Recurrent Neural Networks

THE
UNIVERSITY
OF UTAH

Overview

1. Modeling sequences
2. Recurrent neural networks: An abstraction
3. Usage patterns for RNNs
4. BiDirectional RNNs
5. A concrete example: The Elman RNN
6. The vanishing gradient problem
7. Long short-term memory units

Overview

1. Modeling sequences
2. Recurrent neural networks: An abstraction
3. Usage patterns for RNNs
4. *BiDirectional RNNs*
5. A concrete example: The Elman RNN
6. The vanishing gradient problem
7. Long short-term memory units

Why left to right?

Everything we saw so far models sequences (e.g. words) from left to right

Implicit assumption: If we want to represent a word in a sentence, the words before are useful

Is this right?

Why left to right?

Everything we saw so far models sequences (e.g. words) from left to right

Implicit assumption: If we want to represent a word in a sentence, the words before are useful

Is this right? Not really

For example: For a sequence labeling task, the words after a target word may also be useful in deciding its label

How do we address this?

Bidirectional RNNs

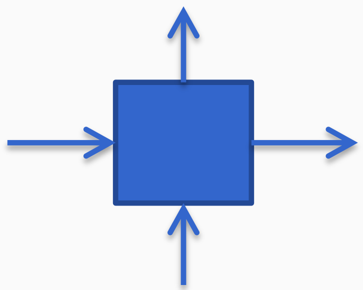
[Schuster and Paliwal 1997]

One answer: Maintain two separate RNNs – one forward and one reverse

BiRNN: A simple example

Forward

John ate cake



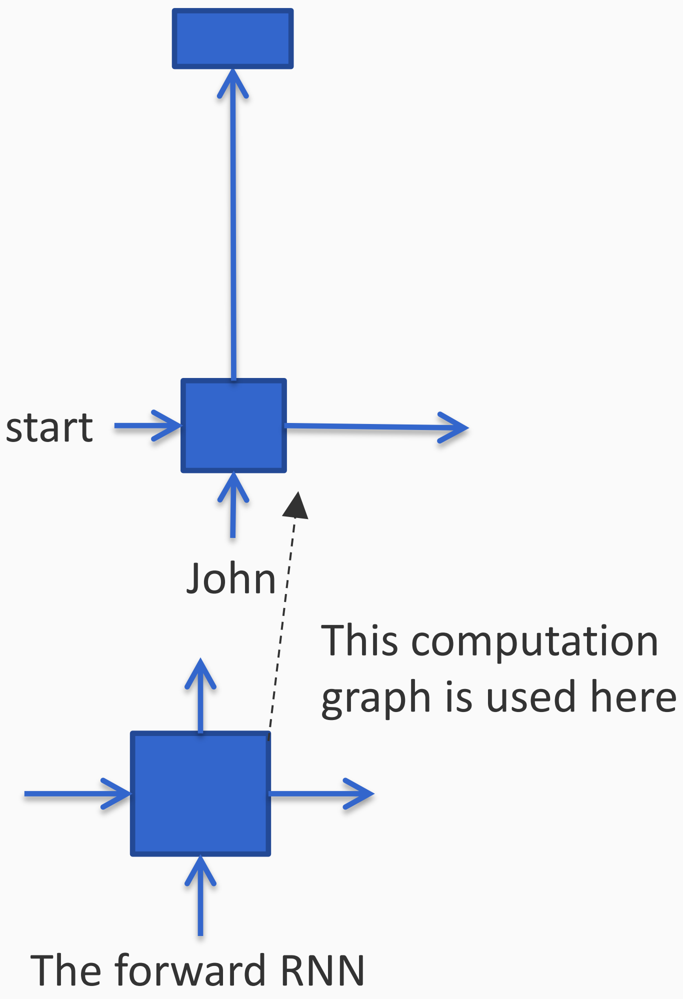
The forward RNN

First, the forward case. We have seen this before.

BiRNN: A simple example

Forward

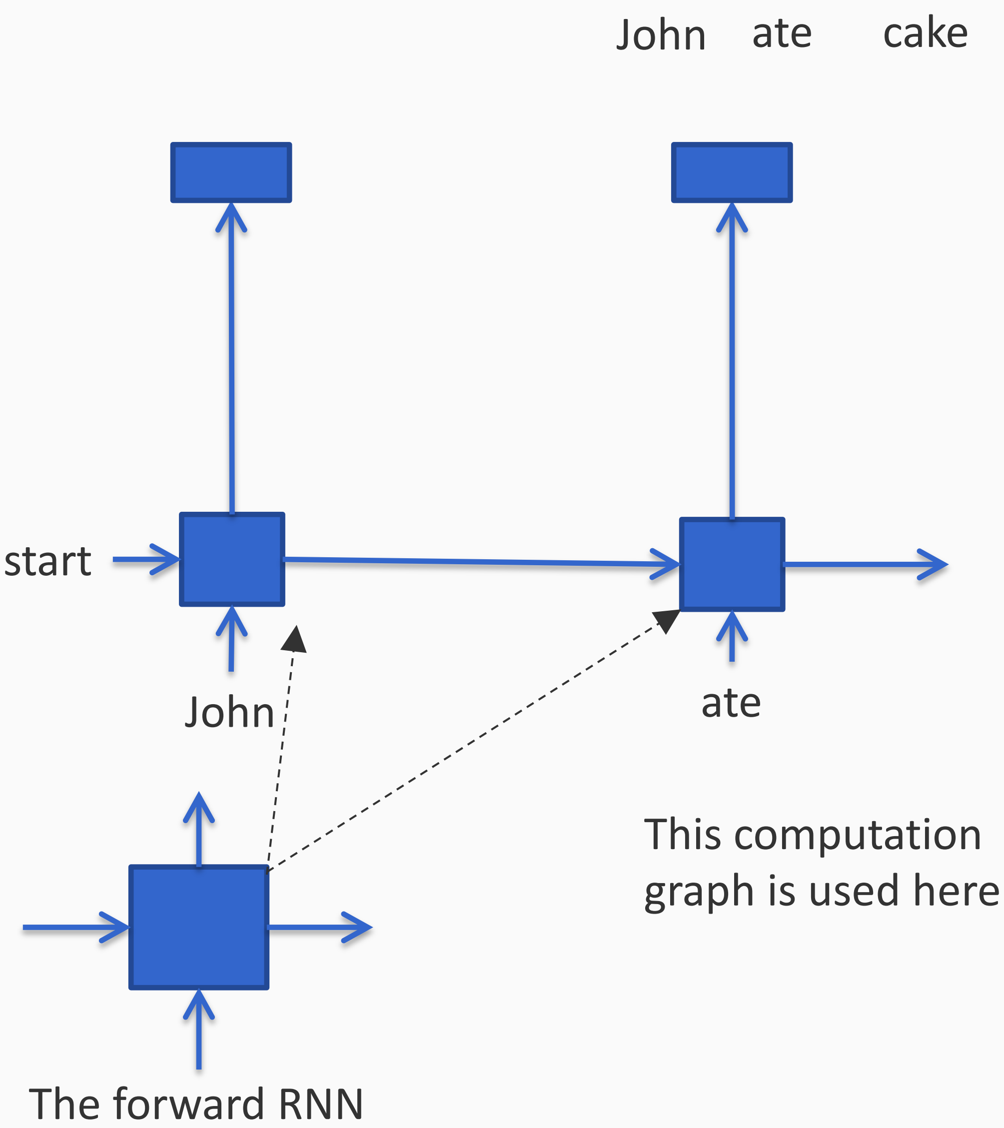
John ate cake



First, the forward case. We have seen this before.

BiRNN: A simple example

Forward

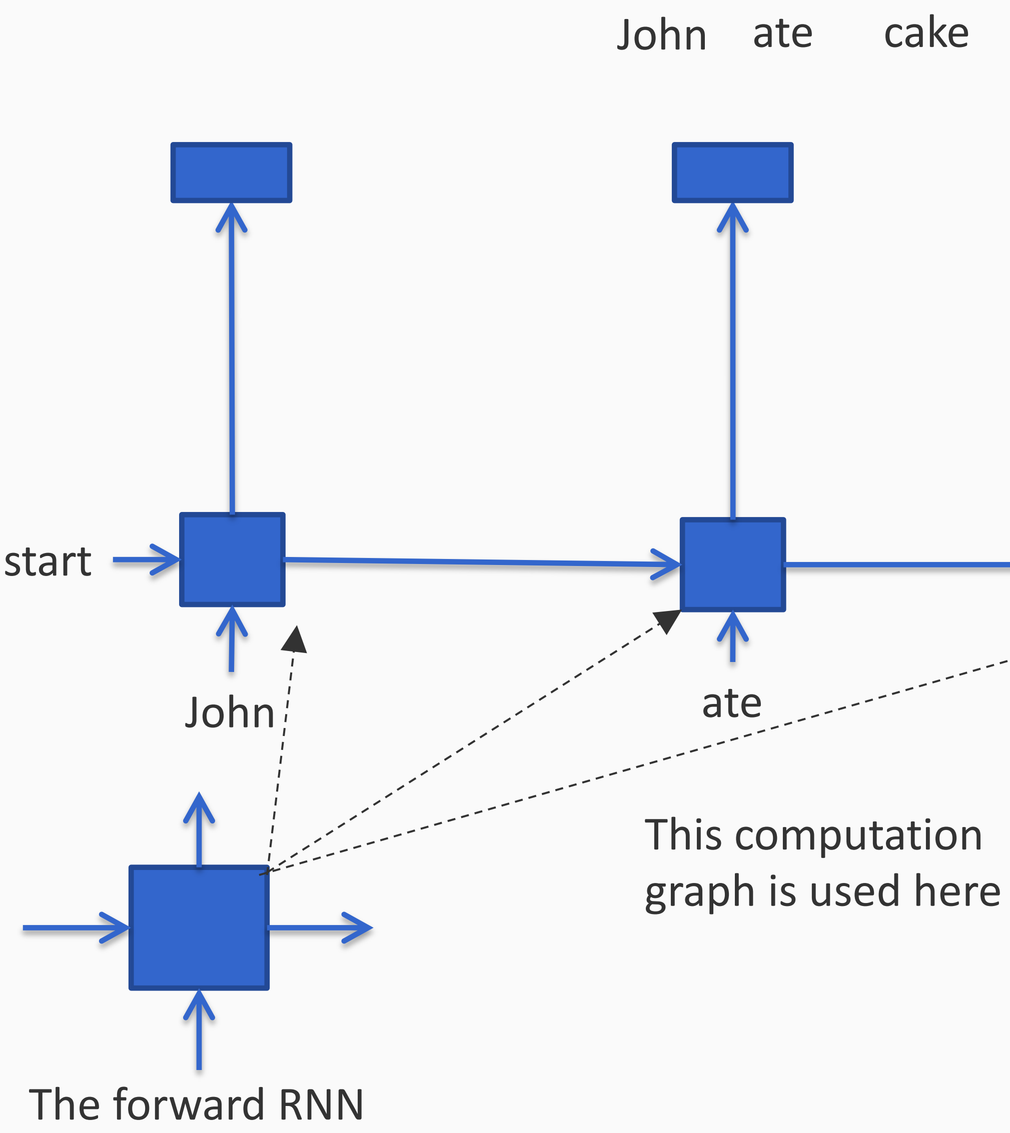


This computation graph is used here

First, the forward case. We have seen this before.

BiRNN: A simple example

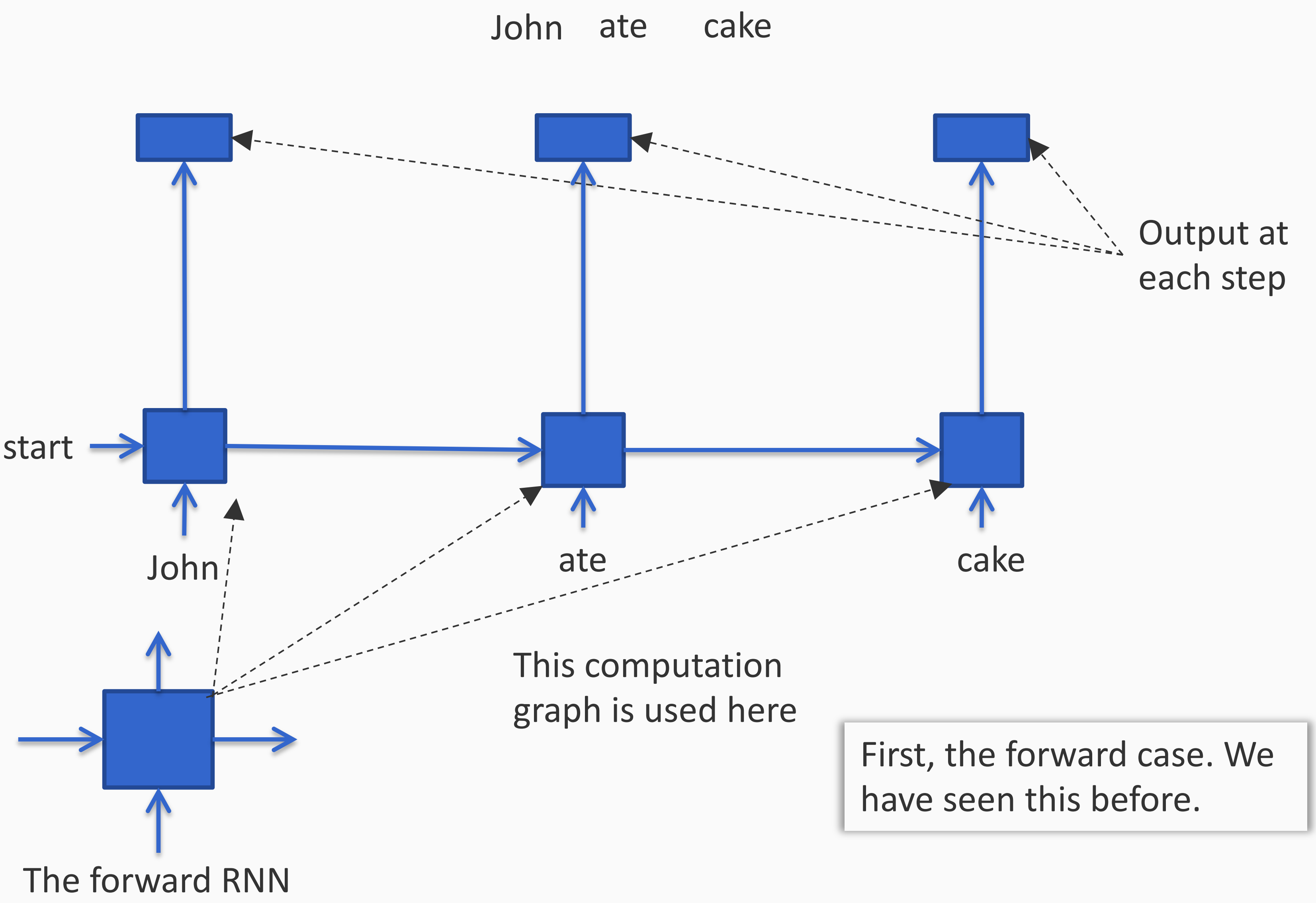
Forward



First, the forward case. We have seen this before.

BiRNN: A simple example

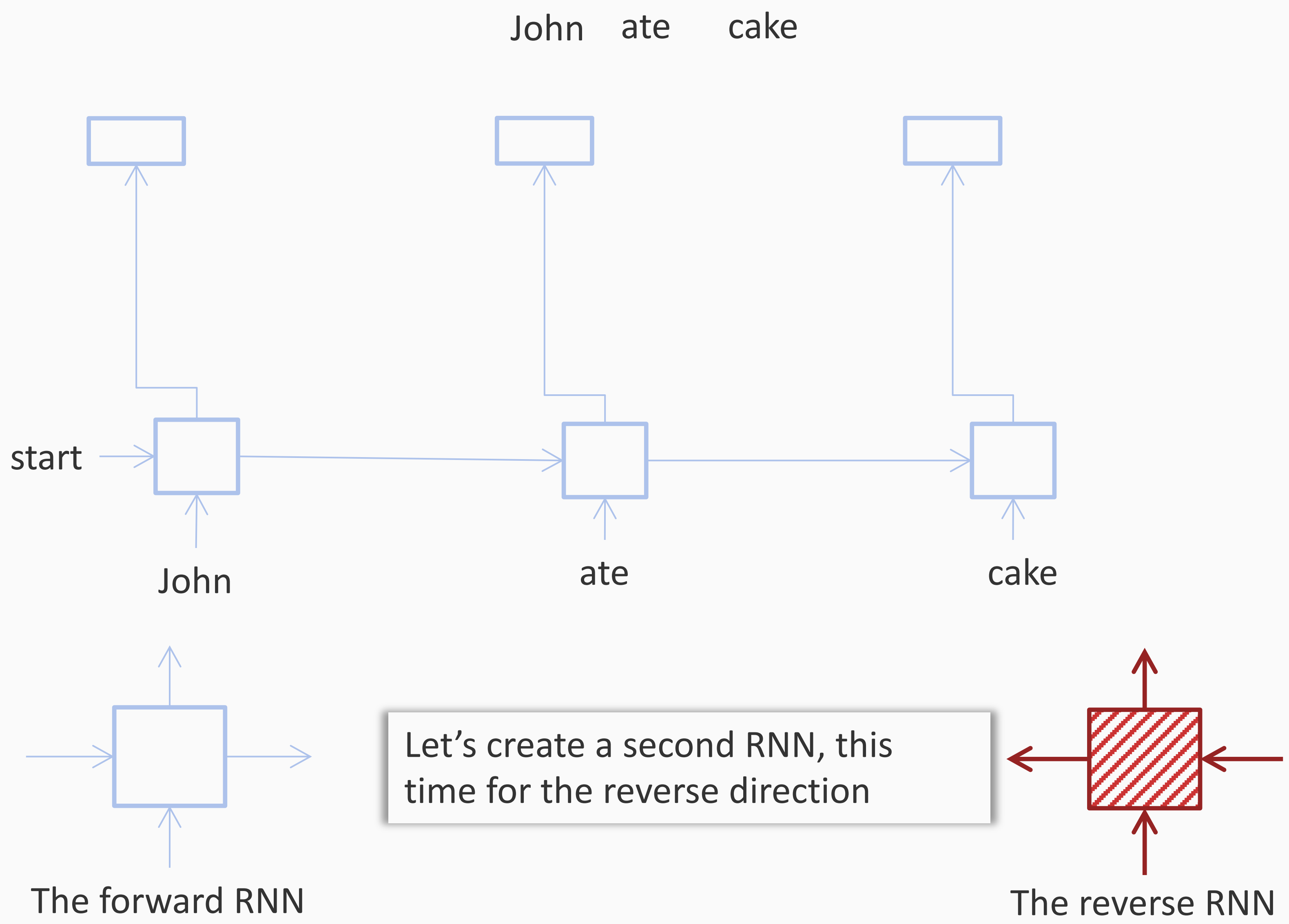
Forward



First, the forward case. We have seen this before.

BiRNN: A simple example

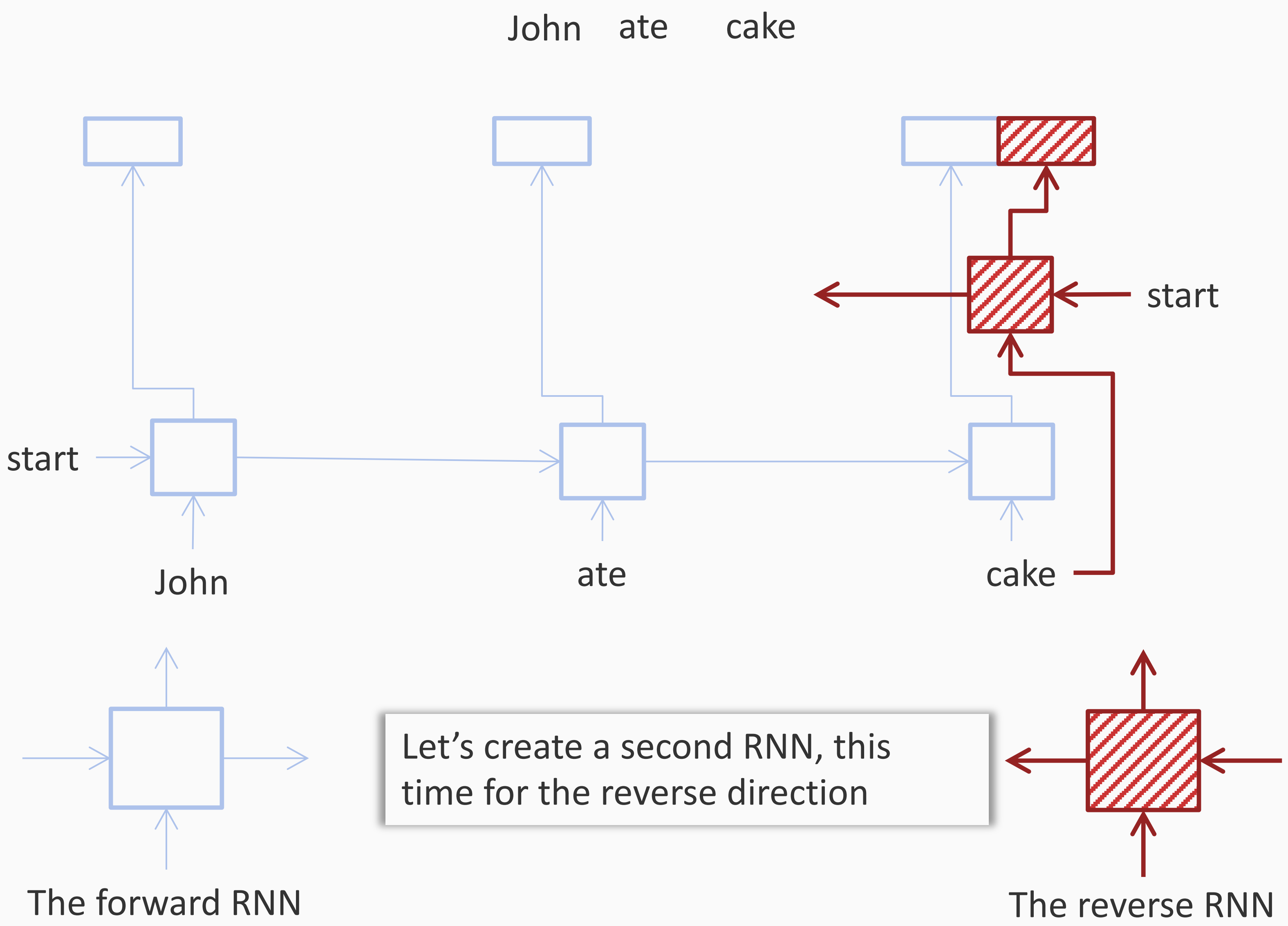
Reverse



Let’s create a second RNN, this time for the reverse direction

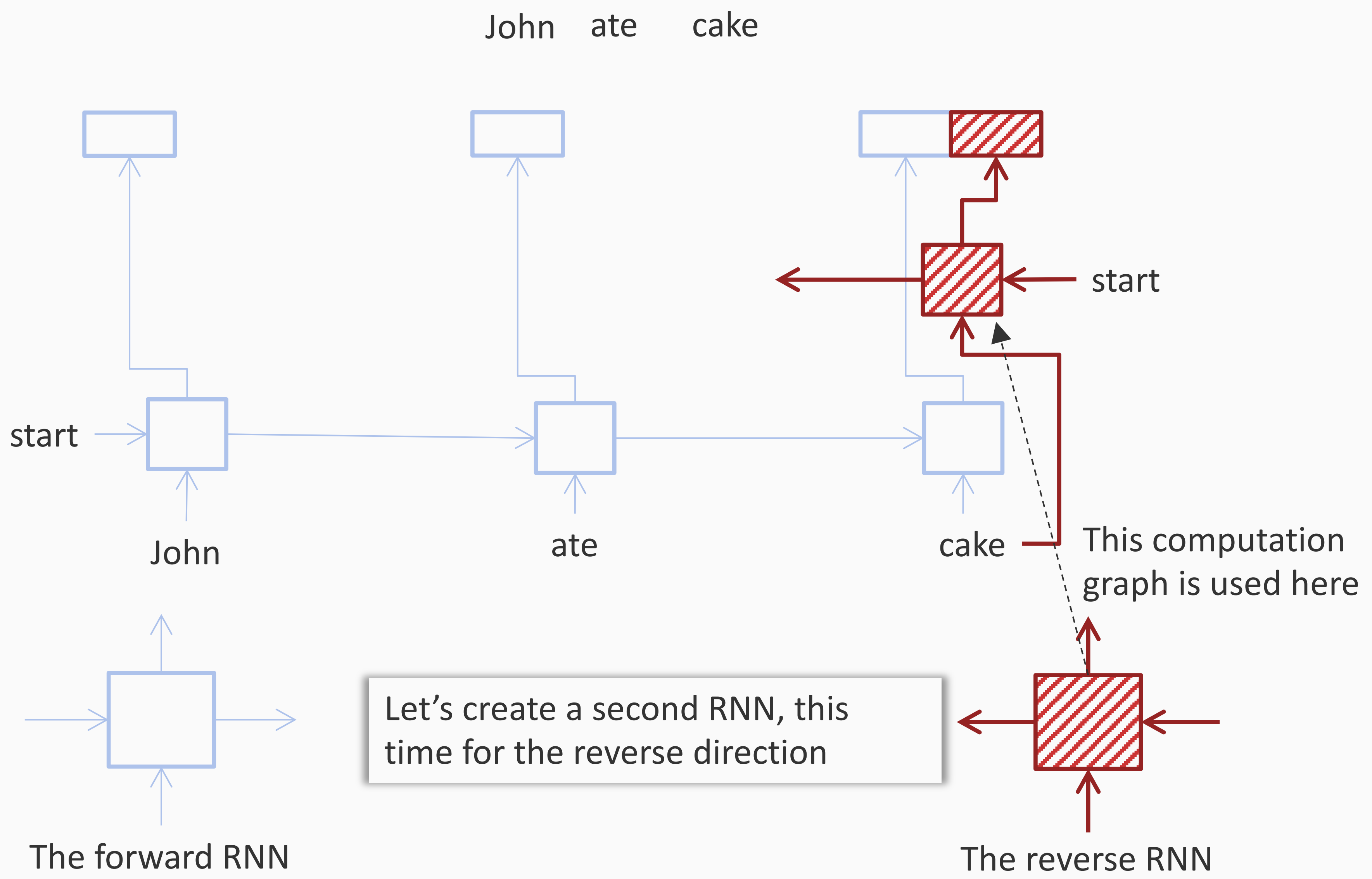
BiRNN: A simple example

Reverse



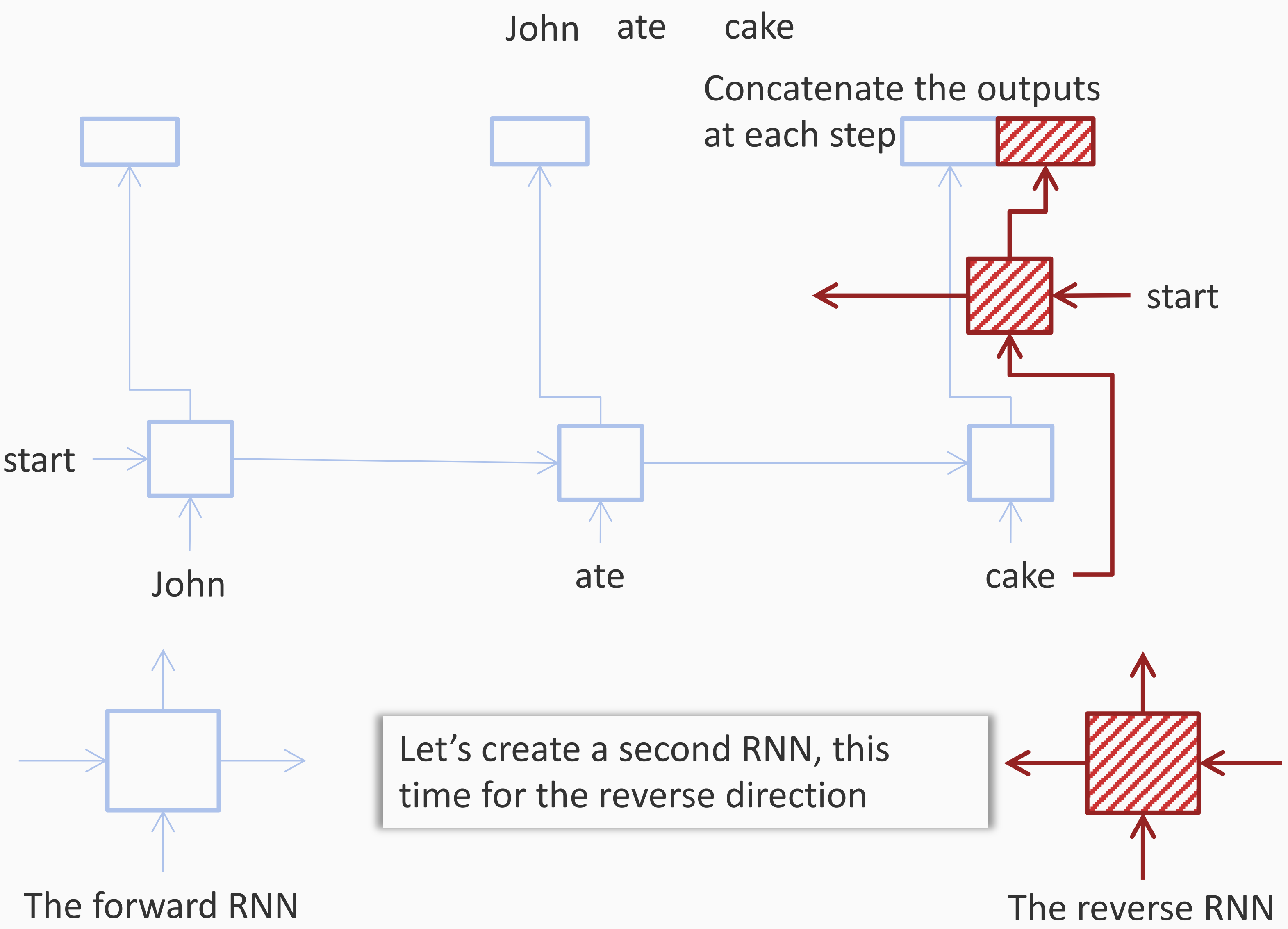
BiRNN: A simple example

Reverse



BiRNN: A simple example

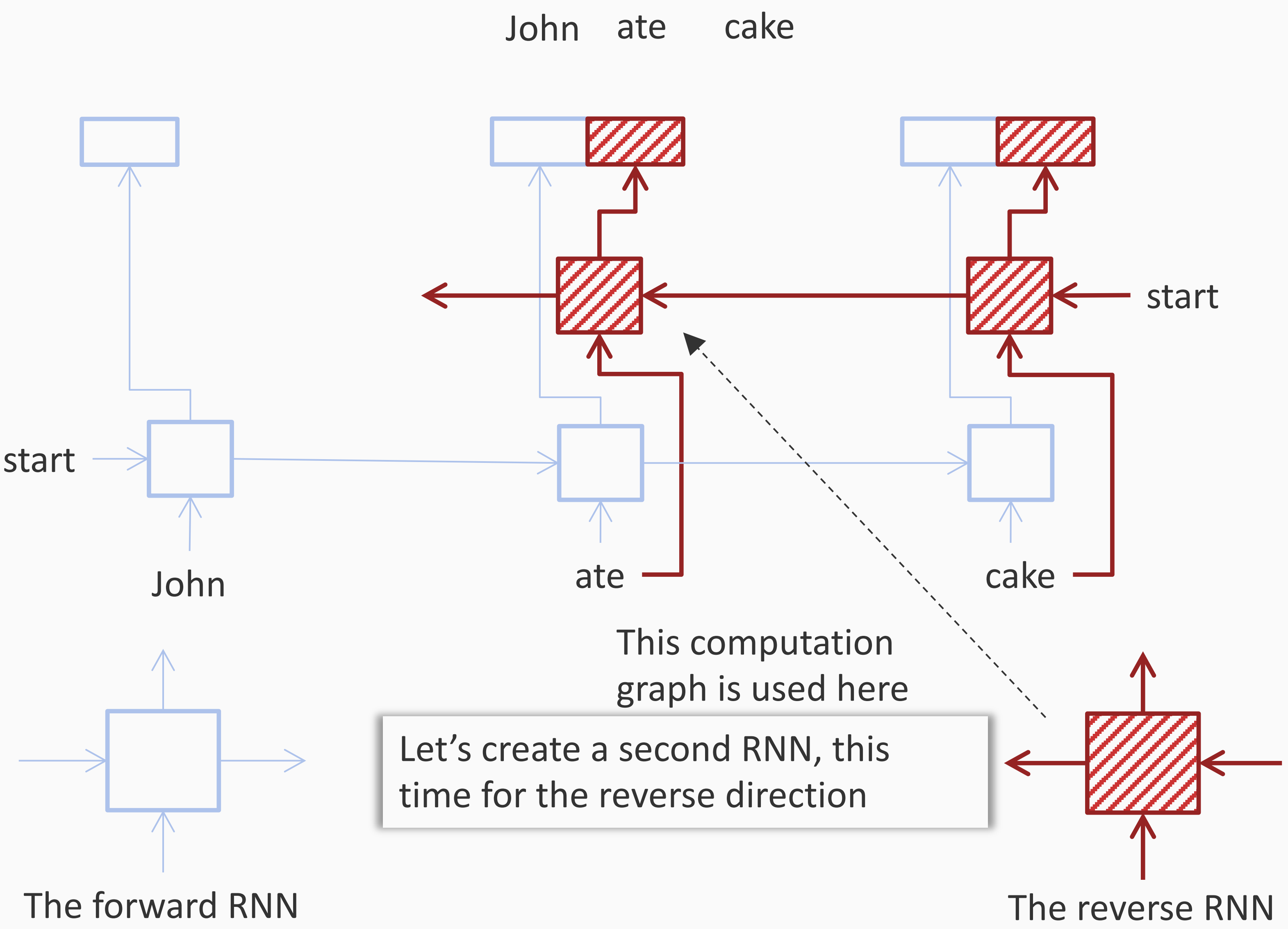
Reverse



Let's create a second RNN, this time for the reverse direction

BiRNN: A simple example

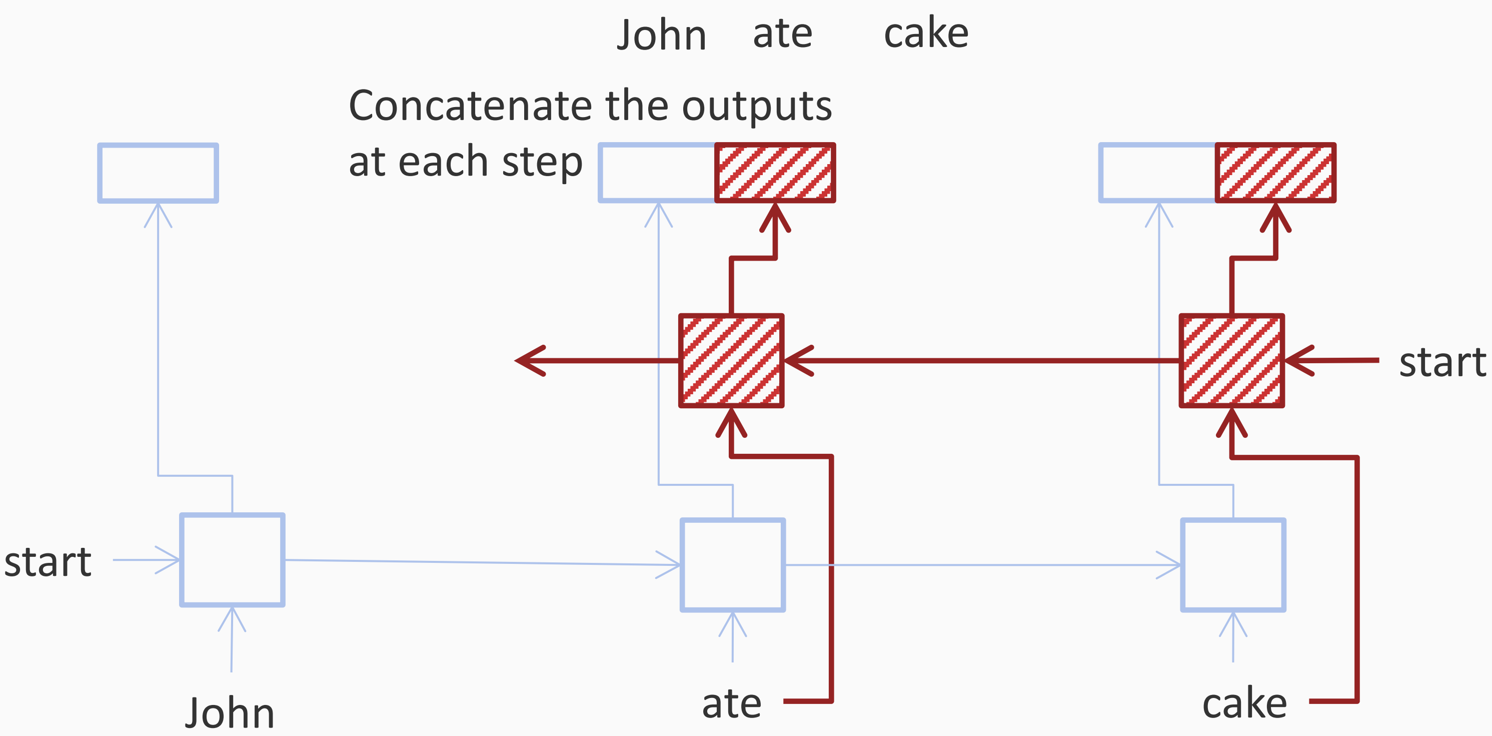
Reverse

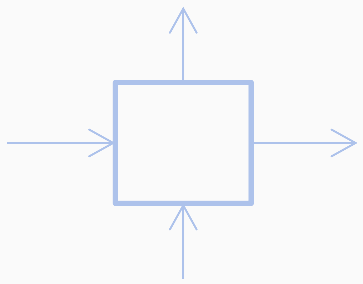


Let's create a second RNN, this time for the reverse direction

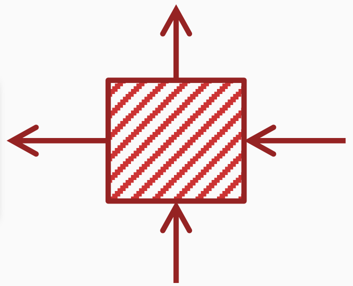
BiRNN: A simple example

Reverse



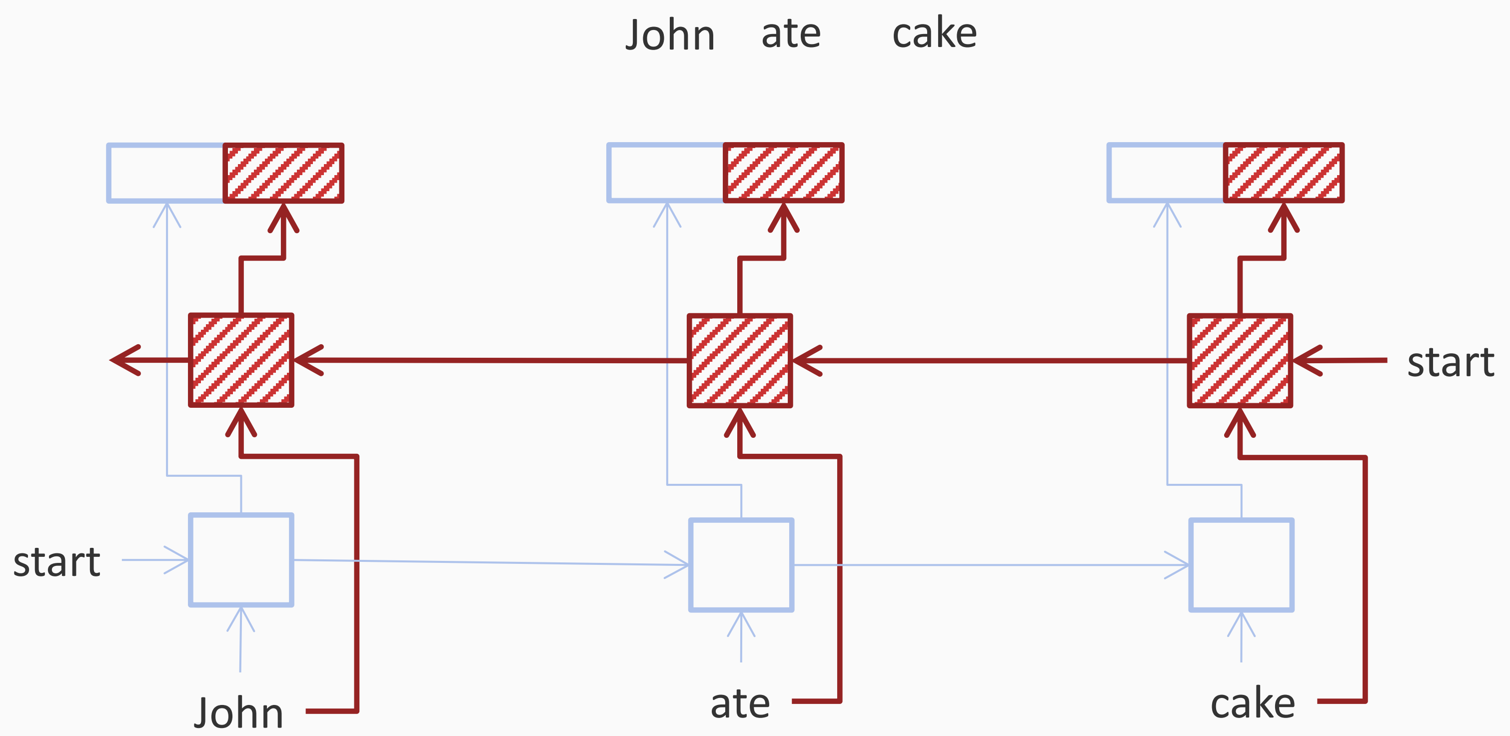


Let's create a second RNN, this time for the reverse direction

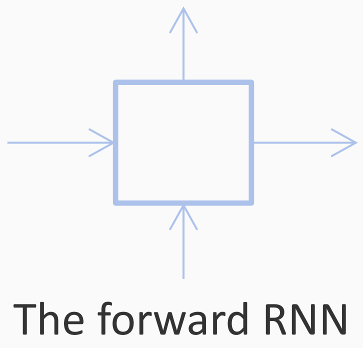


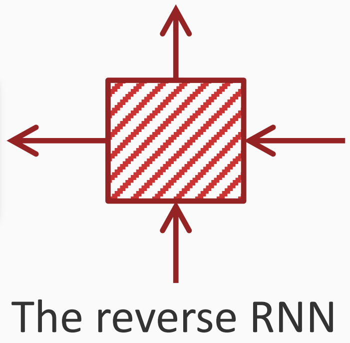
BiRNN: A simple example

Reverse



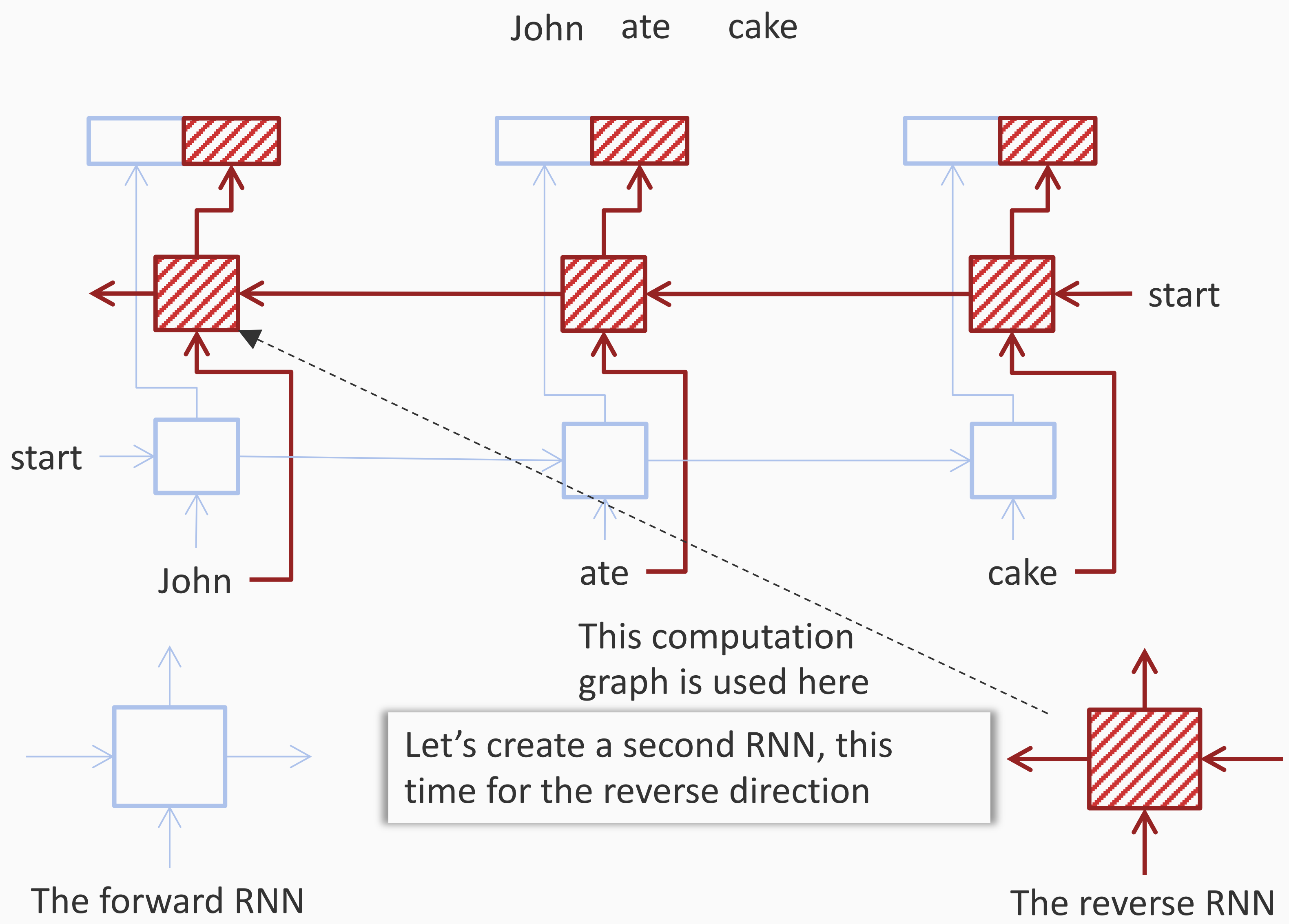
Let's create a second RNN, this time for the reverse direction





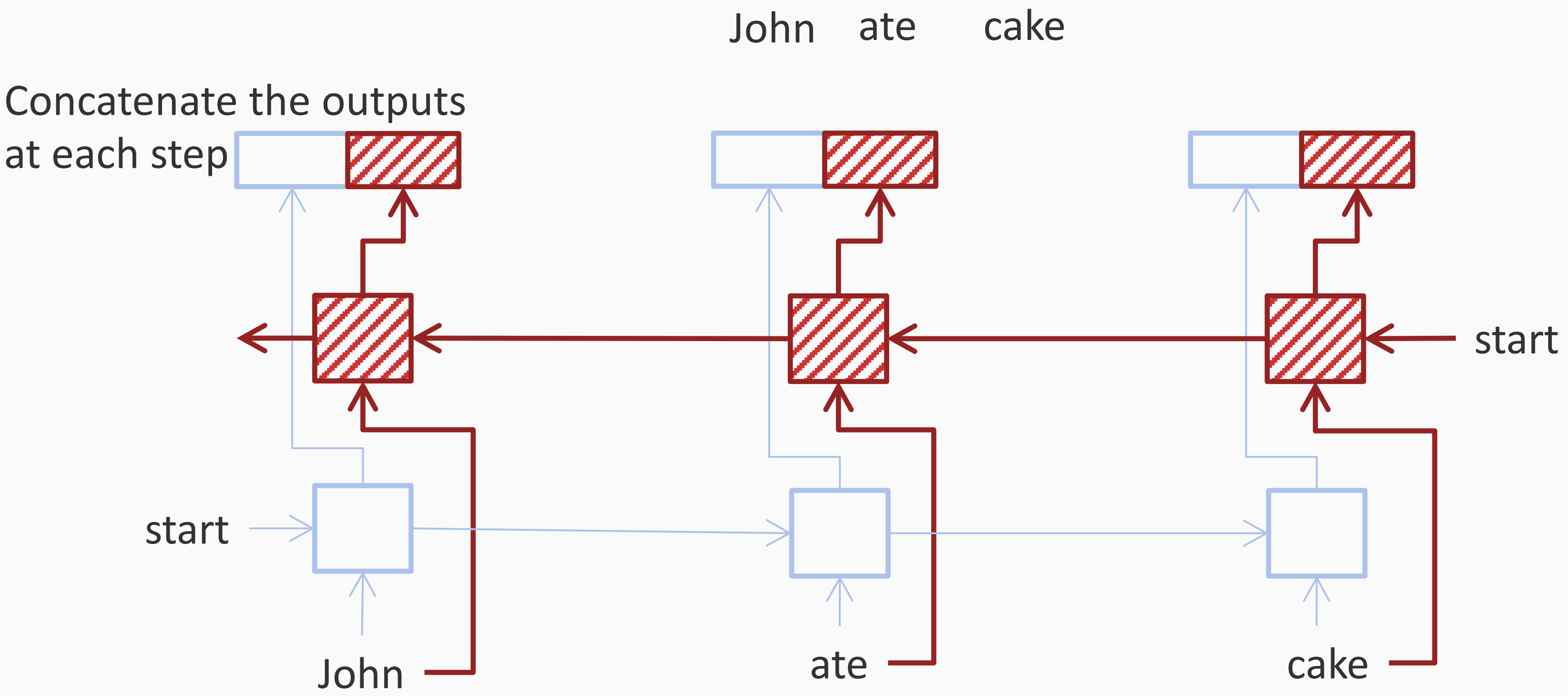
BiRNN: A simple example

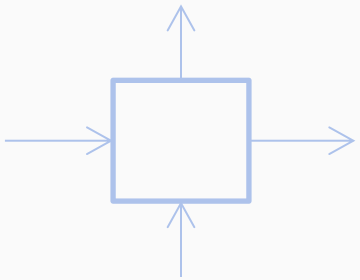
Reverse



BiRNN: A simple example

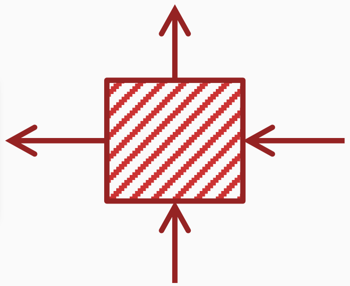
Reverse





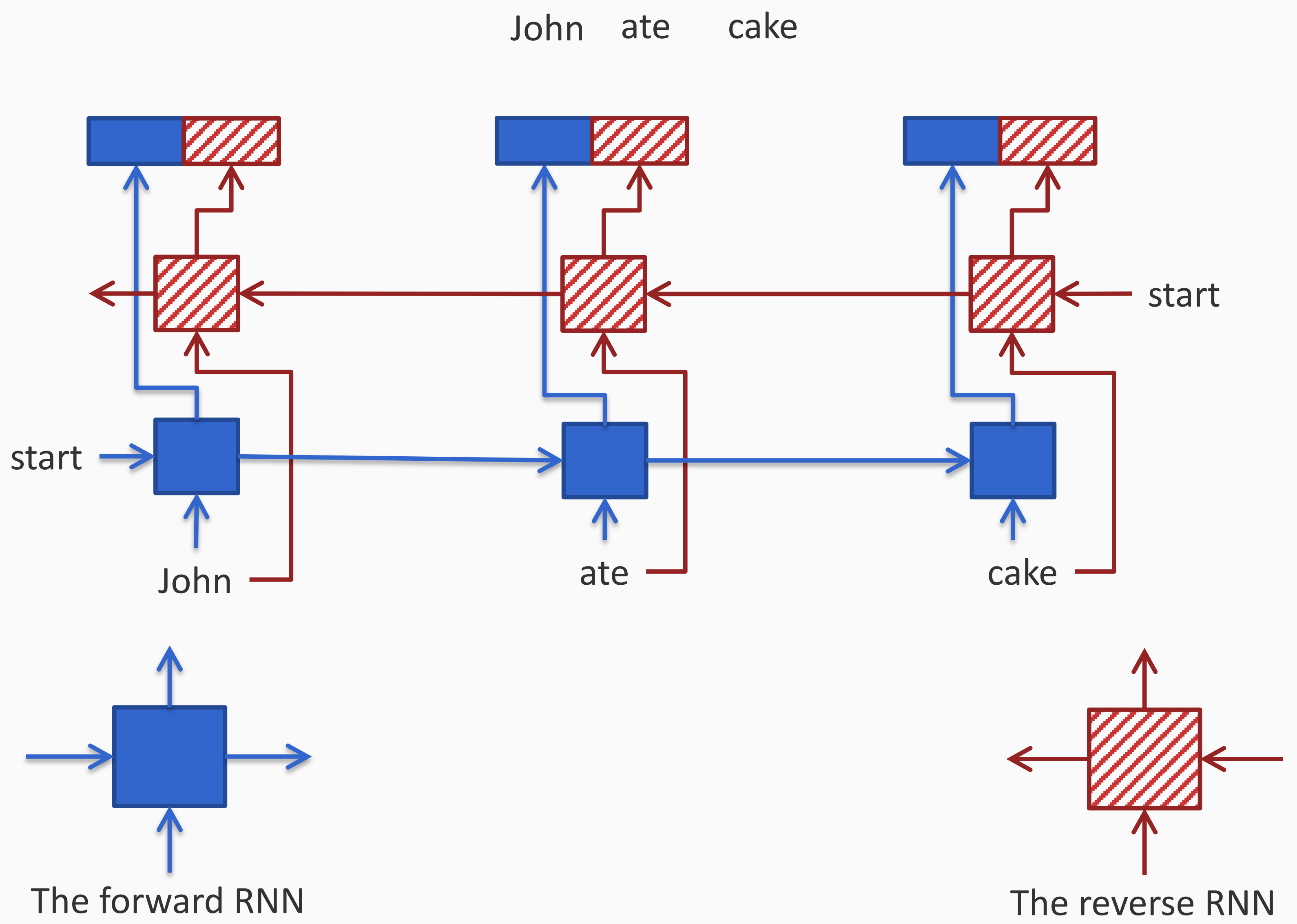
The forward RNN

Let's create a second RNN, this time for the reverse direction



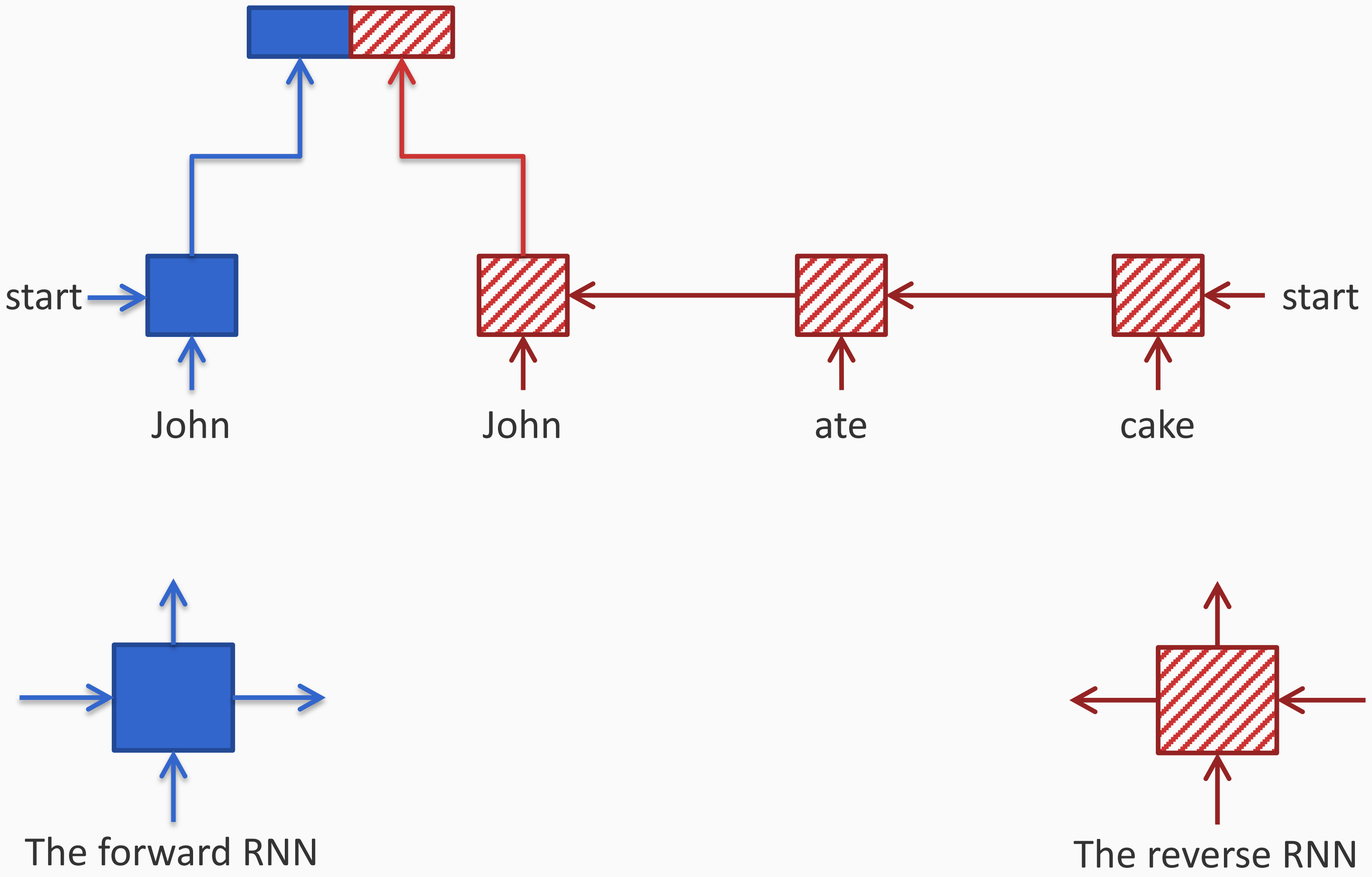
The reverse RNN

BiRNN: Putting both parts together



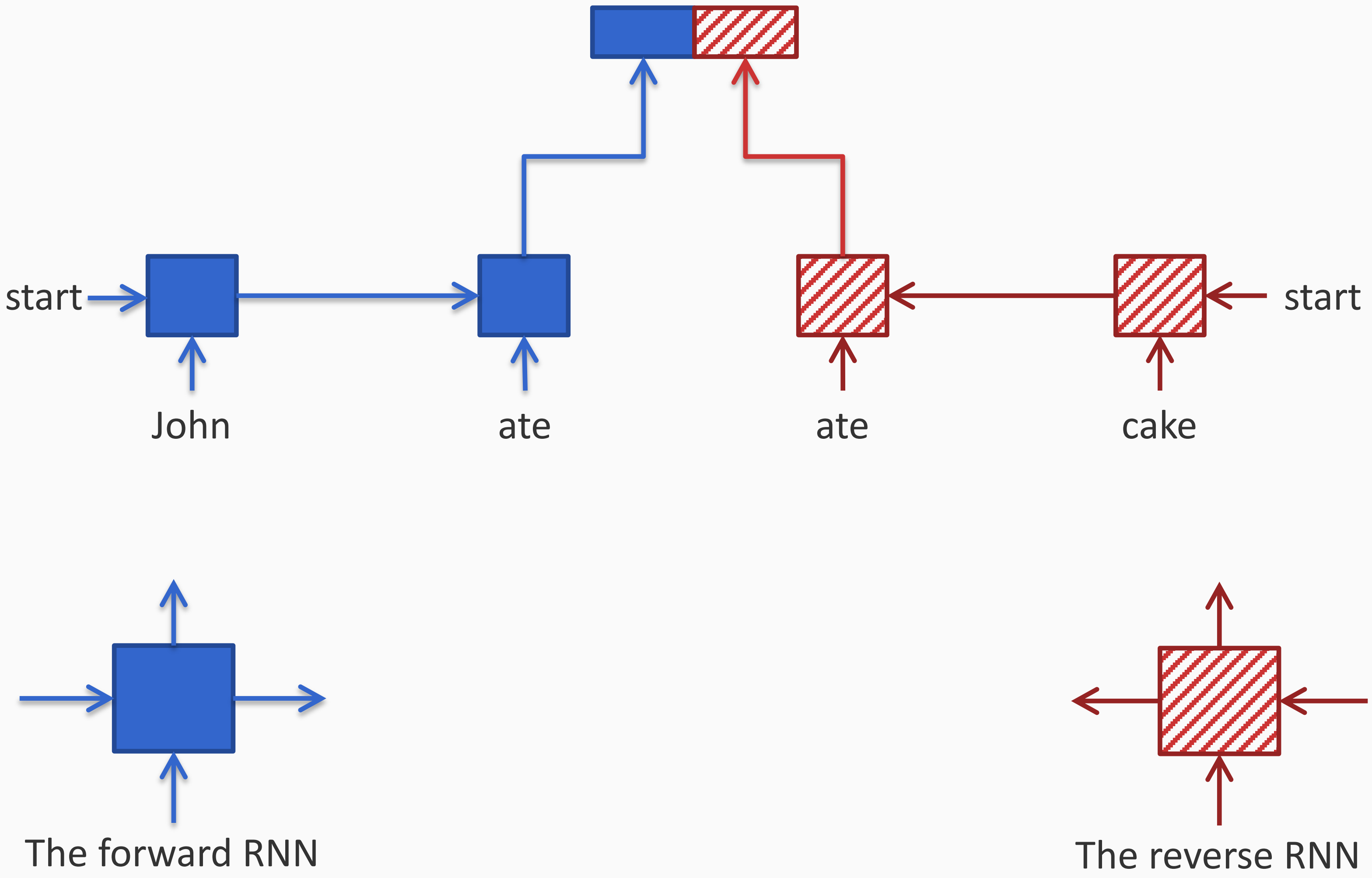
Another way of seeing this

Concatenate to get the representation for the word ***John*** that accounts for both left and right contexts



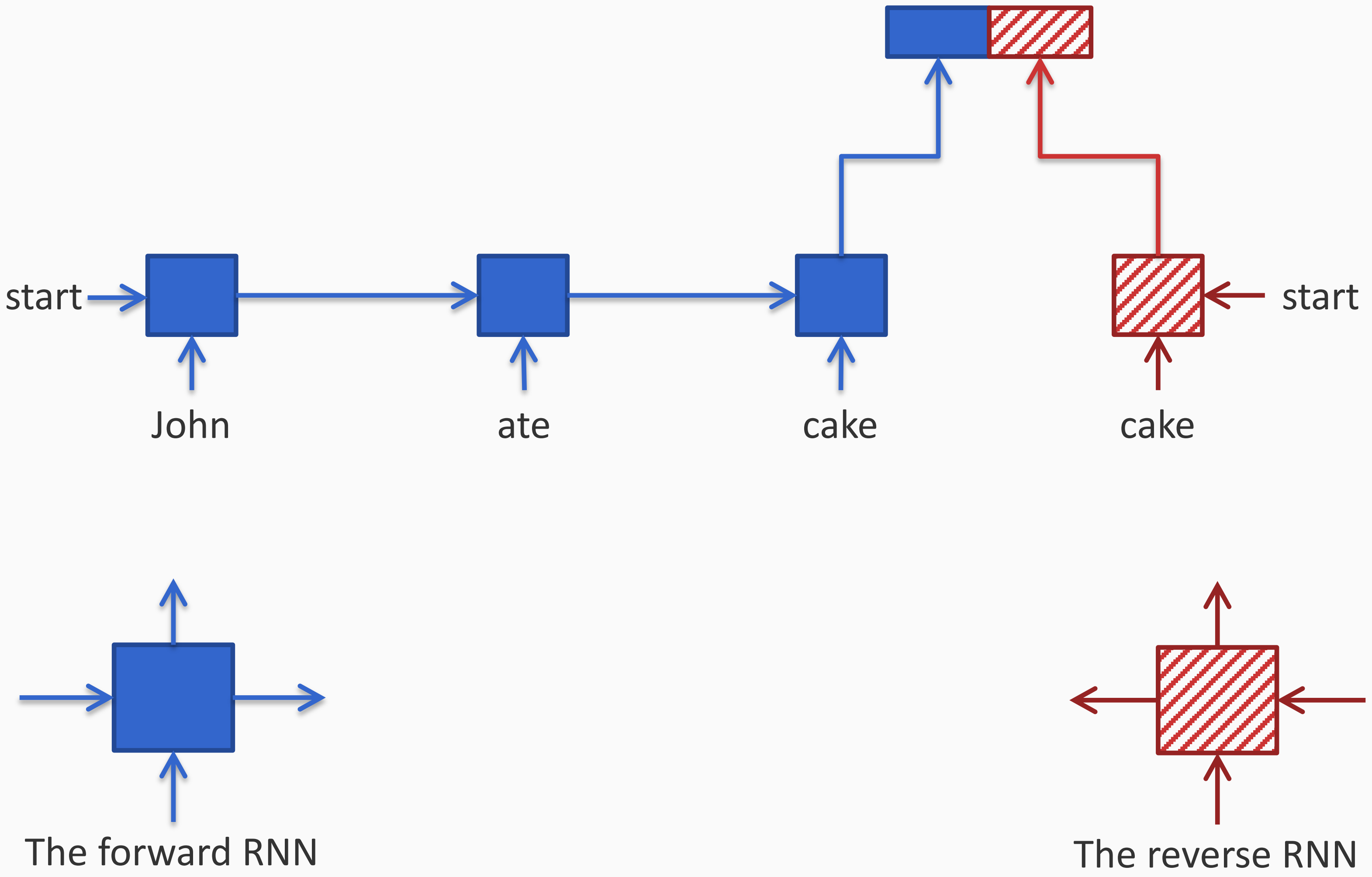
Another way of seeing this

Concatenate to get the representation for the word ***ate*** that accounts for both left and right contexts



Another way of seeing this

Concatenate to get the representation for the word ***cake*** that accounts for both left and right contexts



A Bidirectional RNN

- Two RNNs
  - Forward, defined by functions $R^f(\mathbf{s}_{t-1}^f, \mathbf{x}_t)$ and $O^f(\mathbf{s}_t)$
  - Backward, defined by functions $R^b(\mathbf{s}_{t+1}^b, \mathbf{x}_t)$ and $O^b(\mathbf{s}_t)$

-

A Bidirectional RNN

- Two RNNs
  - Forward, defined by functions $R^f(\mathbf{s}^f_{t-1}, \mathbf{x}_t)$ and $O^f(\mathbf{s}_t)$
  - Backward, defined by functions $R^b(\mathbf{s}^b_{t+1}, \mathbf{x}_t)$ and $O^b(\mathbf{s}_t)$

- The i^{th} output is defined by

$$\mathbf{y}_i = [O^f(\mathbf{s}^f_t), O^b(\mathbf{s}^b_t)]$$

A Bidirectional RNN

- Two RNNs
  - Forward, defined by functions $R^f(\mathbf{s}_{t-1}^f, \mathbf{x}_t)$ and $O^f(\mathbf{s}_t)$
  - Backward, defined by functions $R^b(\mathbf{s}_{t+1}^b, \mathbf{x}_t)$ and $O^b(\mathbf{s}_t)$

- The i^{th} output is defined by

$$\mathbf{y}_i = [O^f(\mathbf{s}_t^f), O^b(\mathbf{s}_t^b)]$$

- Another way to write this

$$\mathrm{biRNN}(\mathbf{x}_{1:n}, t) = [\mathrm{RNN}^f(\mathbf{x}_{1:t}), \mathrm{RNN}^b(\mathbf{x}_{n:t})]$$

BiRNNs: Summary

- Allows capturing both left and right contexts
- Commonly used if RNNs are used as a base encoding layer for text
  - Often stacked
- Specific versions of RNNs give us different BiRNNs
  - BiLSTMs or BiGRUs typically used