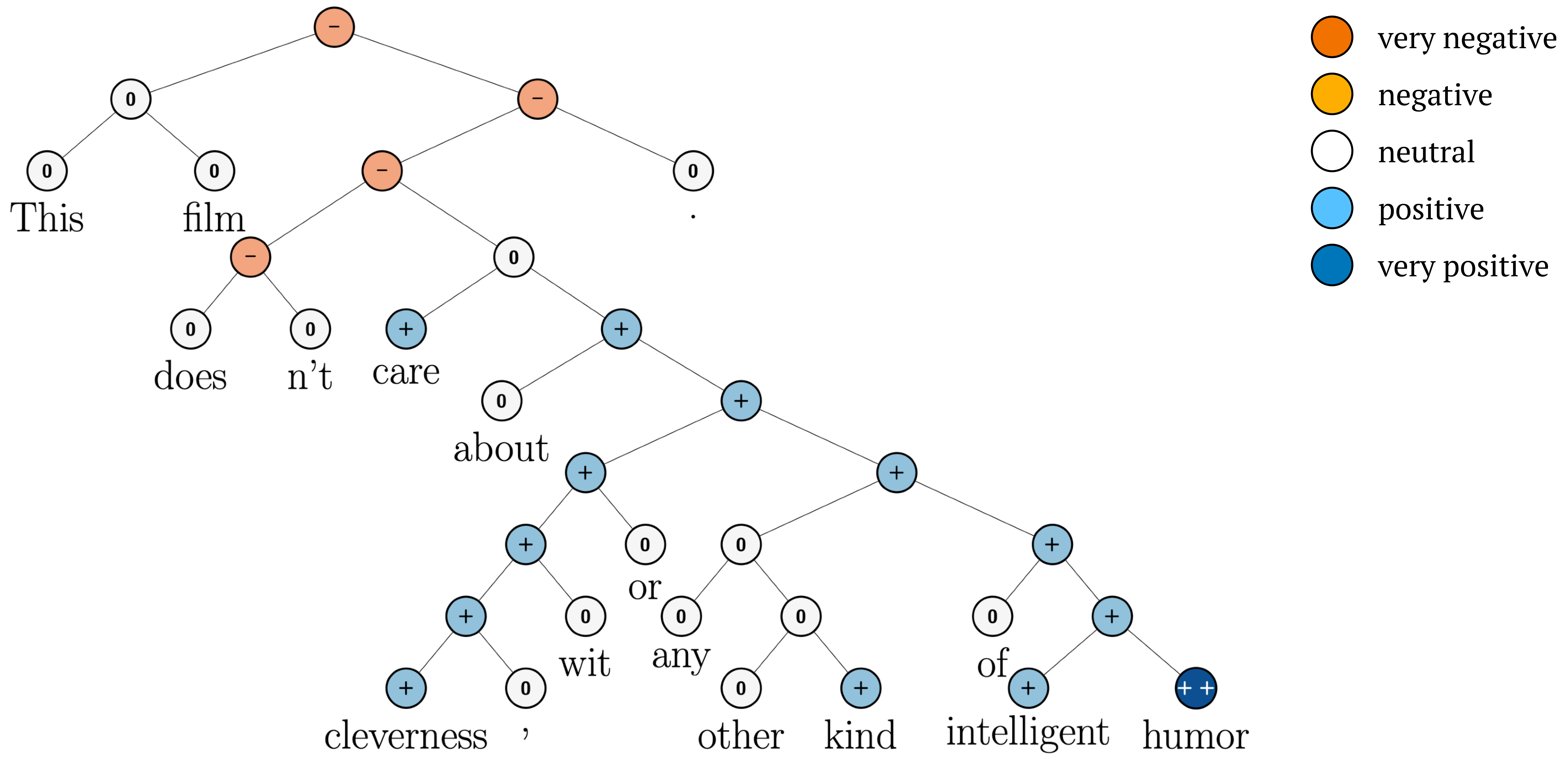
VP-NP



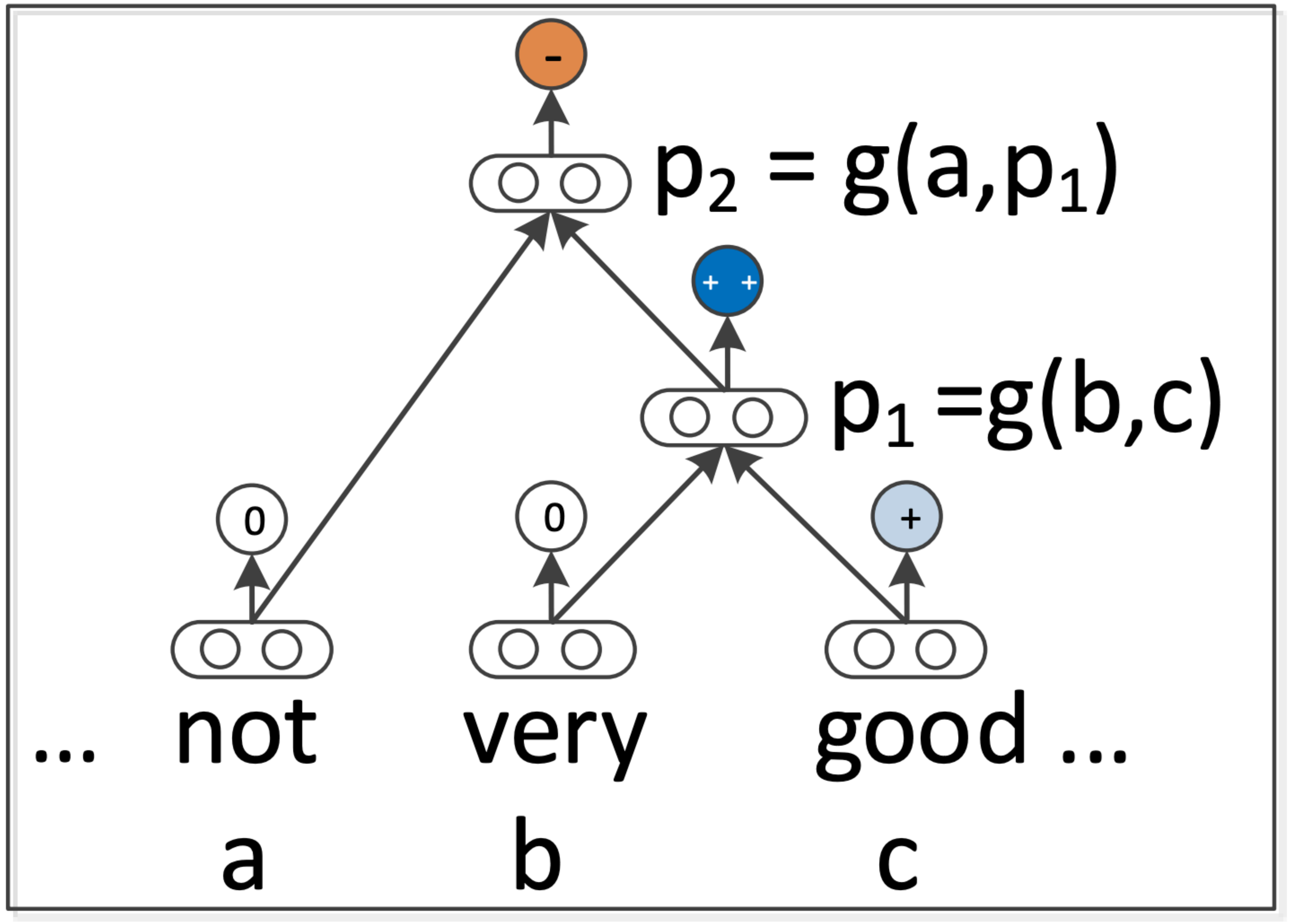
ADJP-NP

Parsing with Compositional Vector Grammars, Socher et al, 2013

Stanford Sentiment Treebank



Training in Recursive Neural Network



$$\mathrm{softmax}(Wa)$$

Classification with 5 classes:

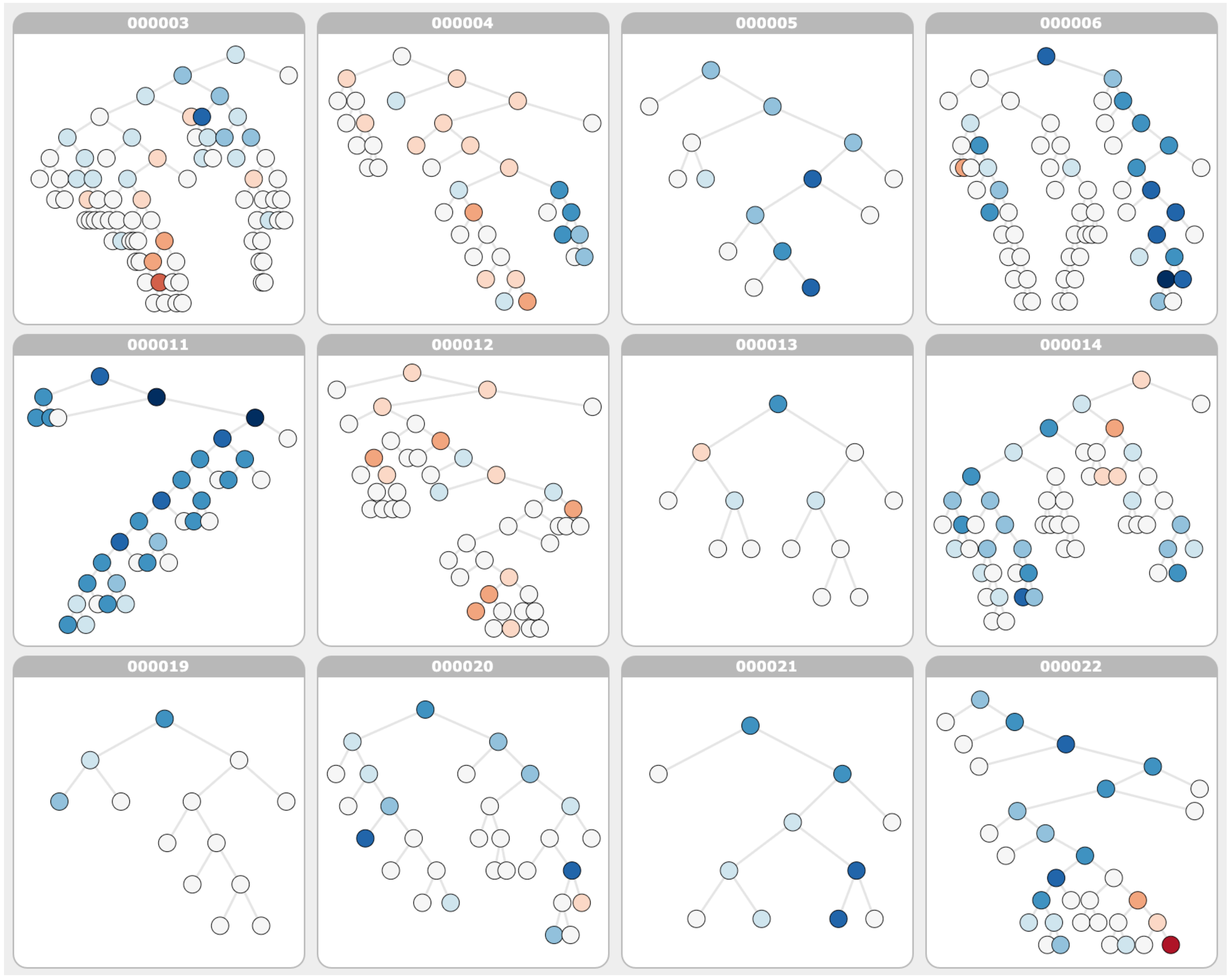
$$W \in \mathbb{R}^{5\times d}$$

Recursive Neural Network

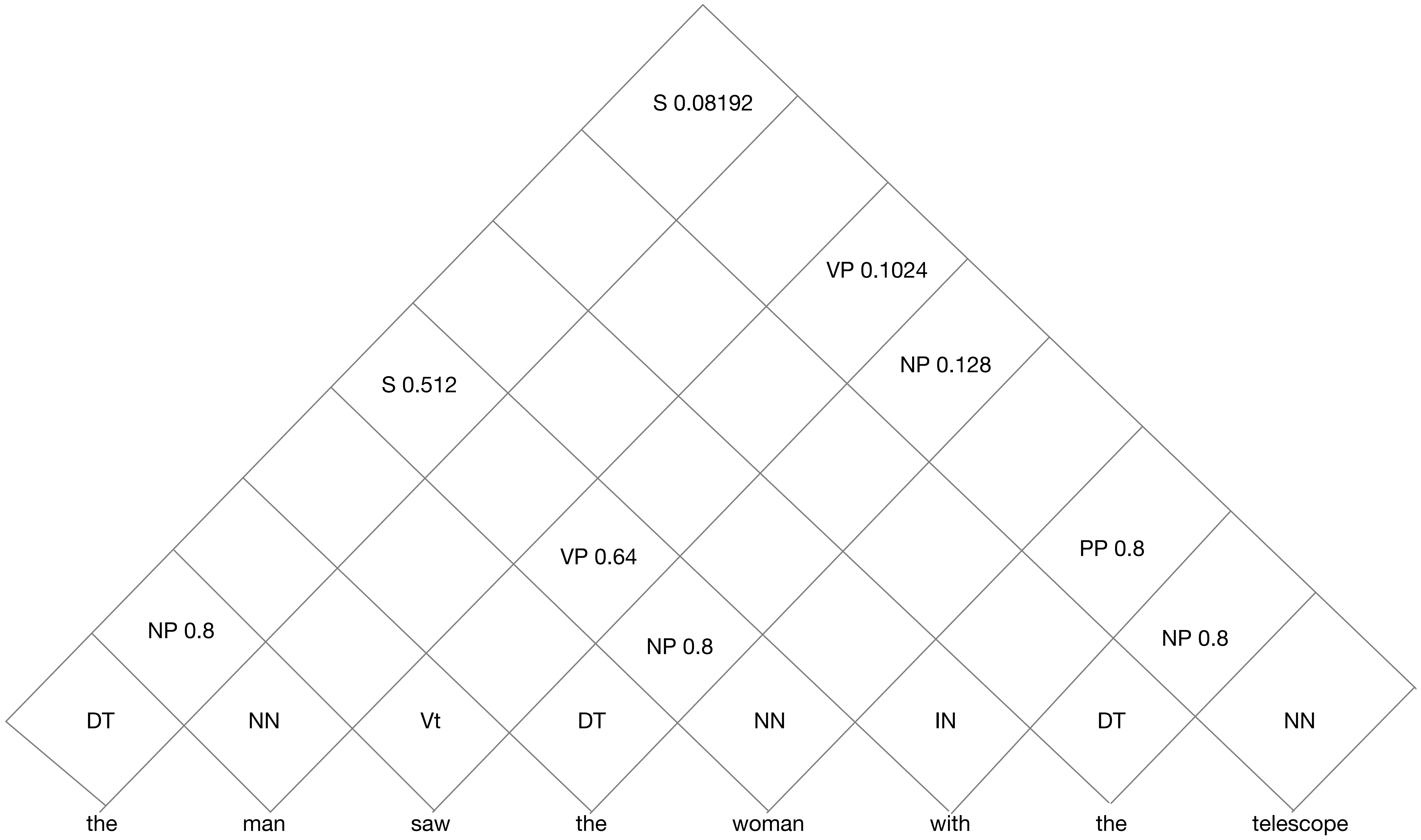
What’s bad about it?

Or, what’s good about Recurrent NN?

hard to batch, parse tree errors, difficult to pretrain (or use pretrained models) …



Parsing with PCFGs (CYK Algorithm)



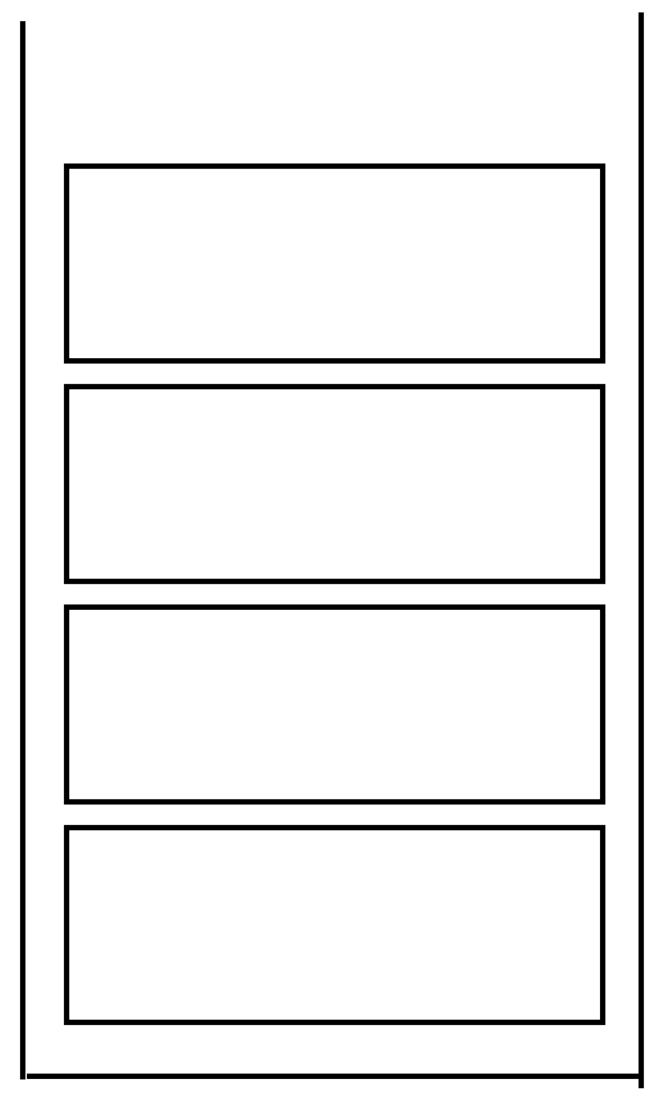
Shift-reduce Parsing

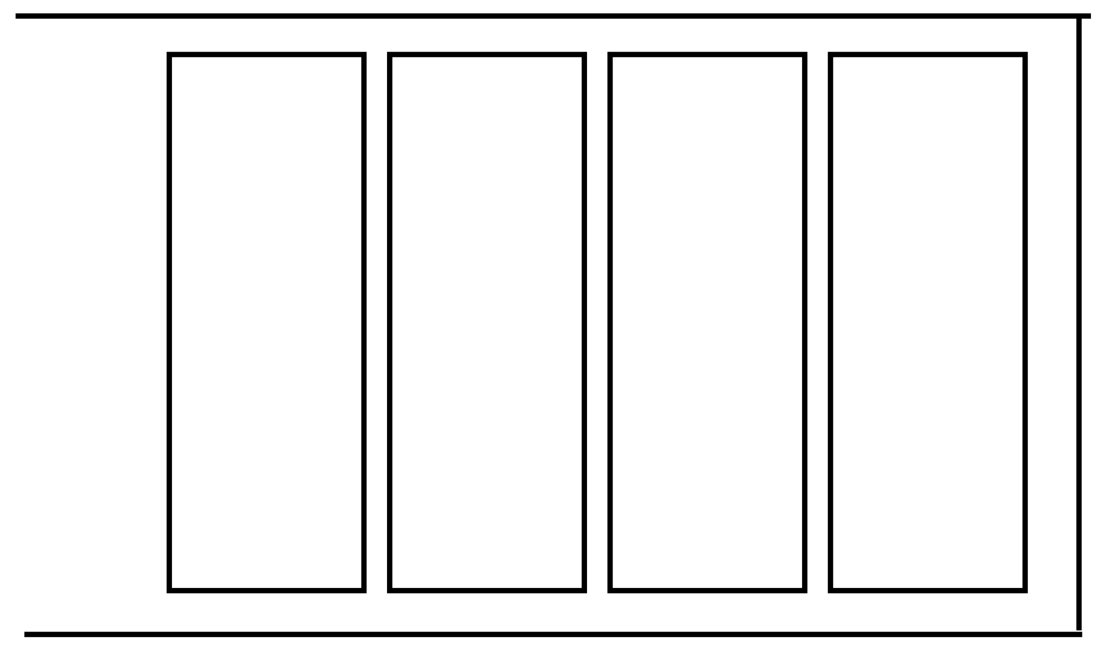
The hungry cat meows .

(S (NP The hungry cat) (VP meows)) .

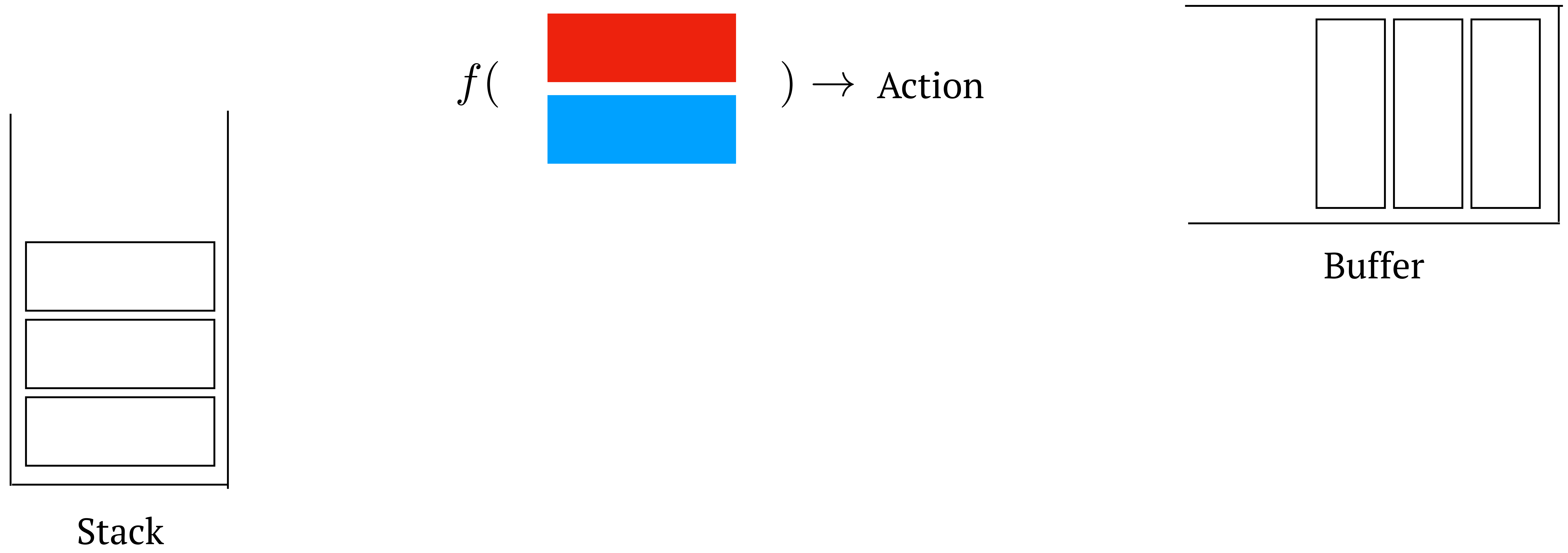
Stack	Buffer	Action

Shift-reduce Parsing


Stack


Buffer

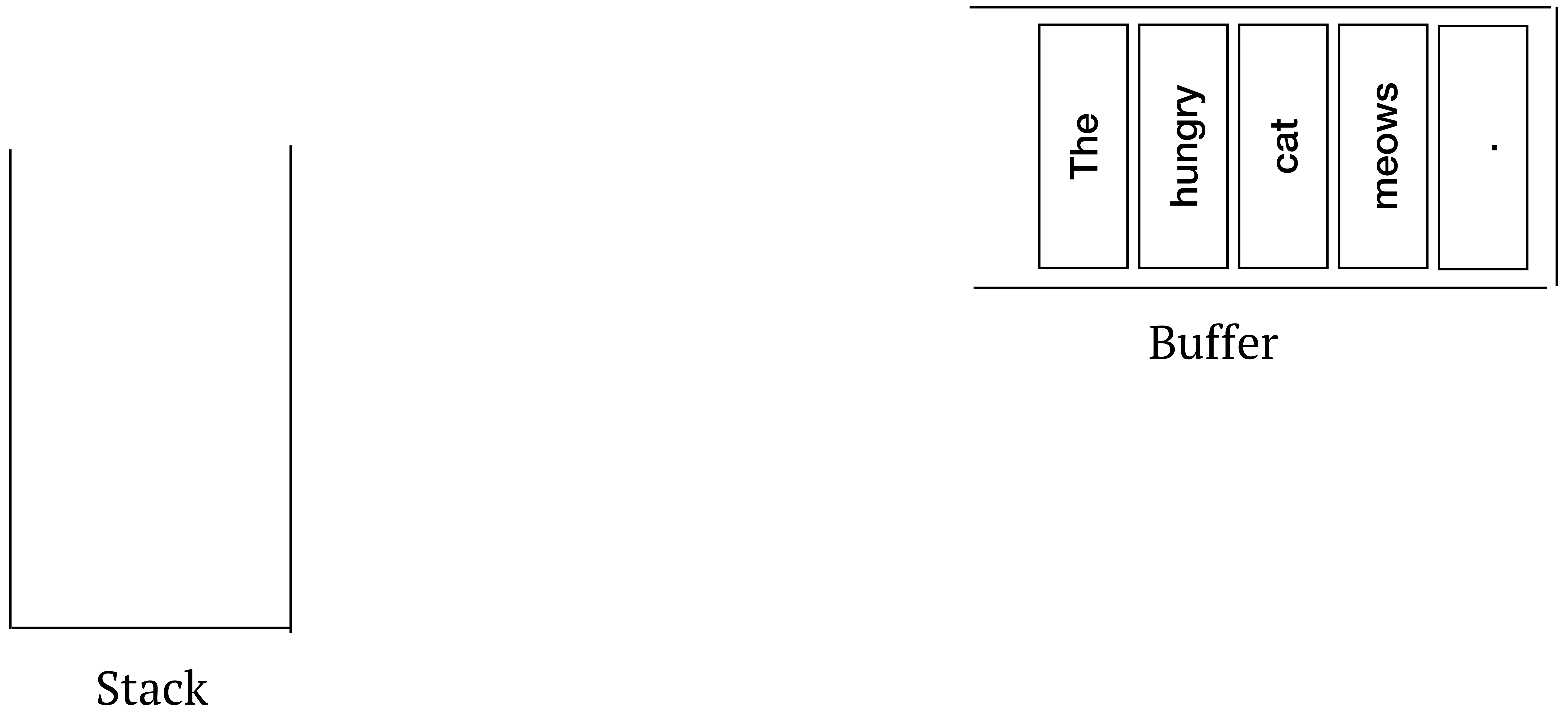
Shift-reduce Parsing



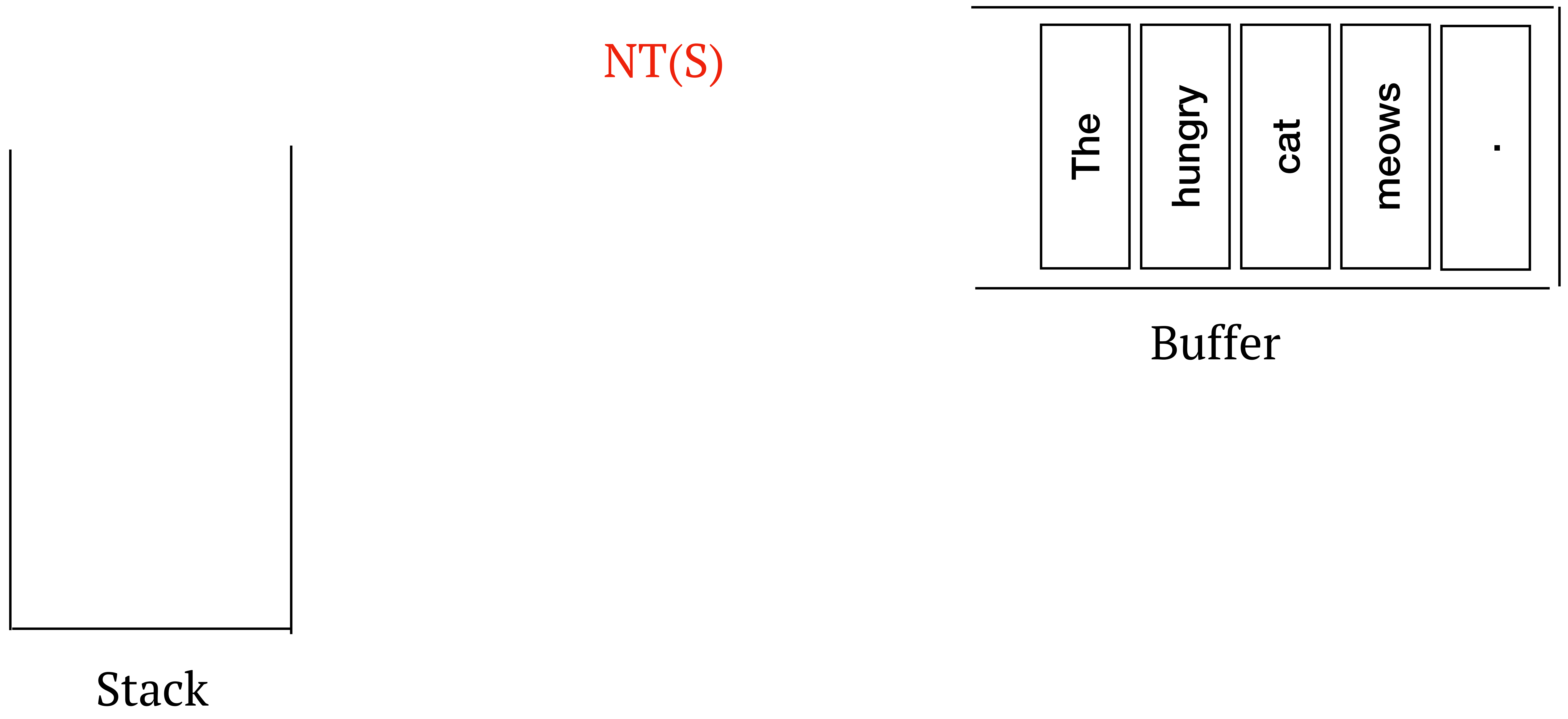
Shift-reduce Parsing

Action		
	NT(S)	
	NT(NP)	push an open non-terminal onto the stack
	NT(VP)	
	SHIFT	shift a symbol from the buffer onto the stack
	REDUCE	repeatedly pops completed subtrees or terminal symbols from the stack until an open nonterminal is encountered, and then this open NT is popped and used as the label of a new constituent that has the popped subtrees as its children. This new completed constituent is pushed onto the stack as a single composite item.

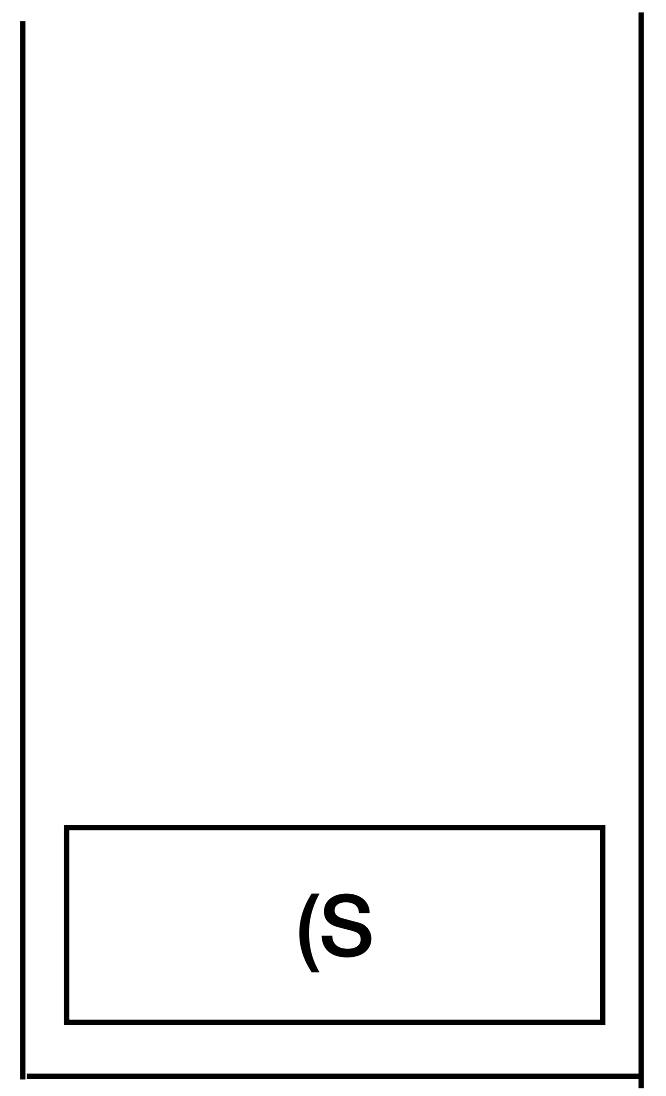
Shift-reduce Parsing



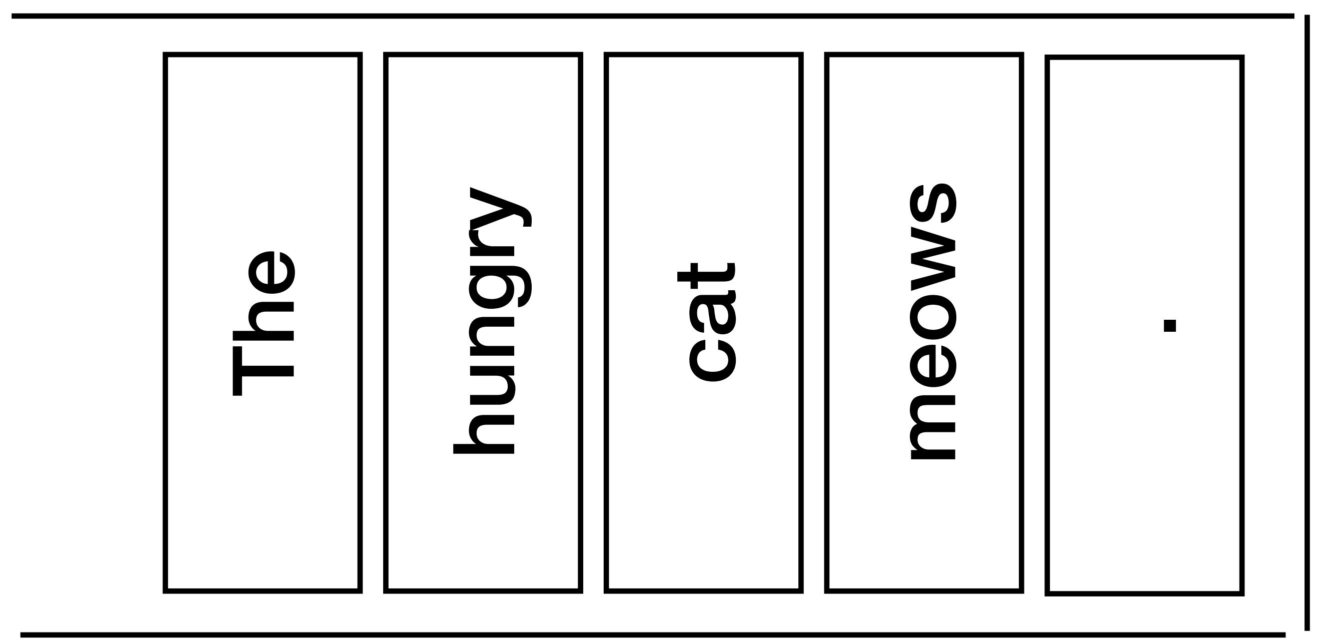
Shift-reduce Parsing



Shift-reduce Parsing

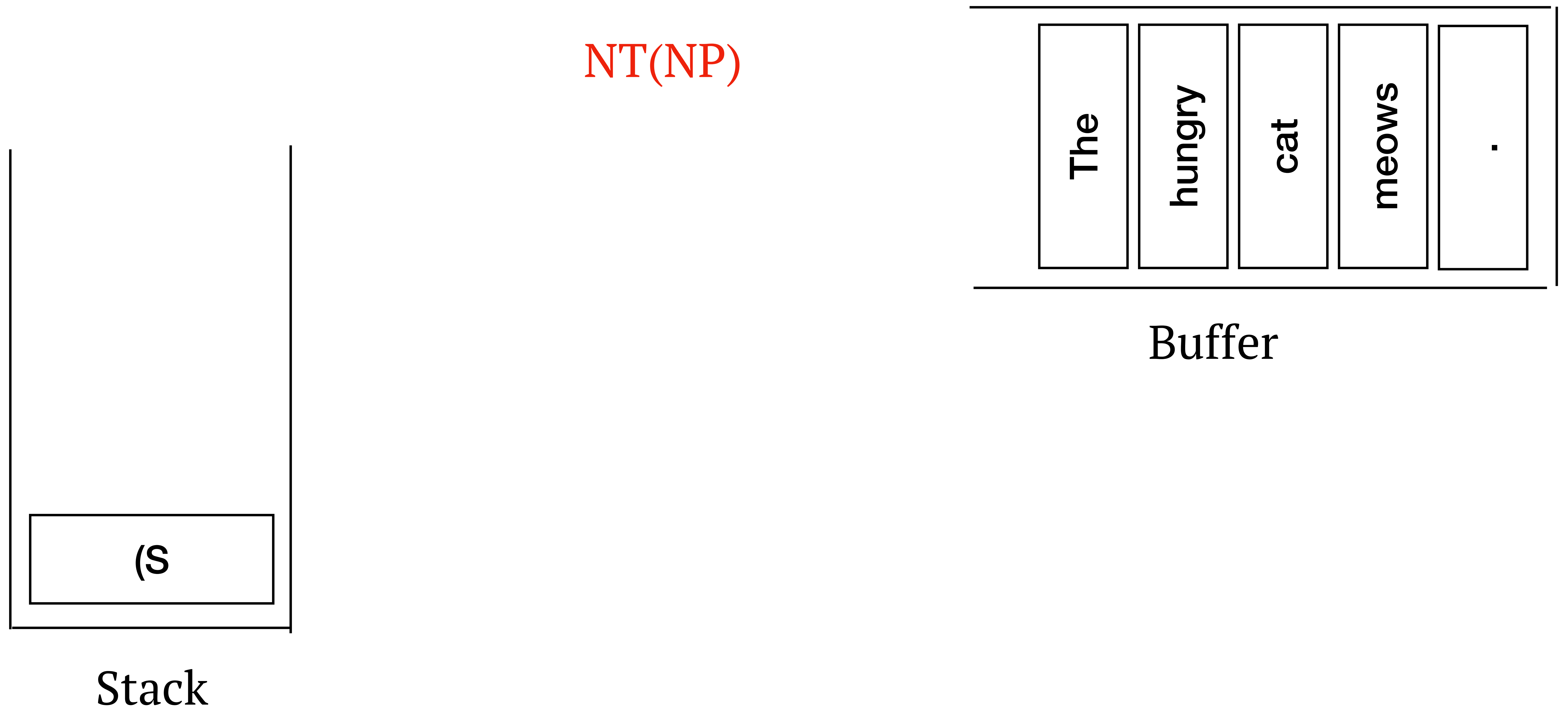


Stack

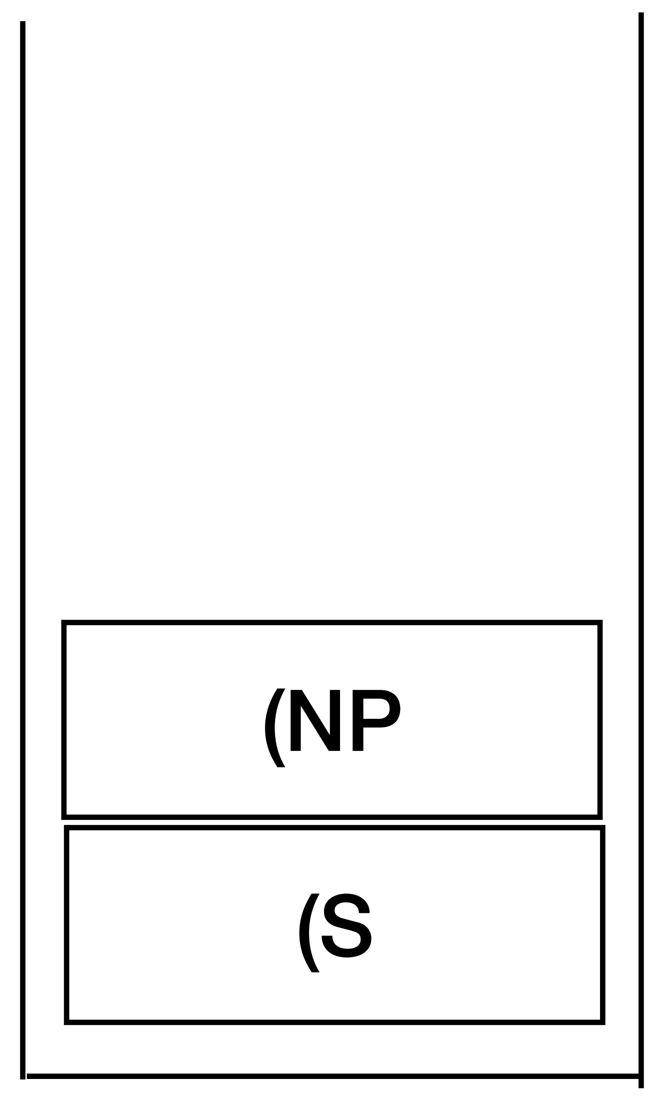


Buffer

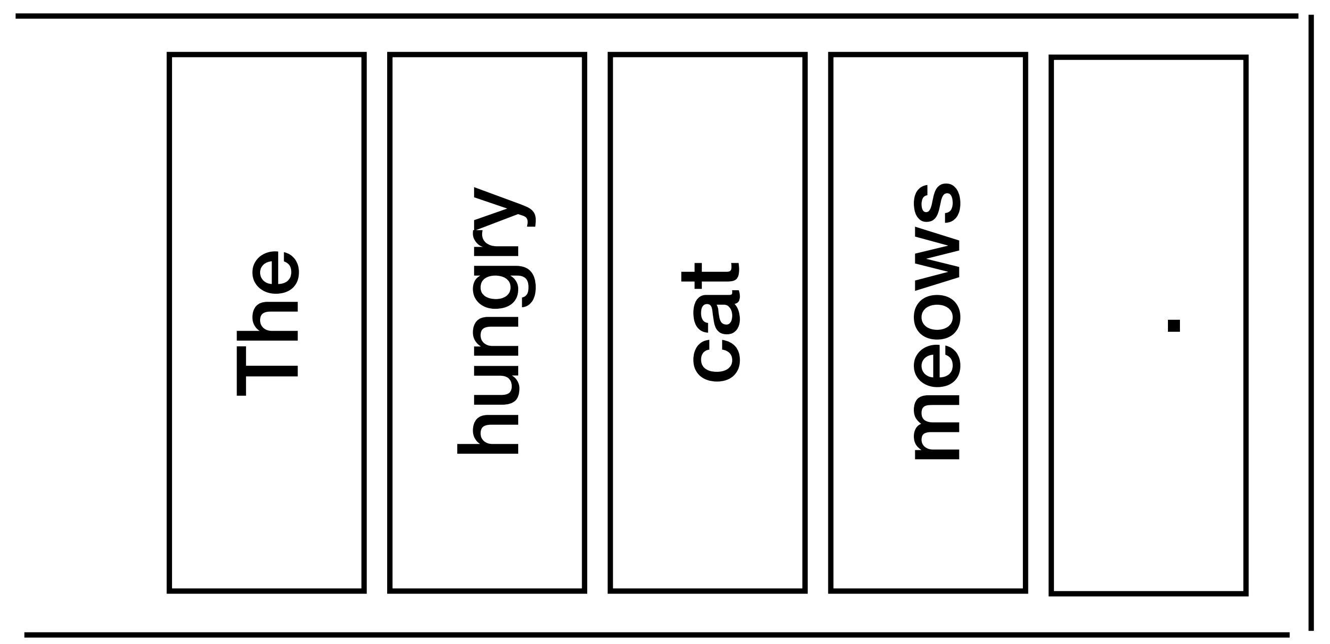
Shift-reduce Parsing



Shift-reduce Parsing

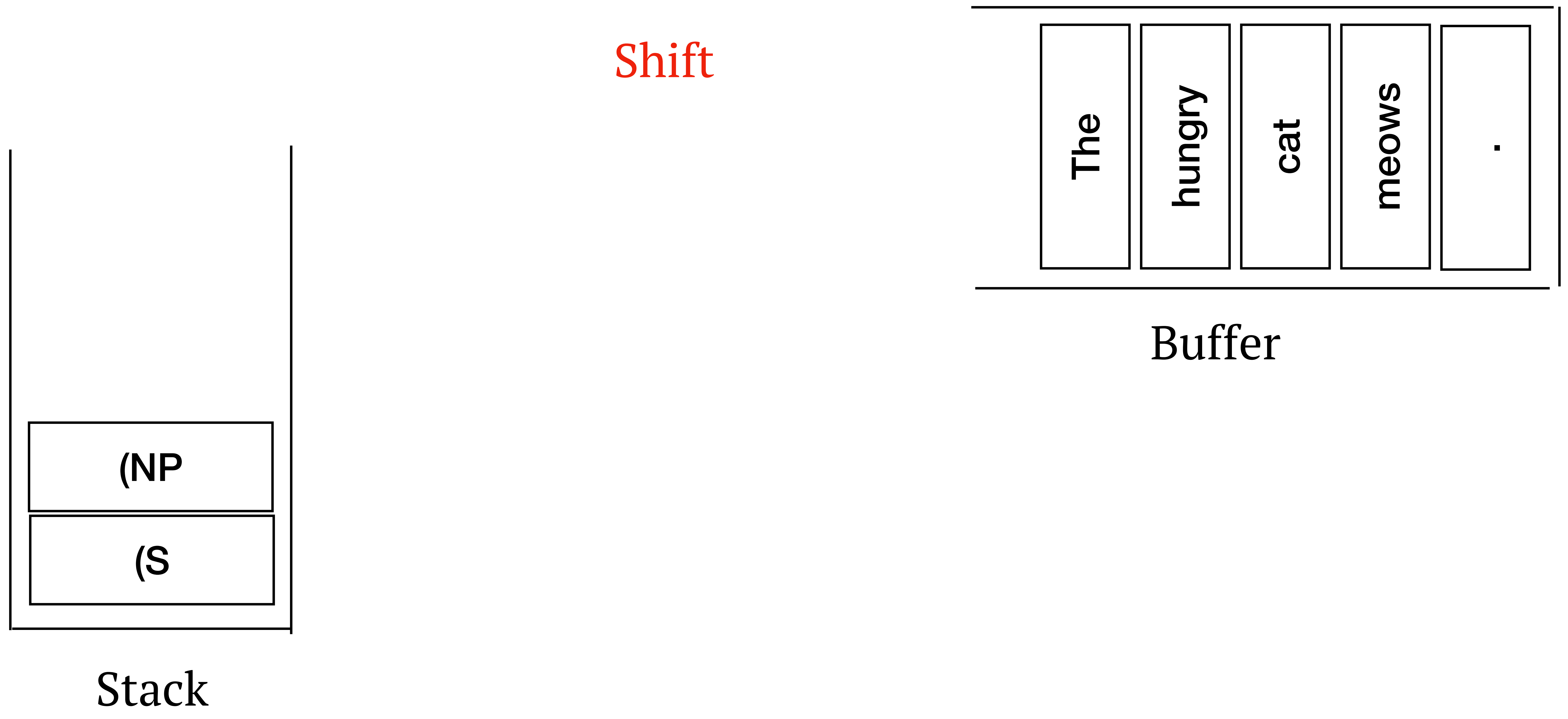


Stack

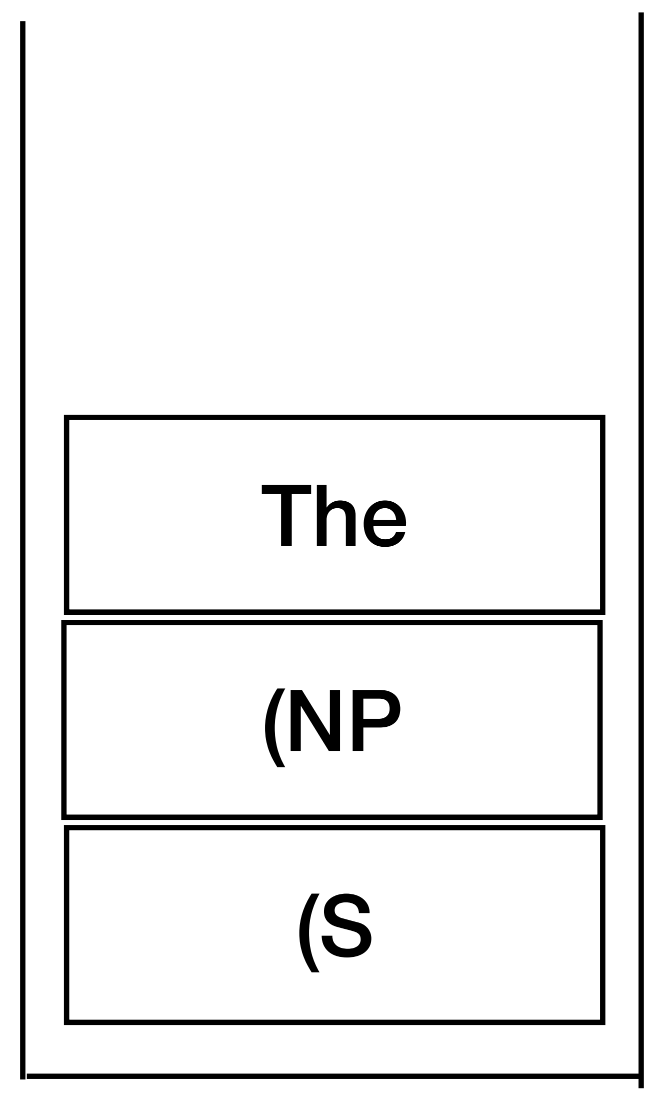


Buffer

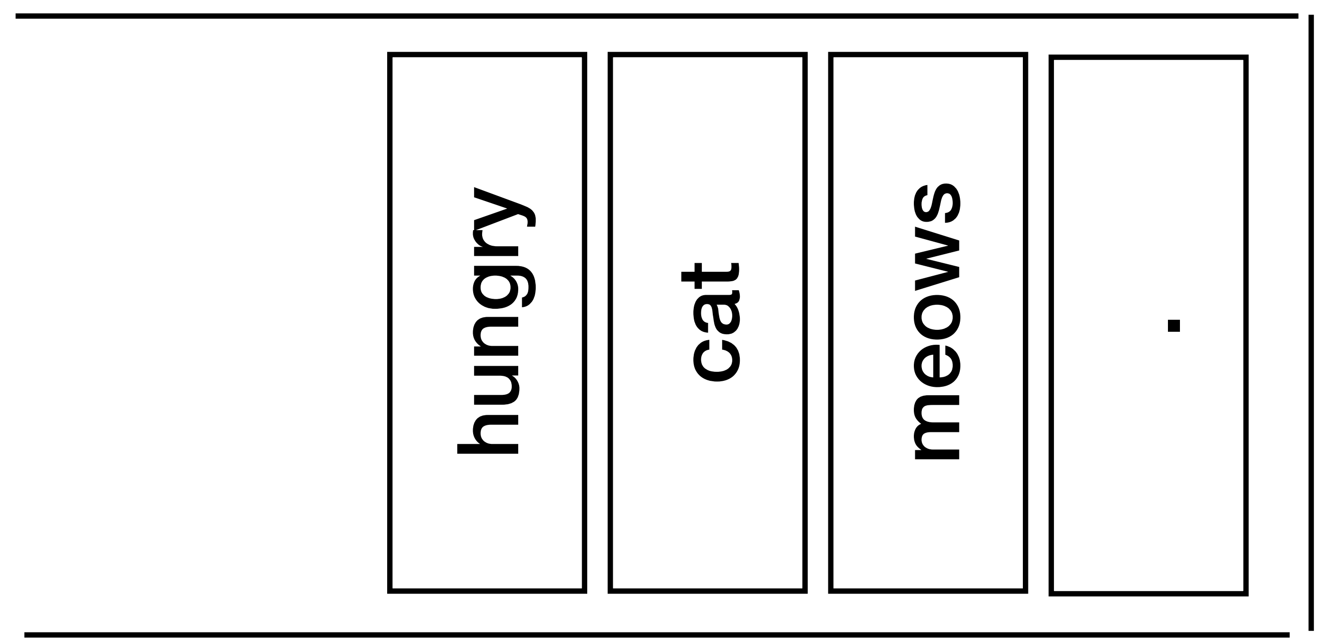
Shift-reduce Parsing



Shift-reduce Parsing

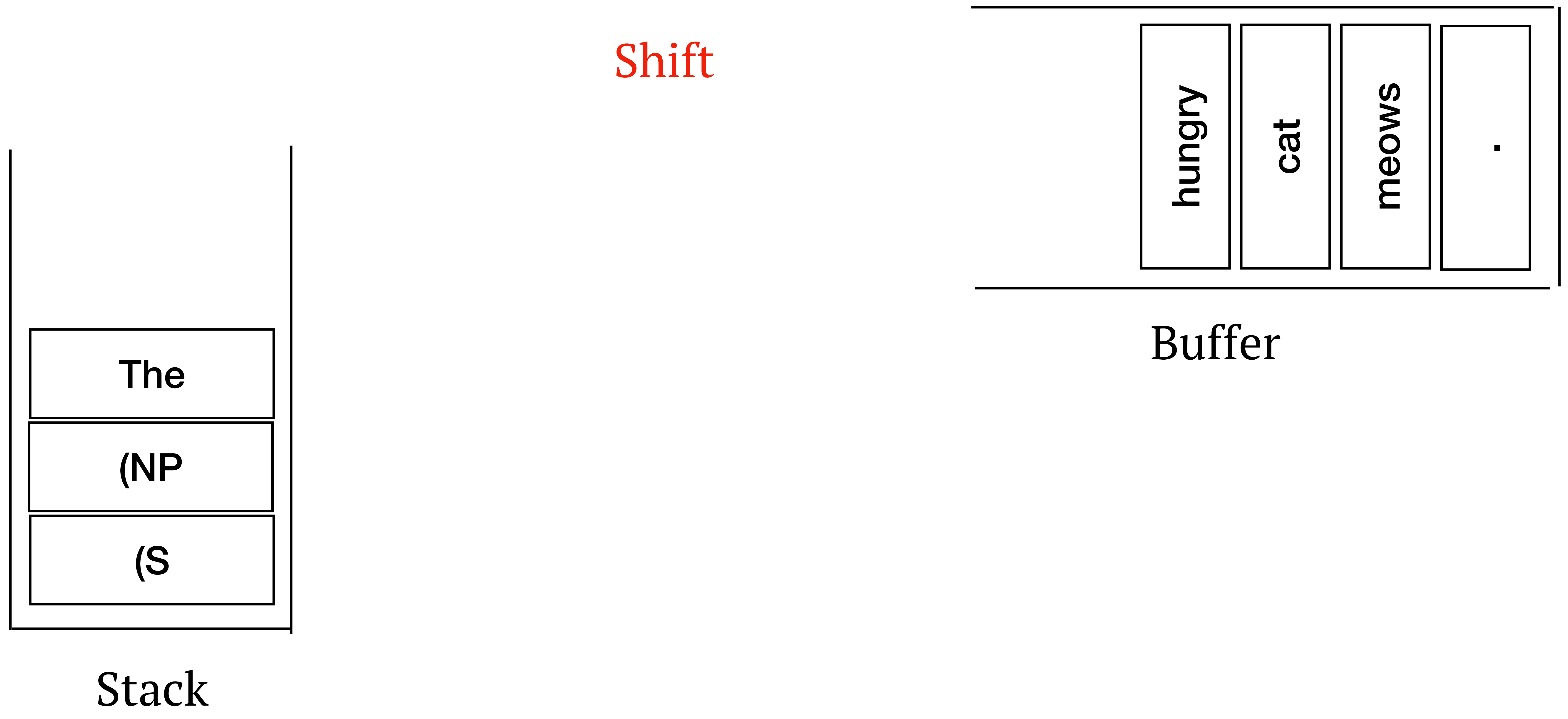


Stack

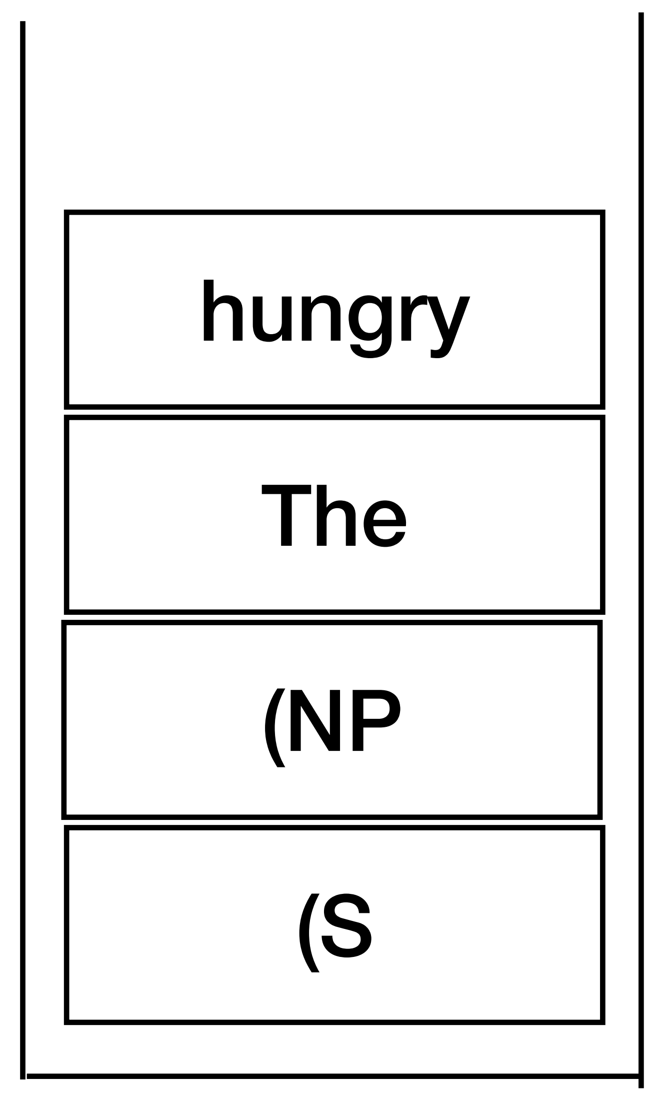


Buffer

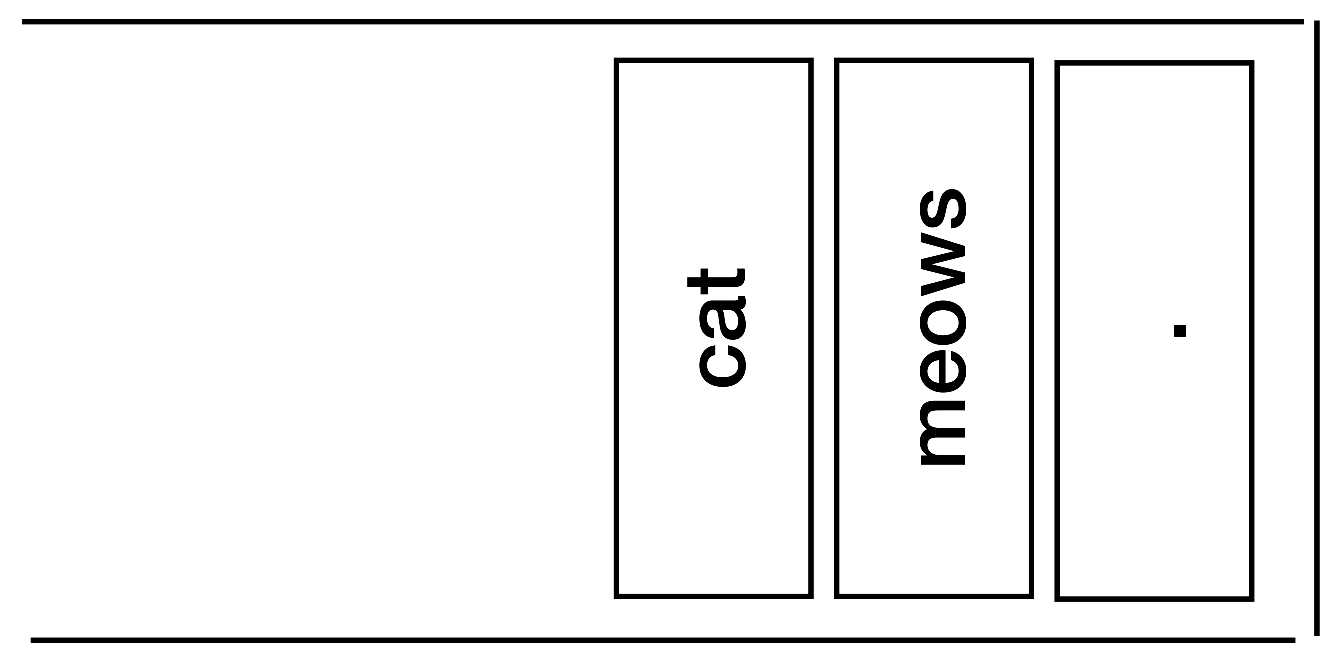
Shift-reduce Parsing



Shift-reduce Parsing



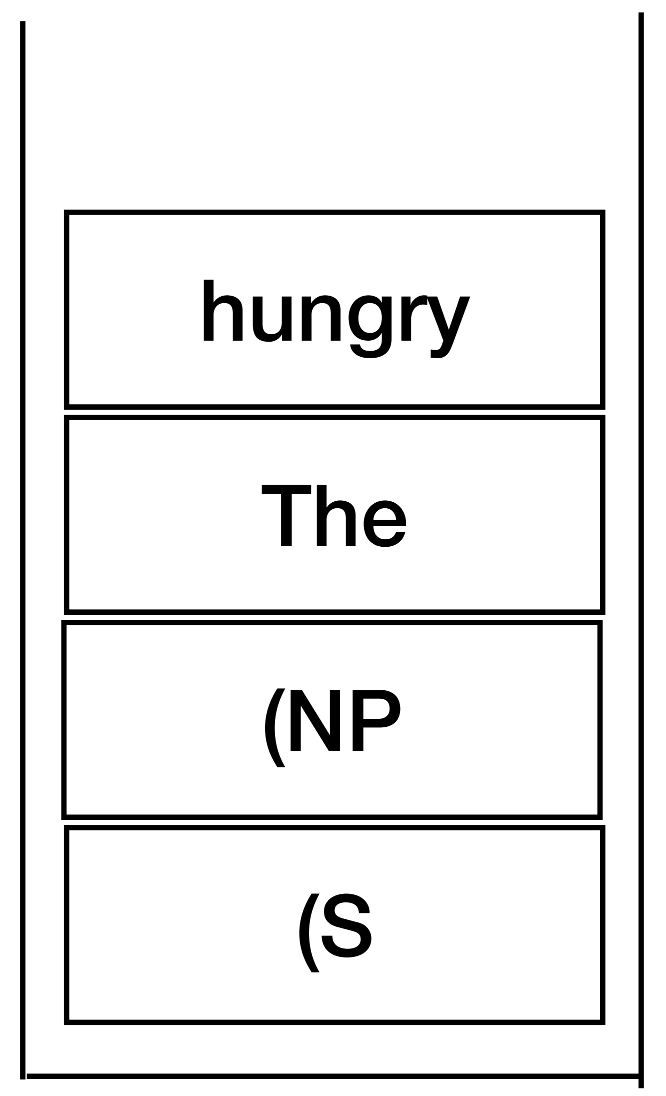
Stack



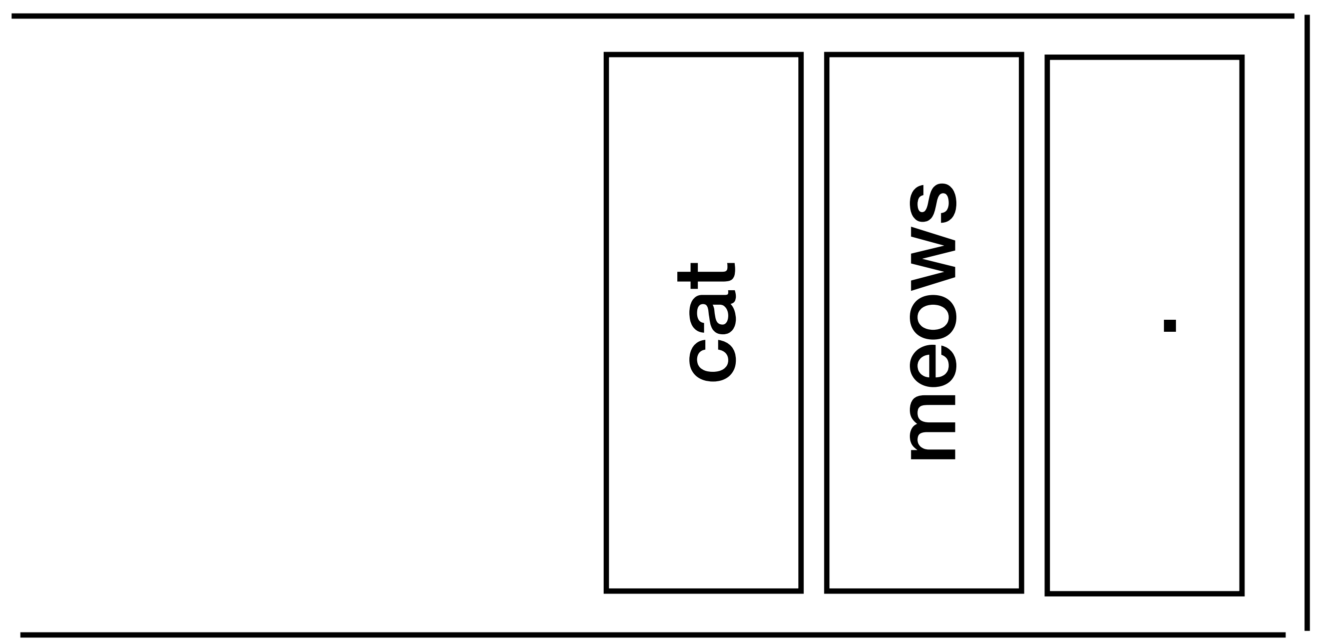
Buffer

Shift-reduce Parsing

Shift

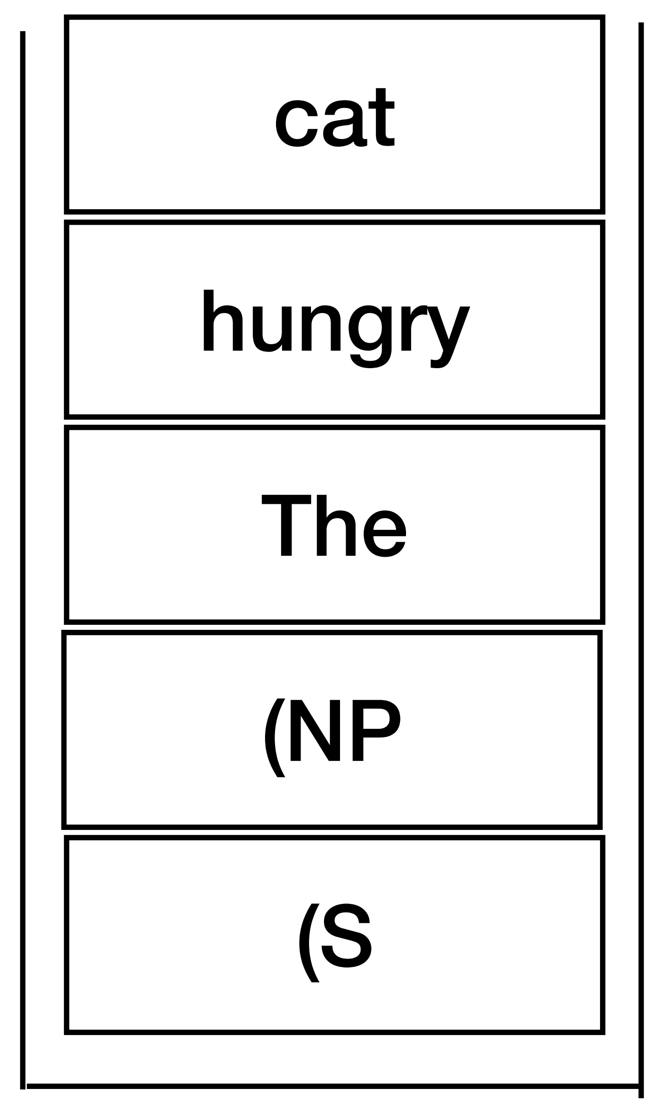


Stack

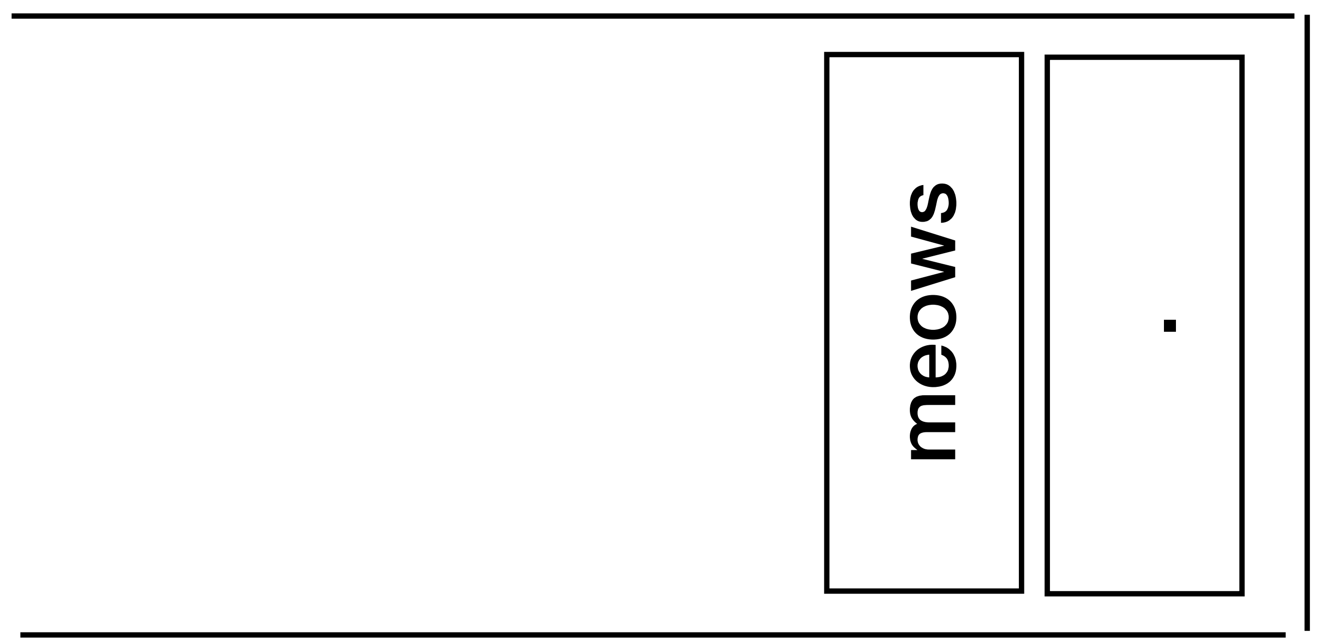


Buffer

Shift-reduce Parsing

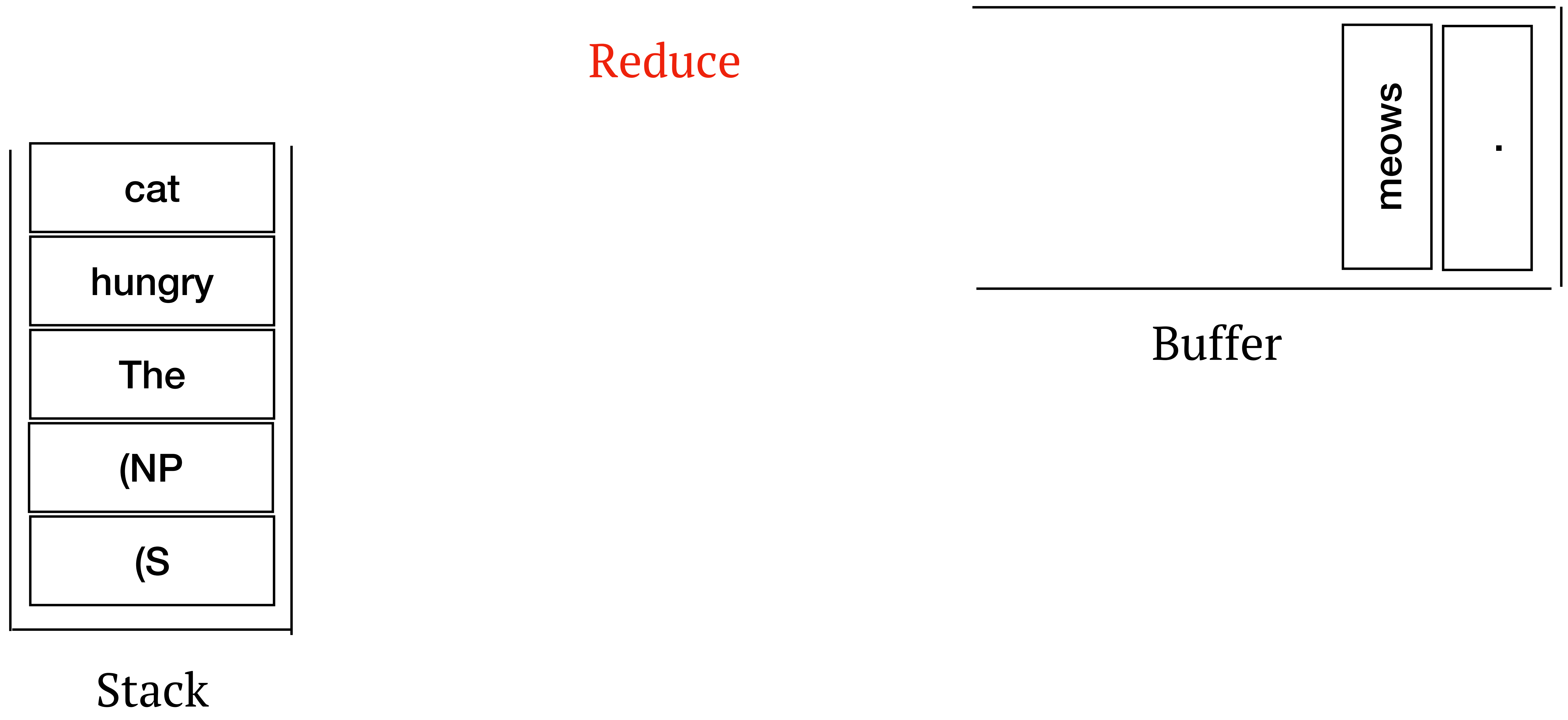


Stack

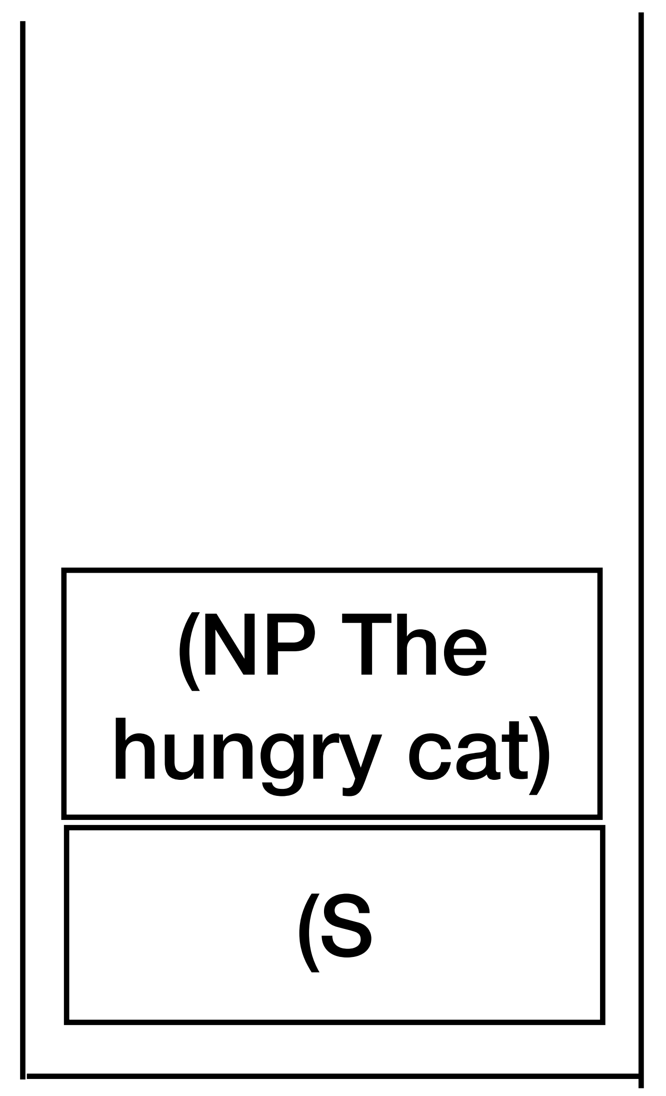


Buffer

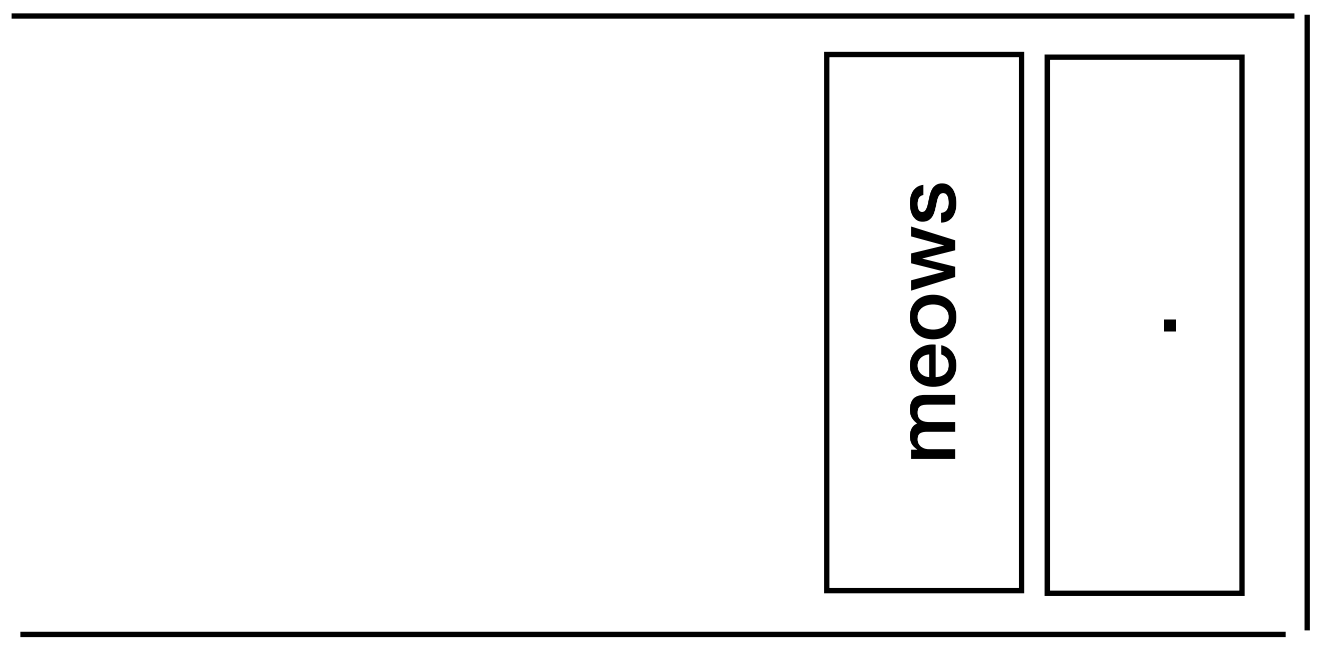
Shift-reduce Parsing



Shift-reduce Parsing

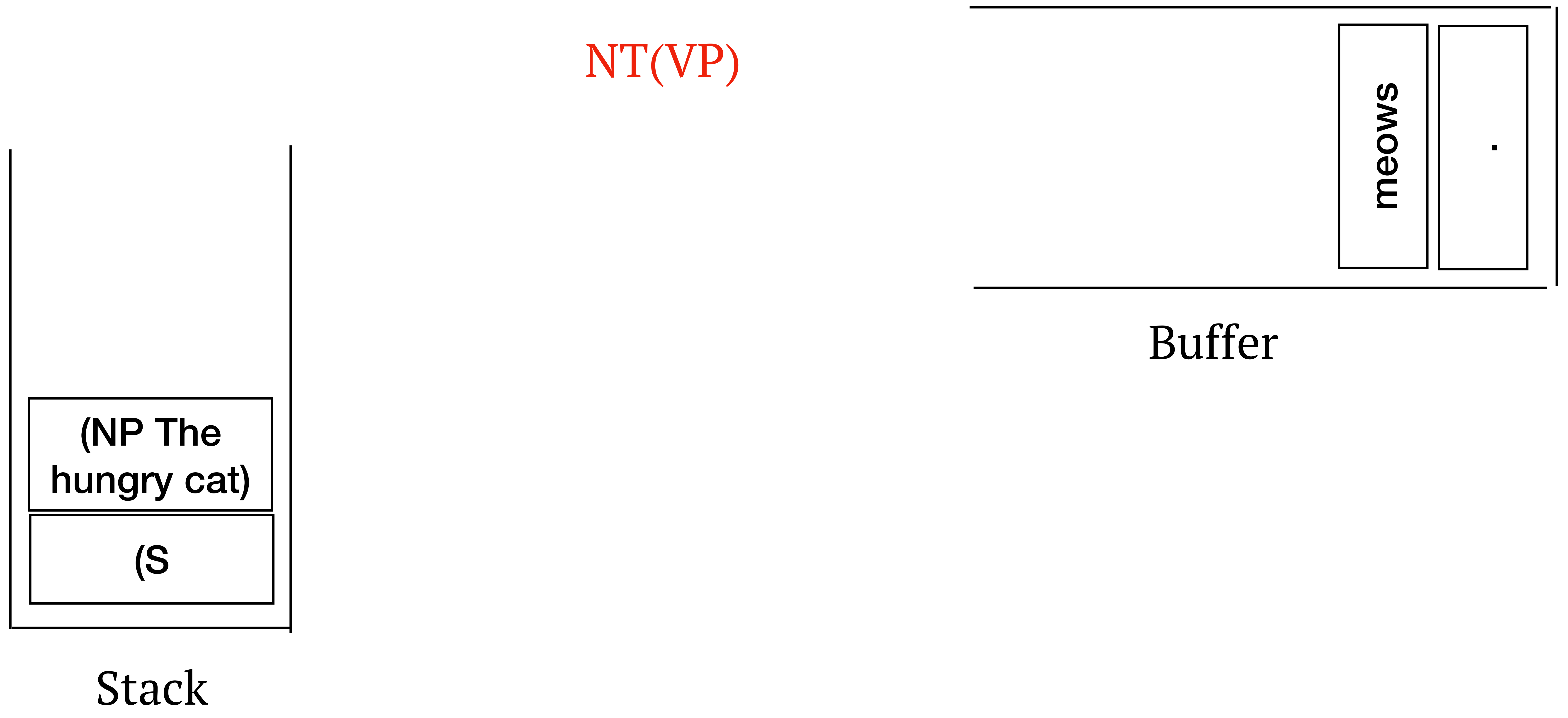


Stack

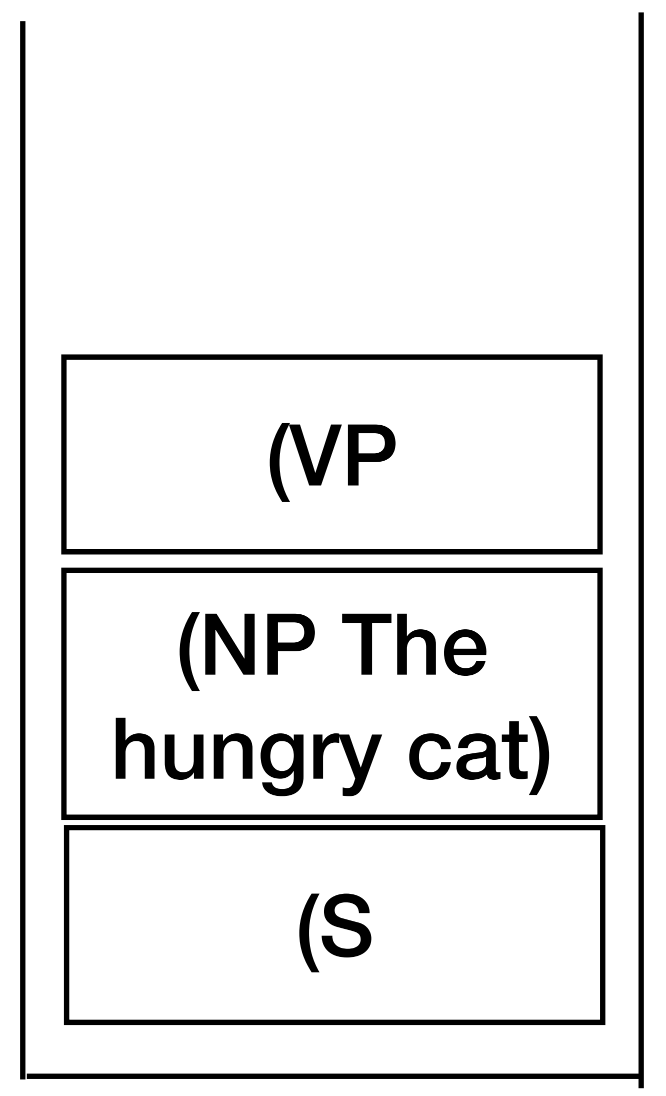


Buffer

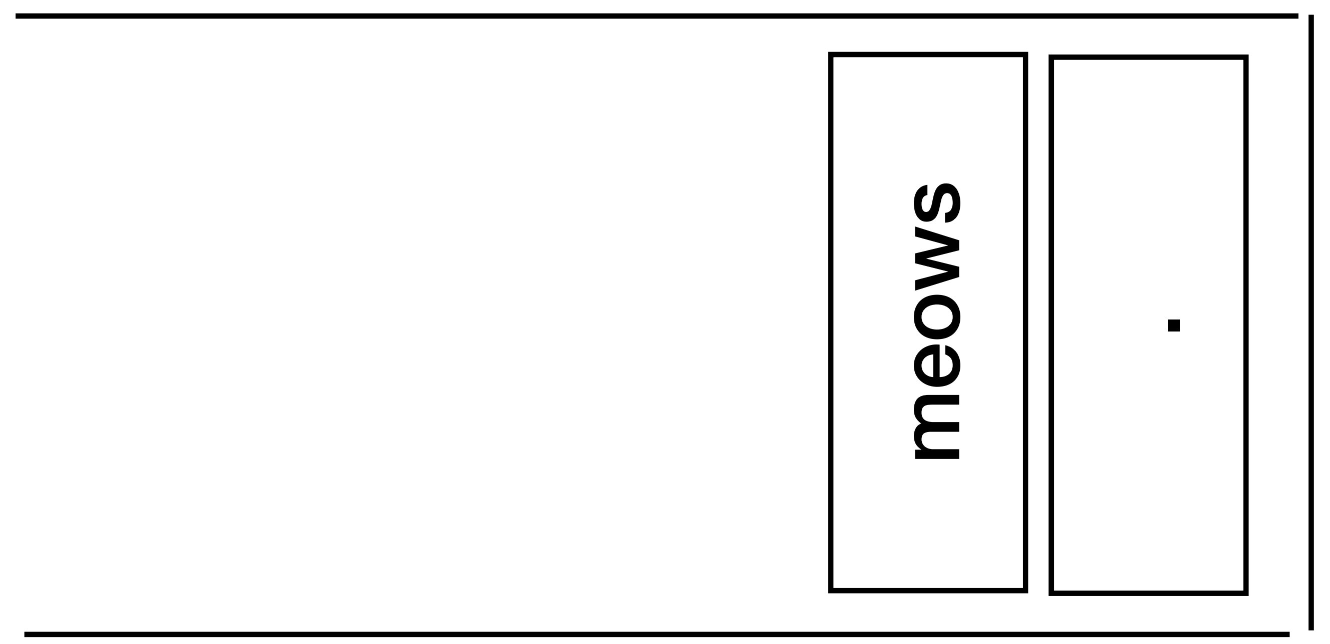
Shift-reduce Parsing



Shift-reduce Parsing

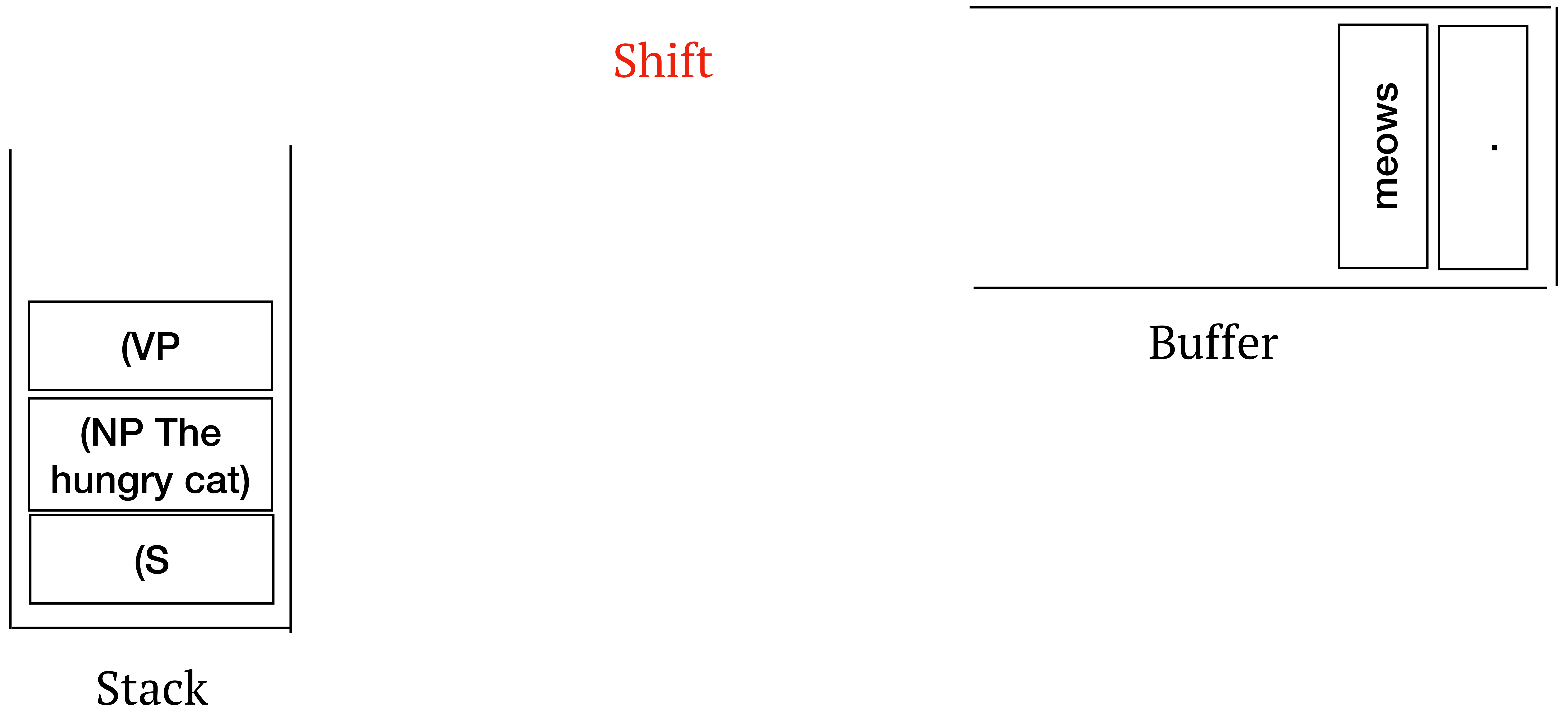


Stack

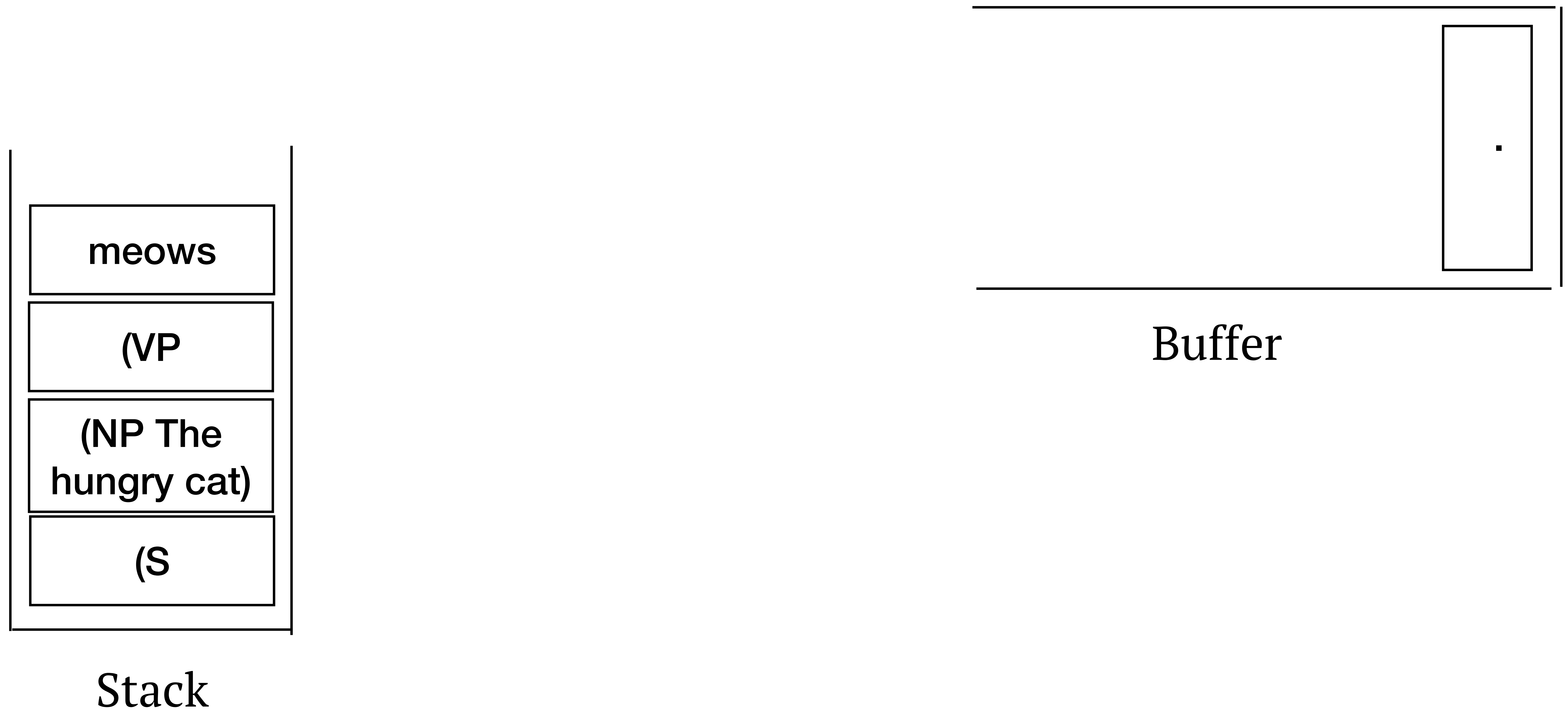


Buffer

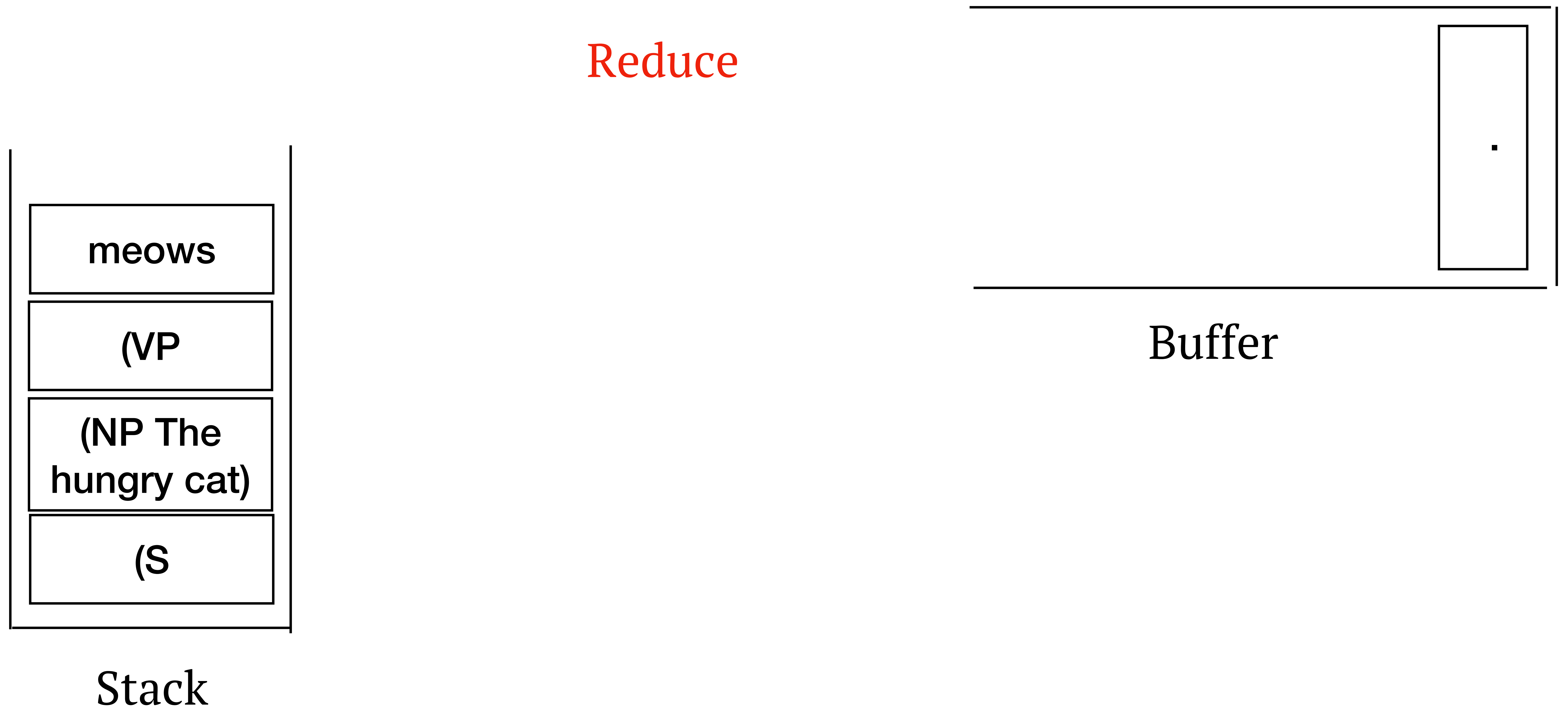
Shift-reduce Parsing



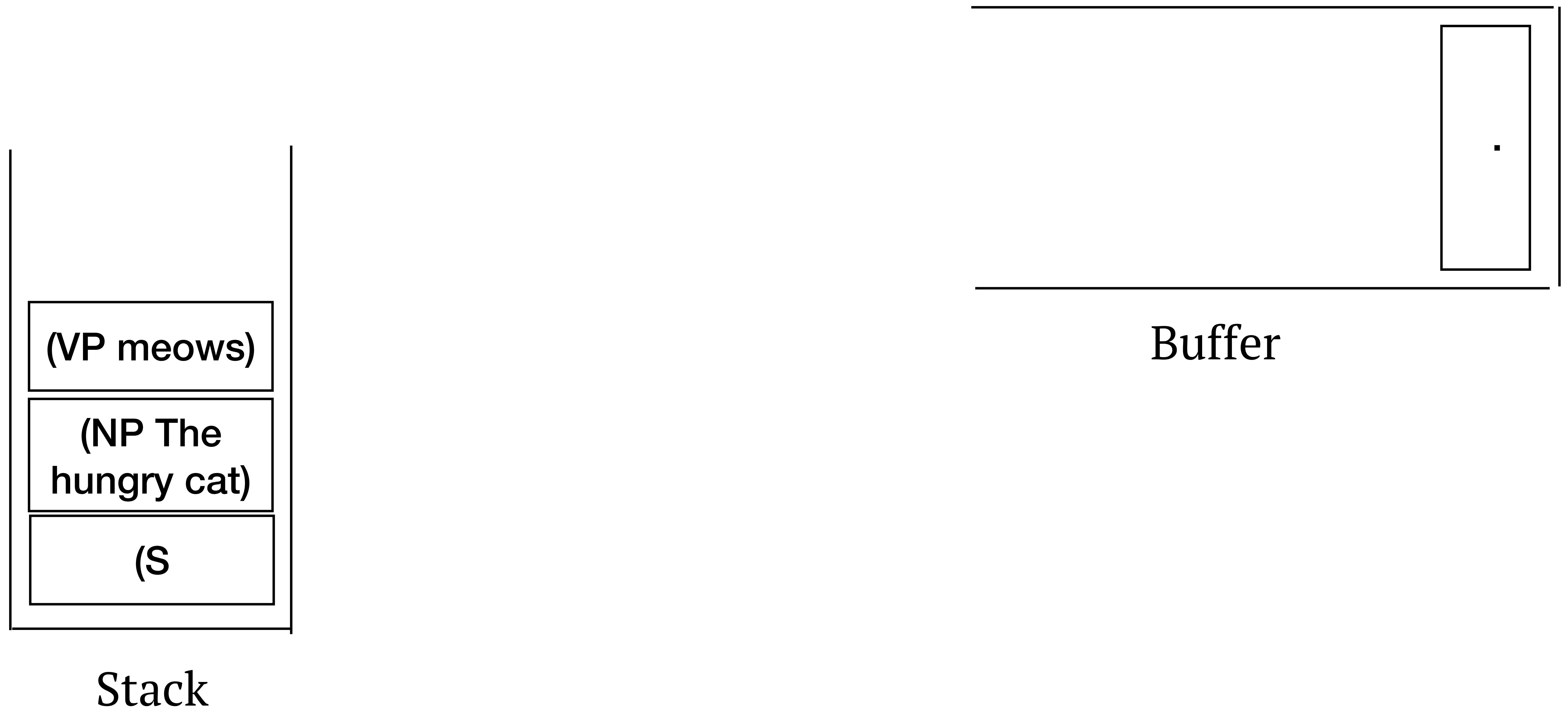
Shift-reduce Parsing



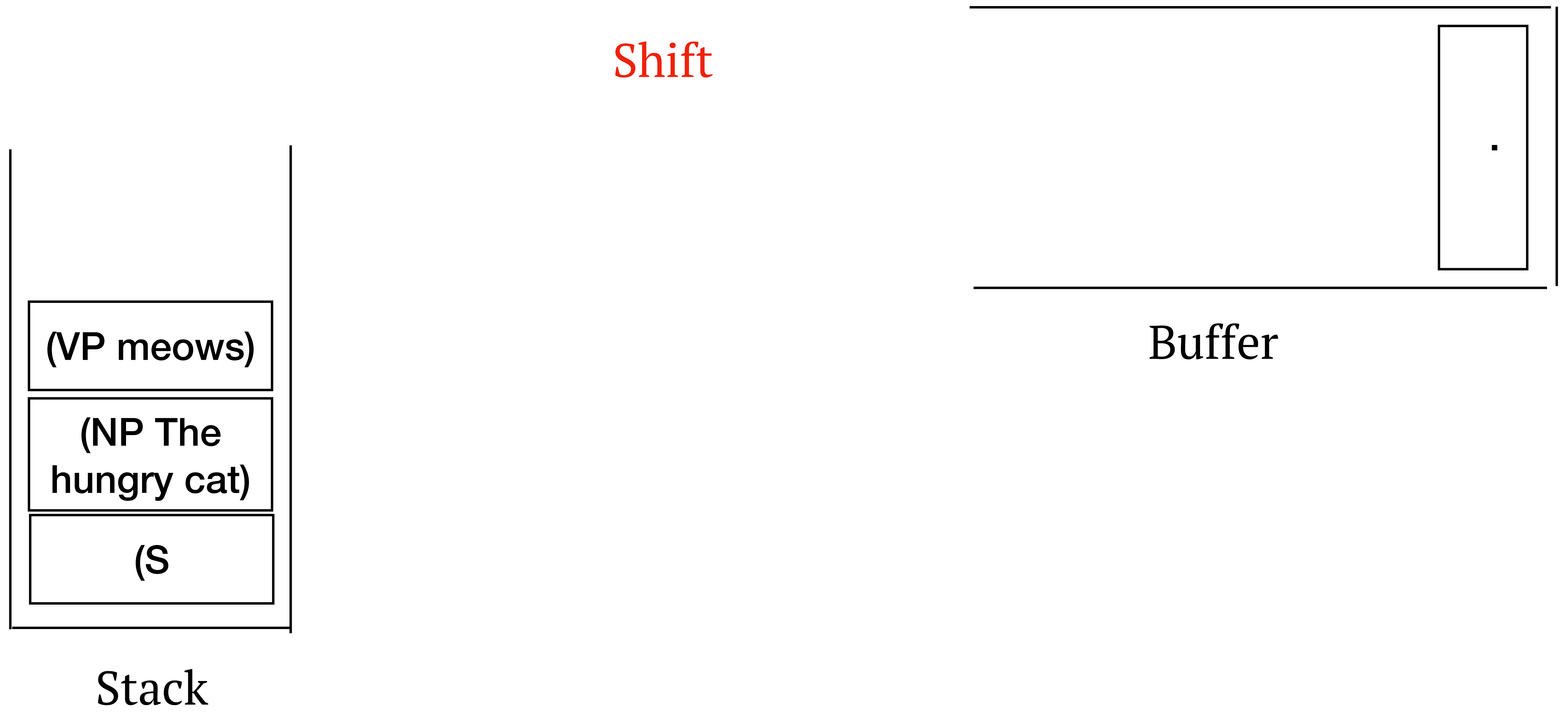
Shift-reduce Parsing



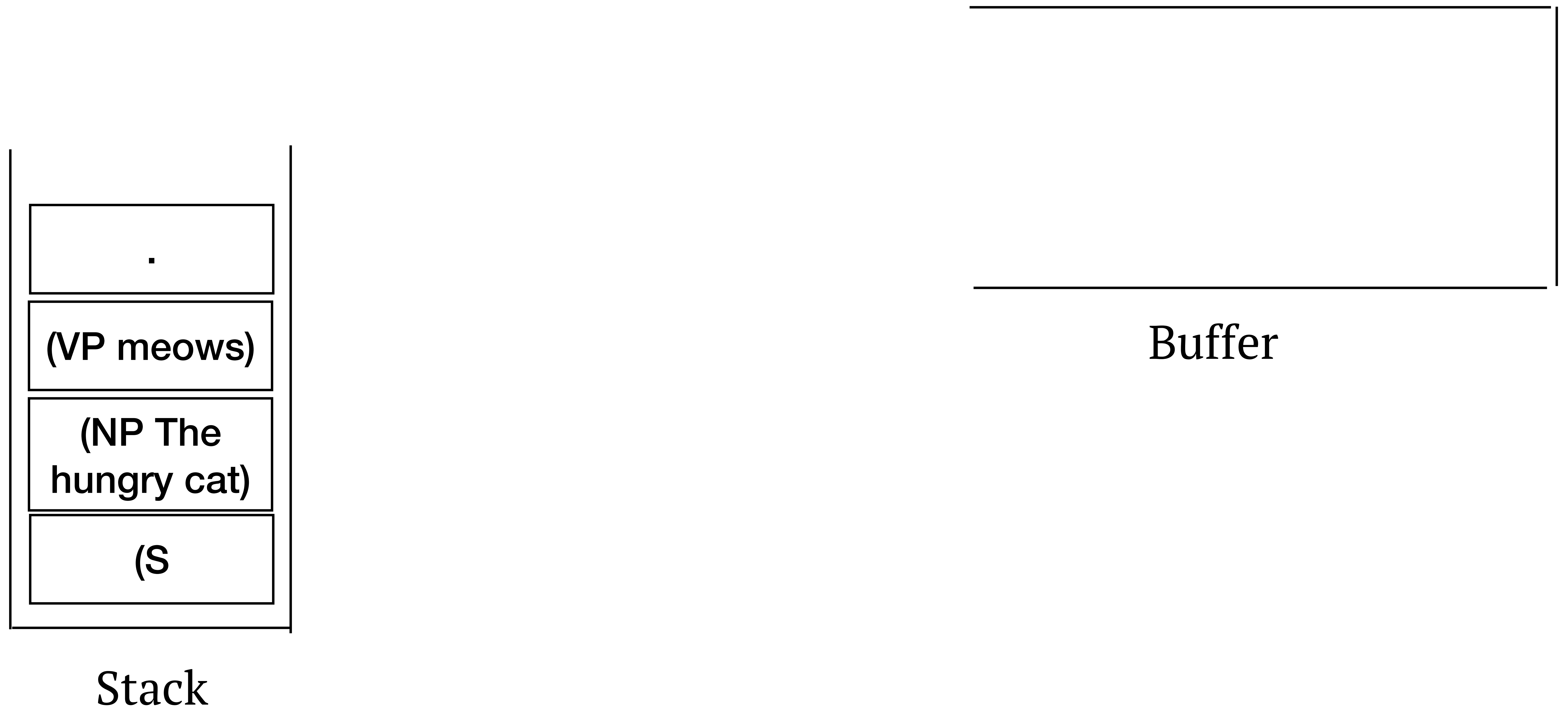
Shift-reduce Parsing



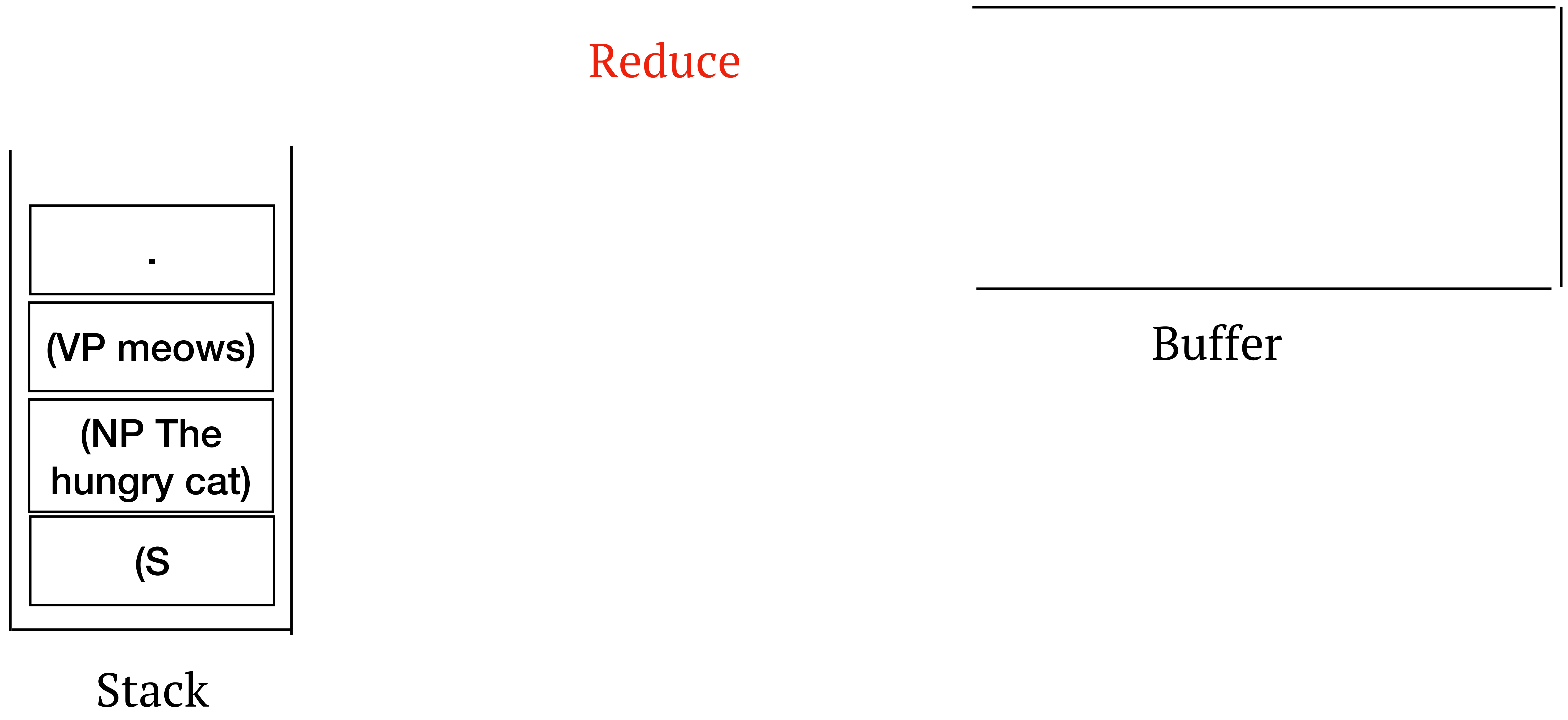
Shift-reduce Parsing



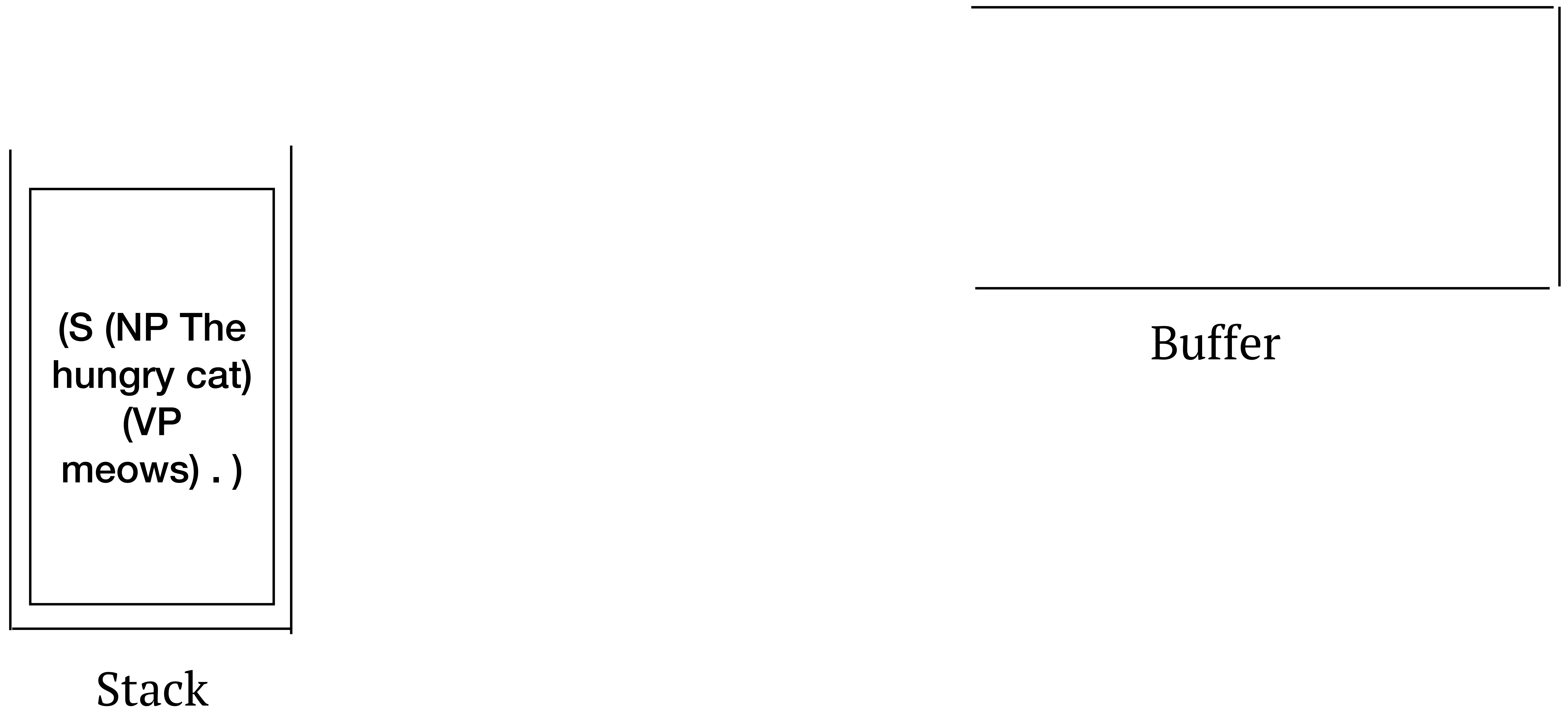
Shift-reduce Parsing



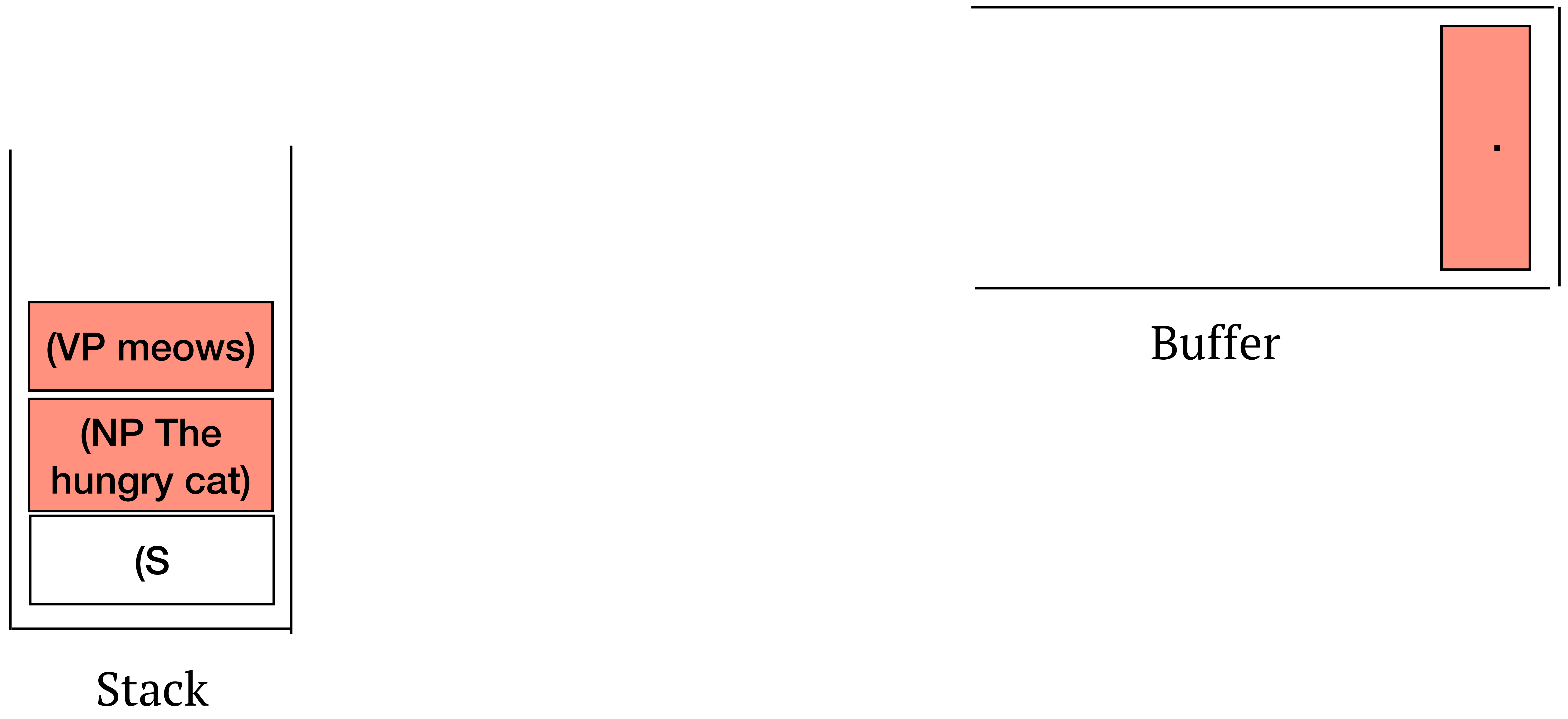
Shift-reduce Parsing



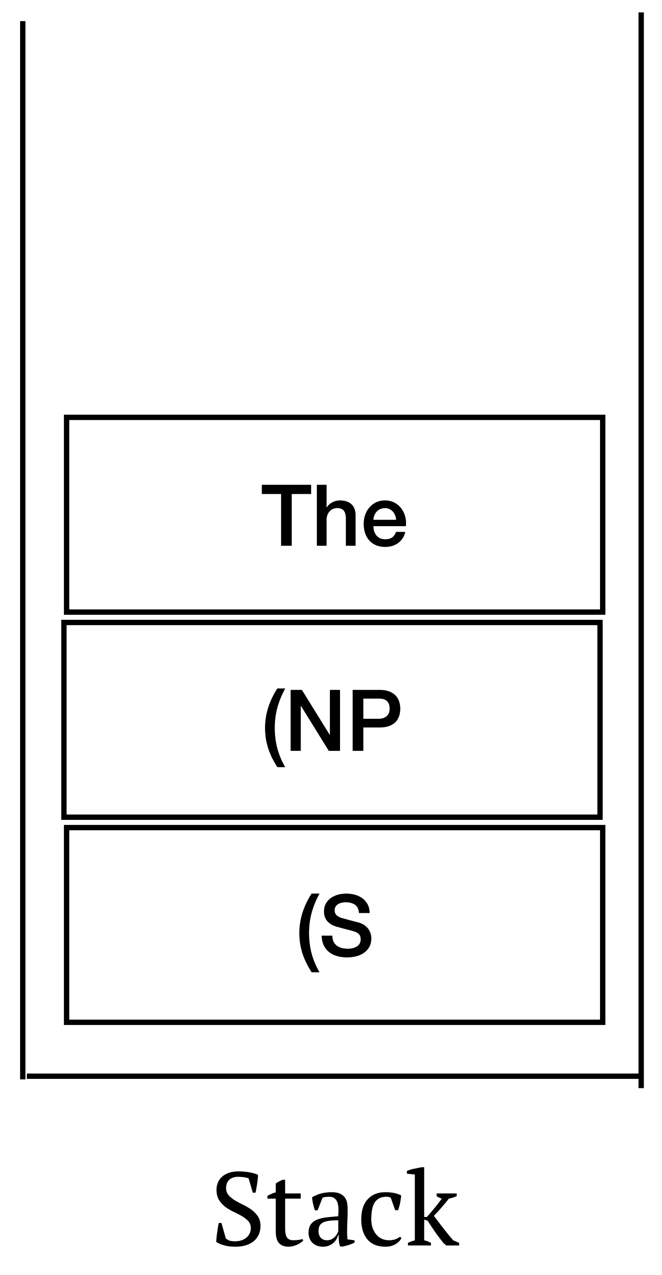
Shift-reduce Parsing

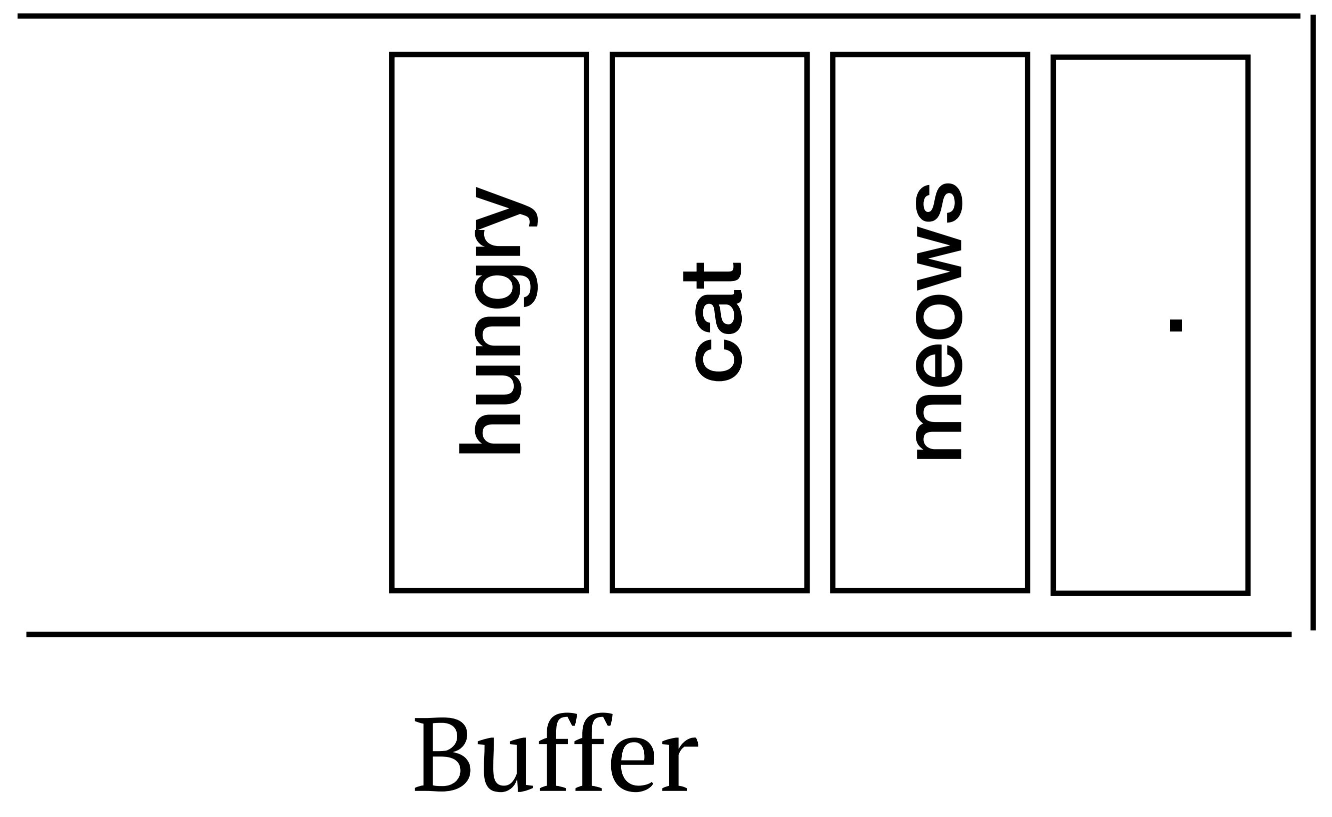


How to make decisions?



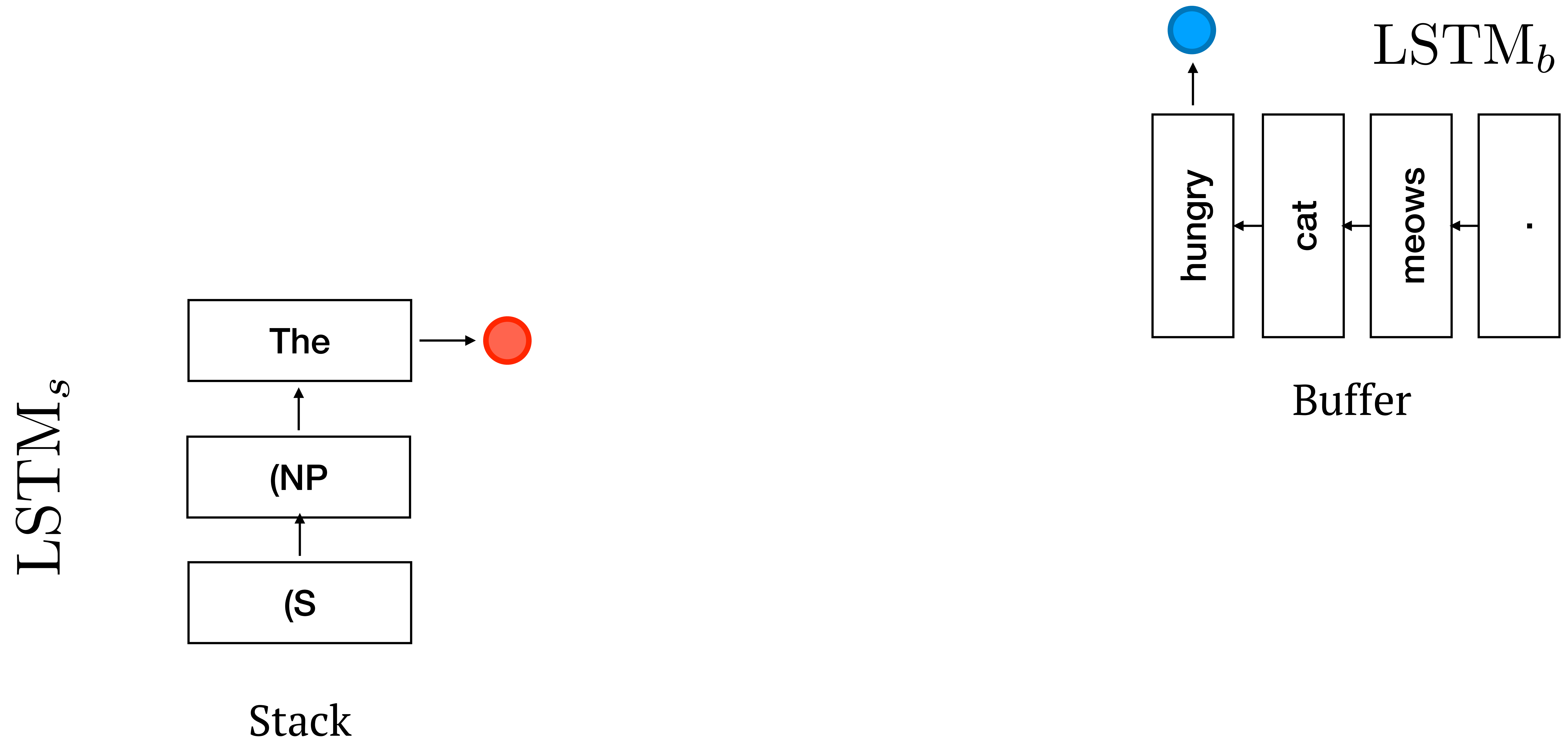
Stack LSTMs





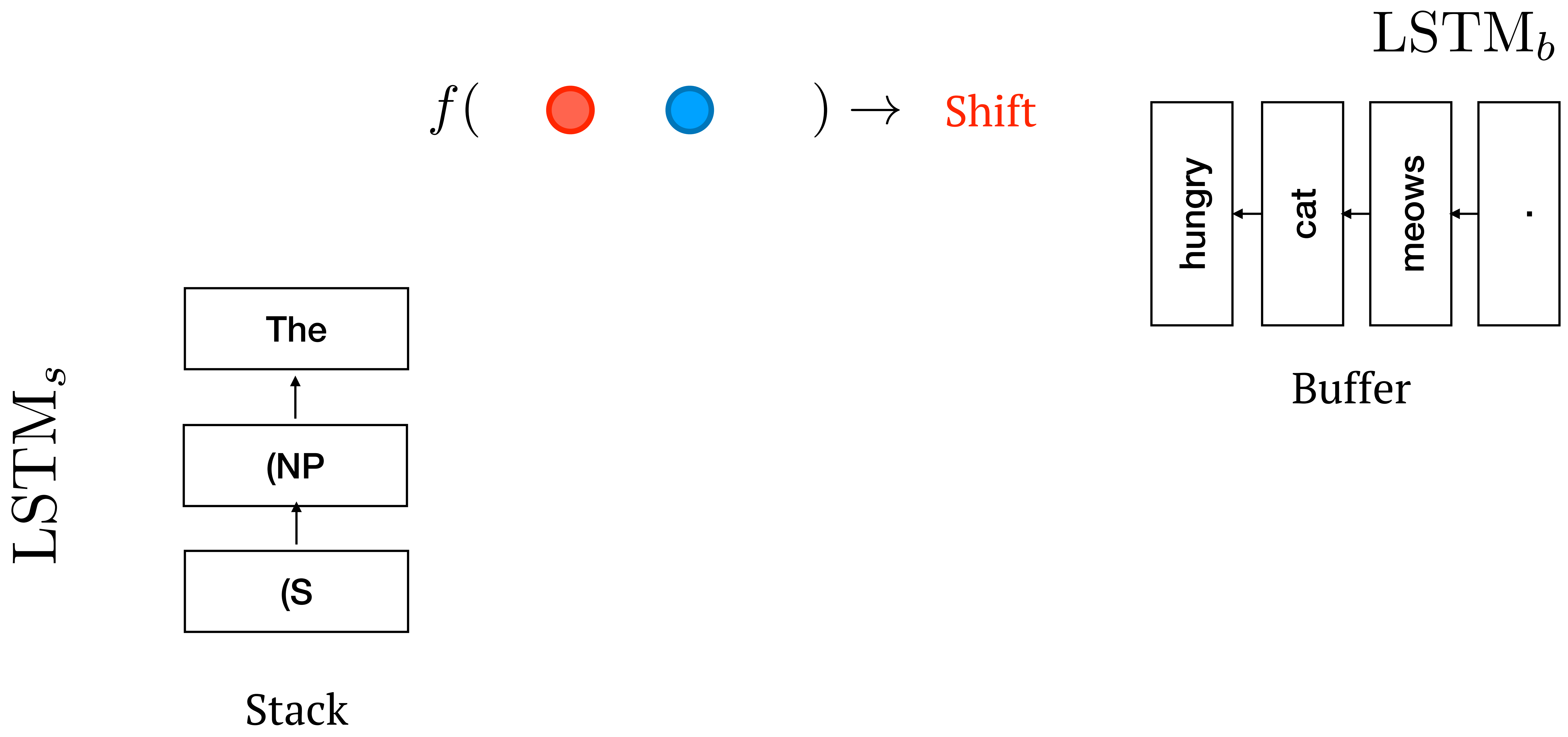
(Dyer et al, 2015)

Stack LSTMs



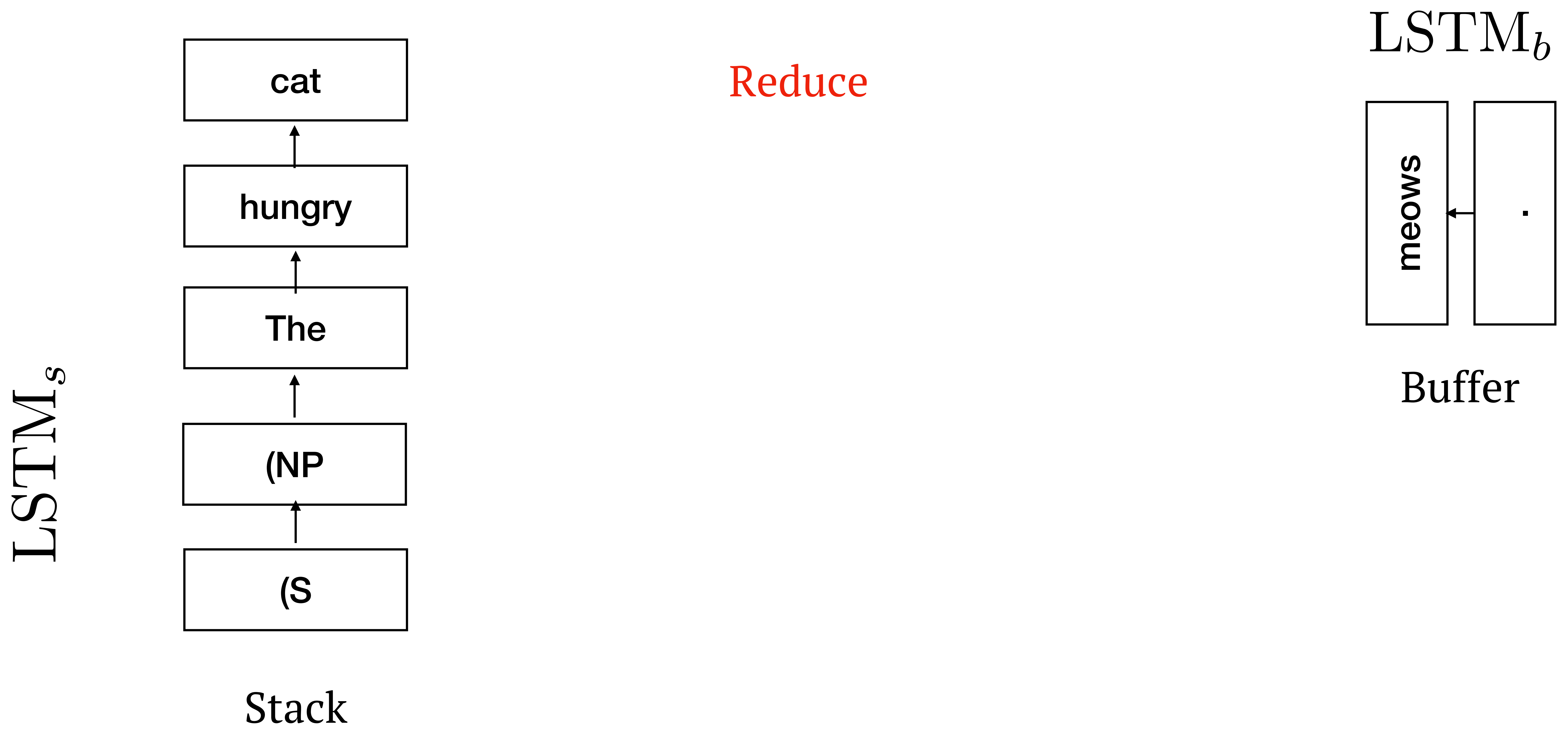
(Dyer et al, 2015)

Stack LSTMs



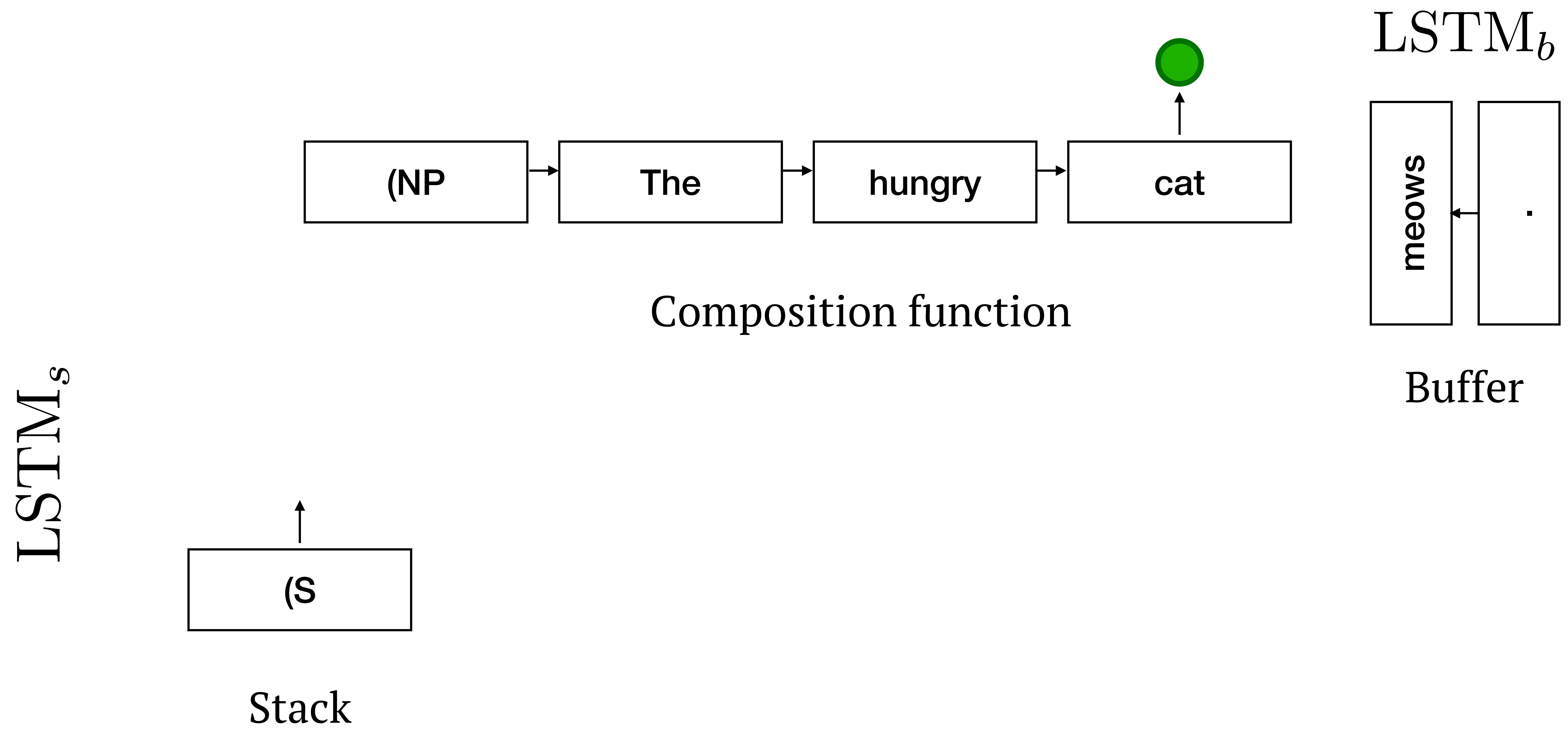
(Dyer et al, 2015)

Stack LSTMs



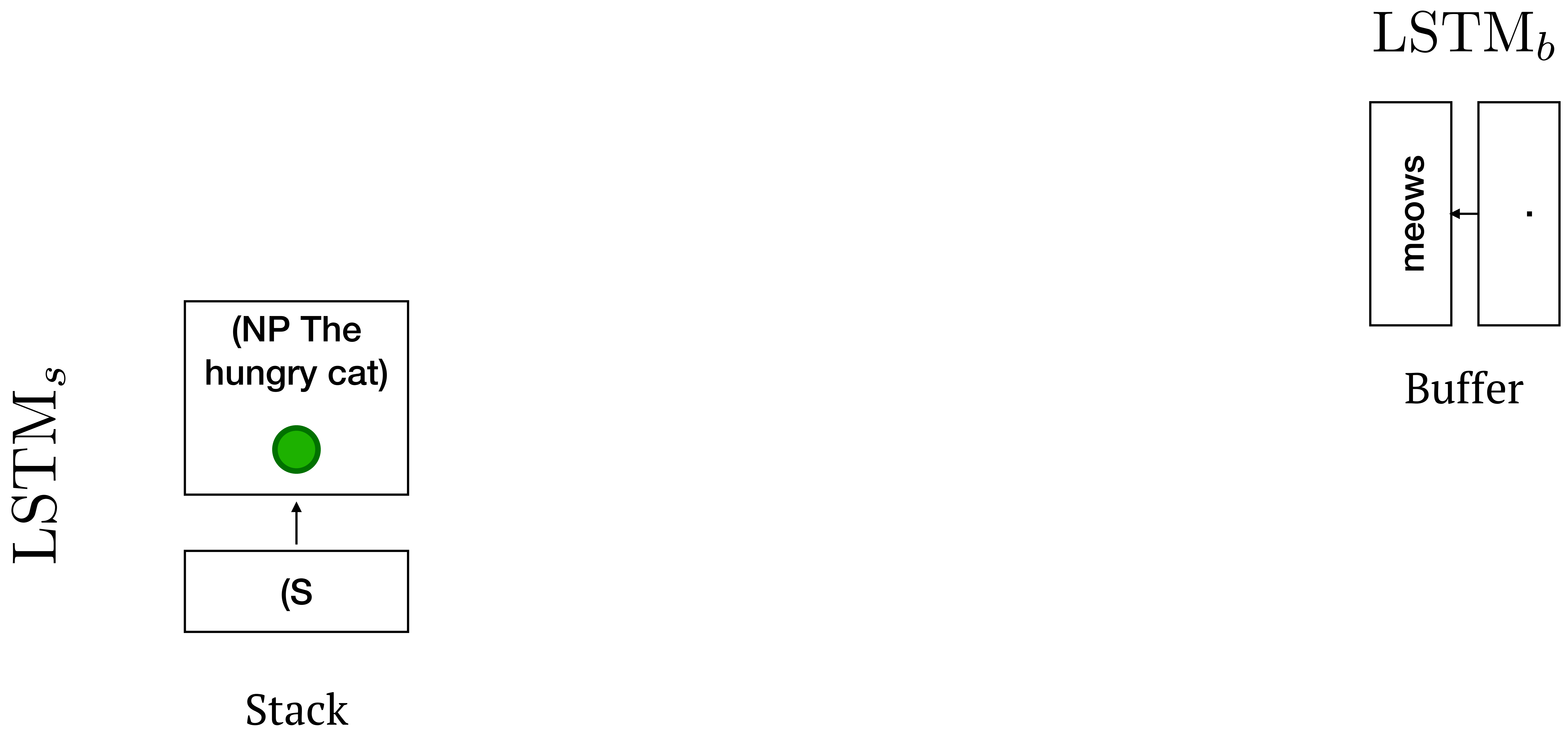
(Dyer et al, 2015)

Stack LSTMs



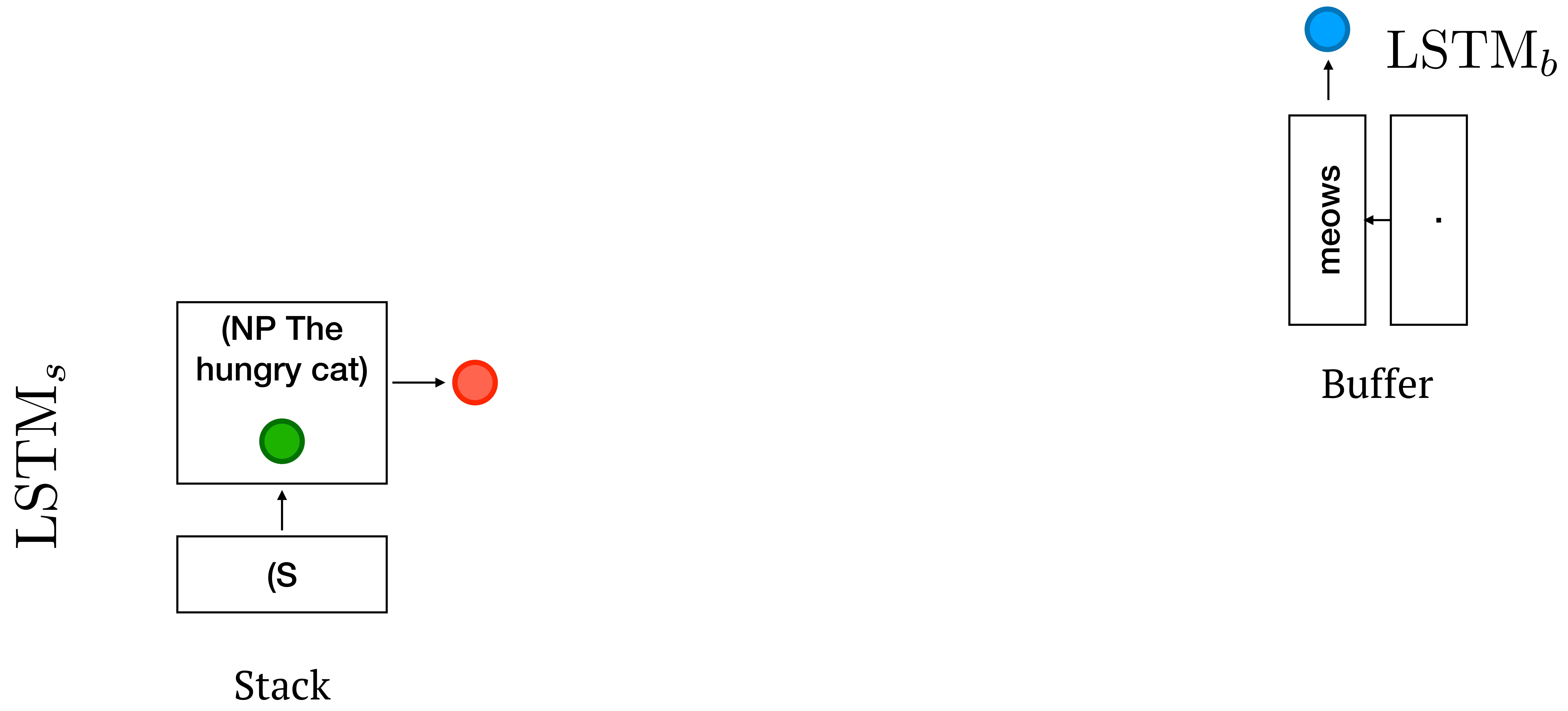
(Dyer et al, 2015)

Stack LSTMs



(Dyer et al, 2015)

Stack LSTMs



(Dyer et al, 2015)