



# What You Need To Know For “Attention Is All You Need”

SW융합대학 소프트웨어학과 김진호 (32221210)

# Agenda

## Background Knowledge

Feed-forward Neural Network (FFNN)

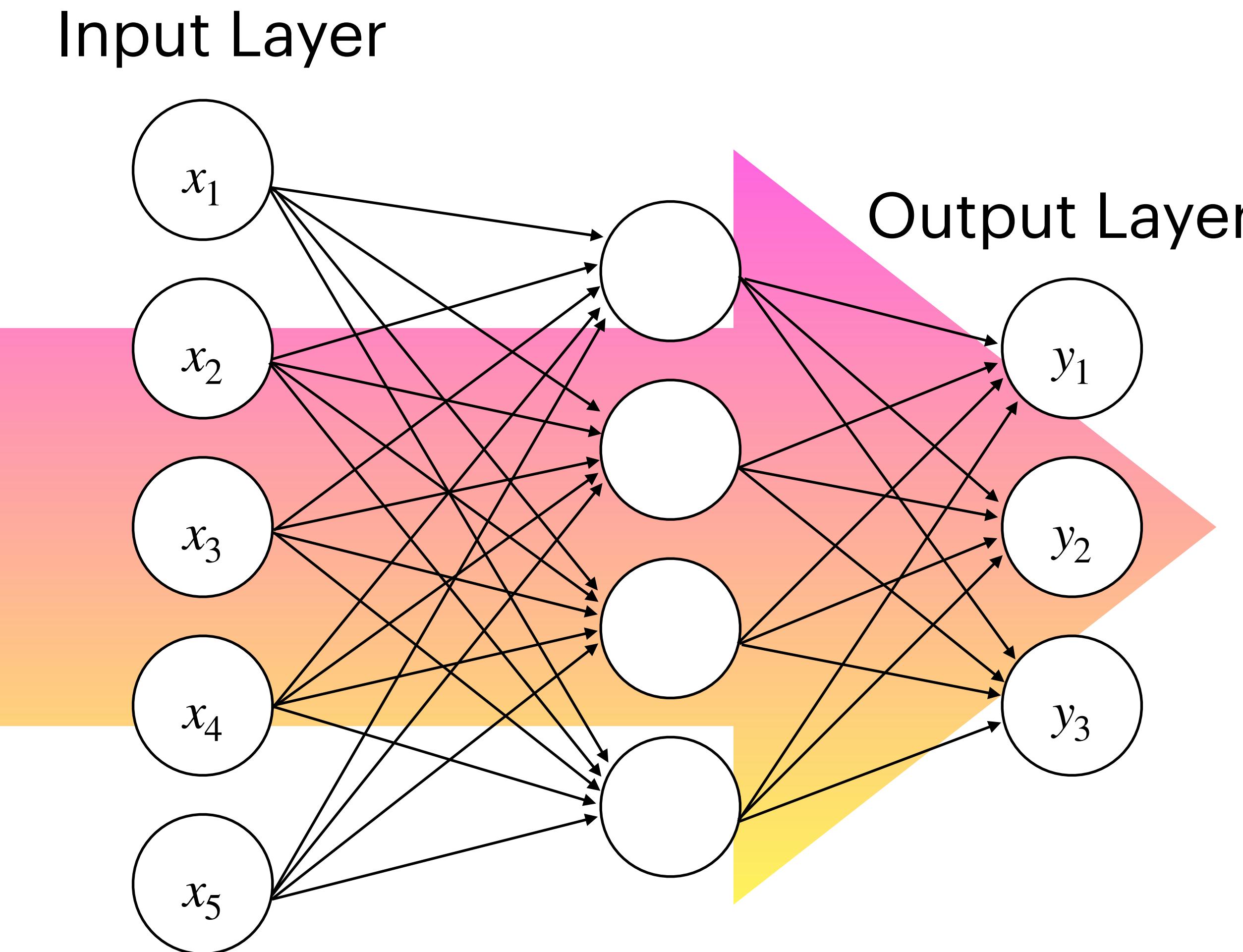
Recurrent Neural Network (RNN)

Sequence to Sequence Learning

Attention Mechanism

# Background Knowledge

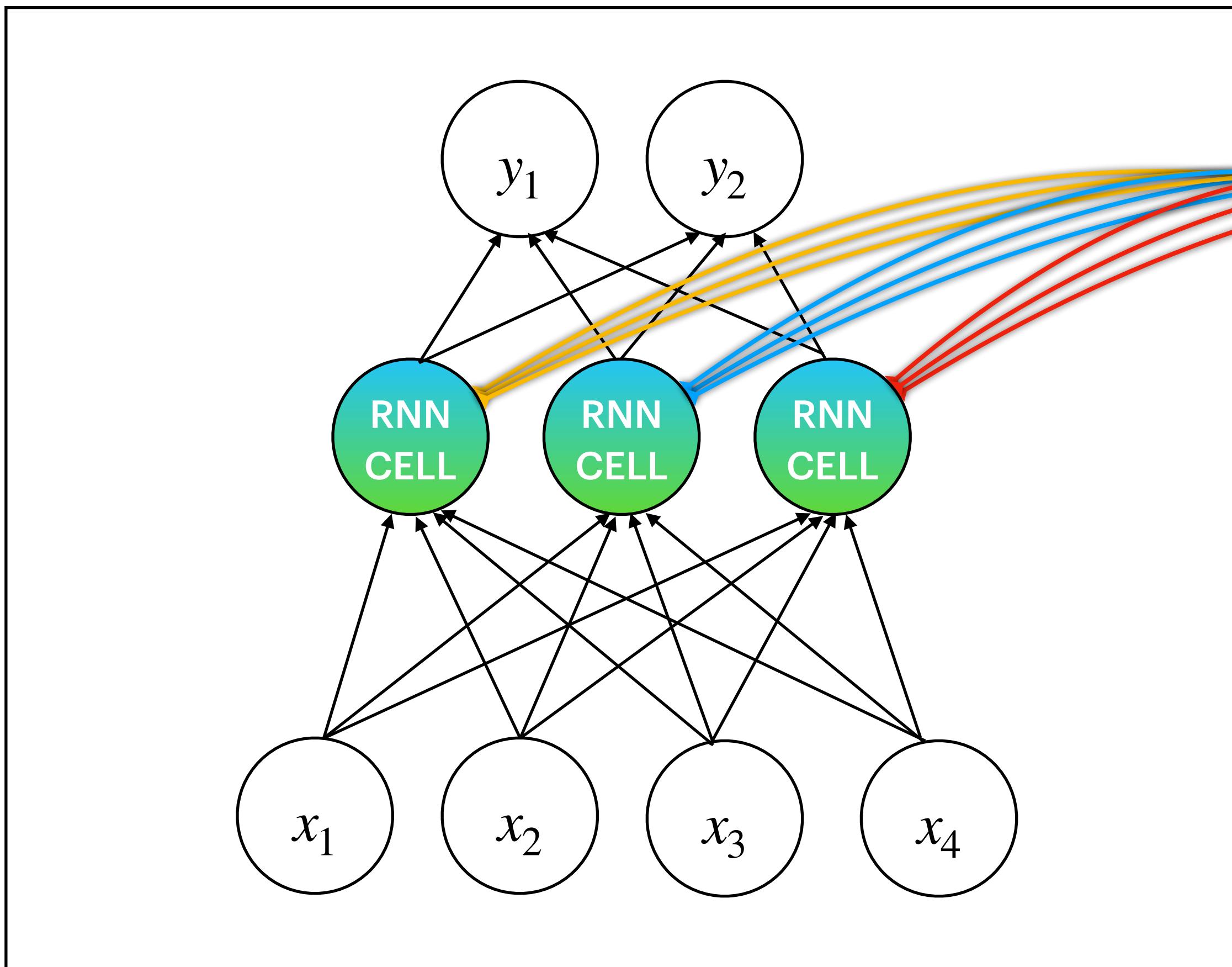
## 1) Feed-forward Neural Network



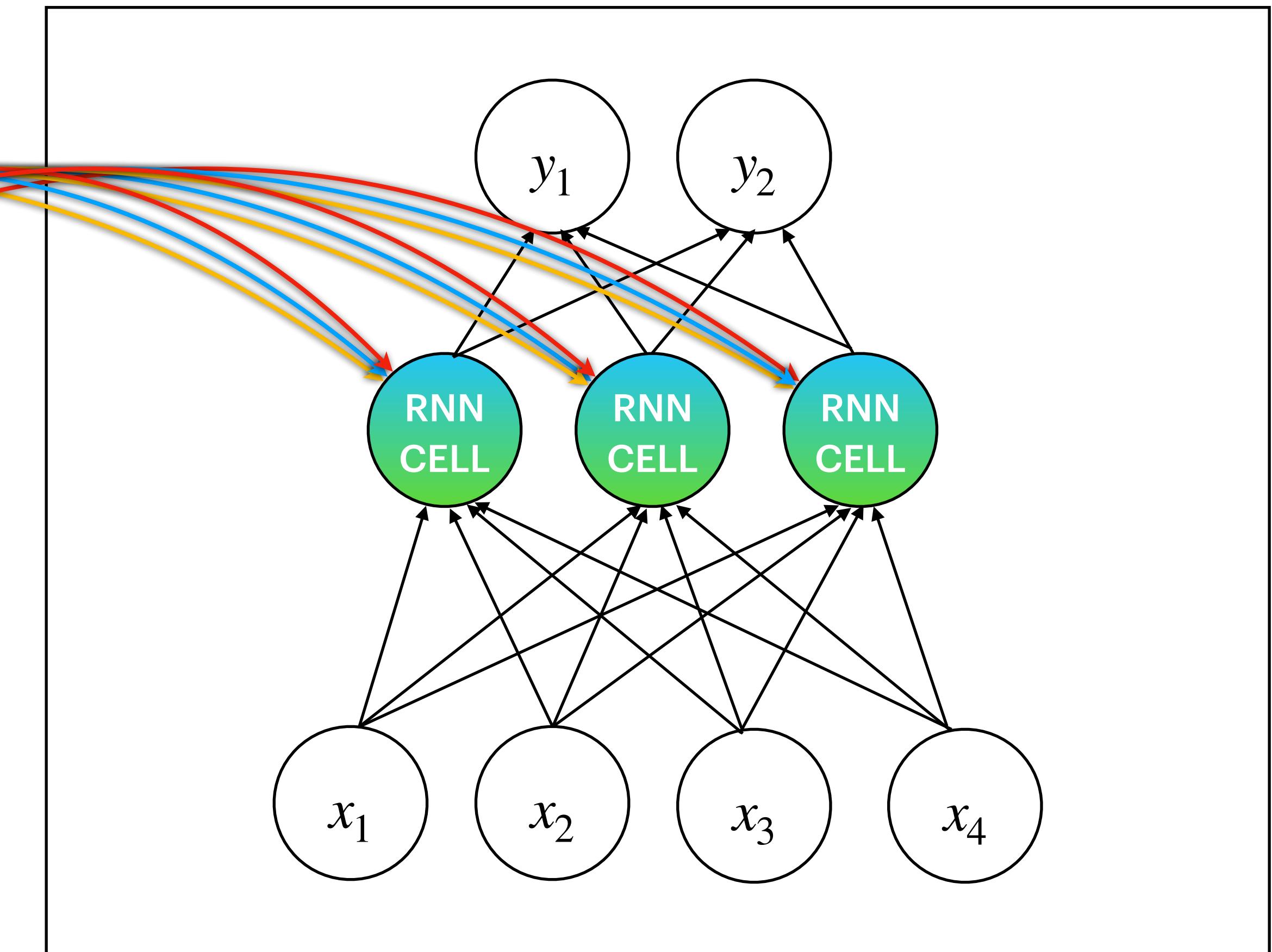
# Background Knowledge

## 2) Recurrent Neural Network

Time Step 1

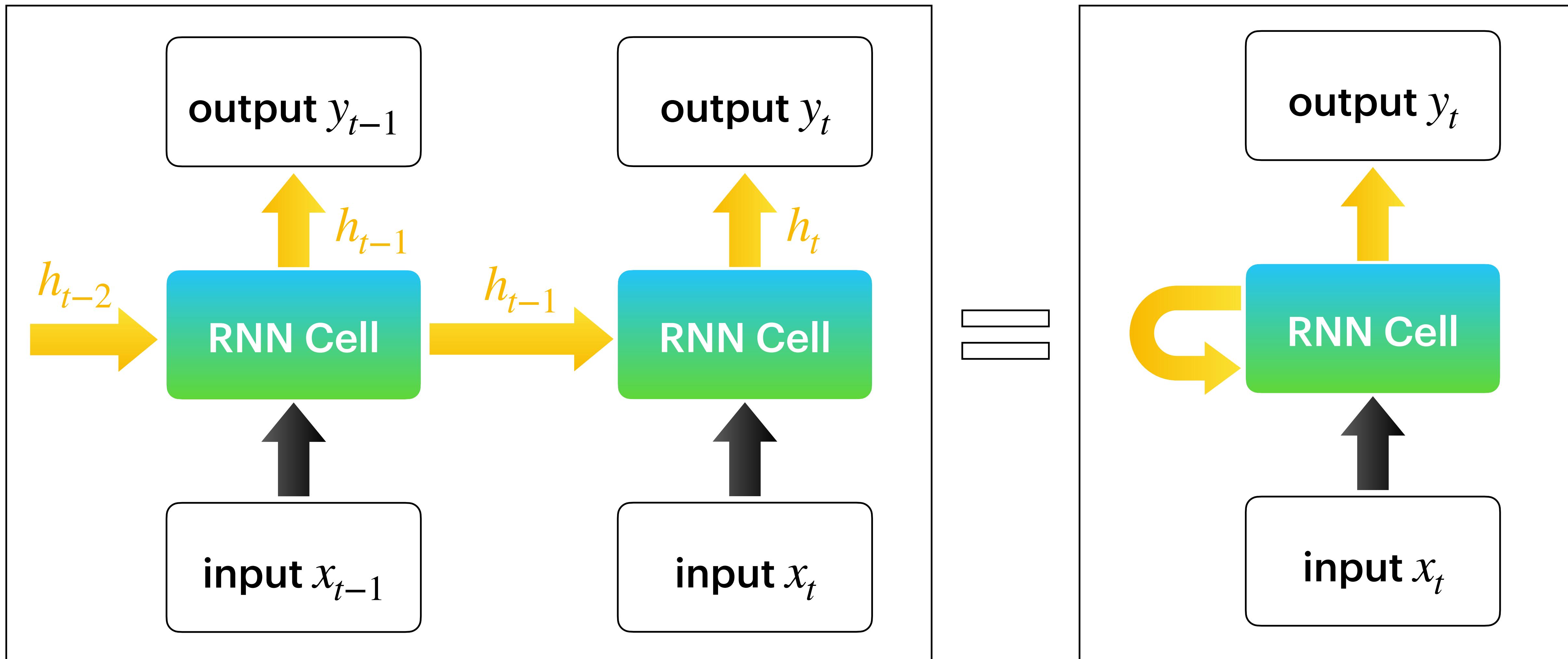


Time Step 2



# Background Knowledge

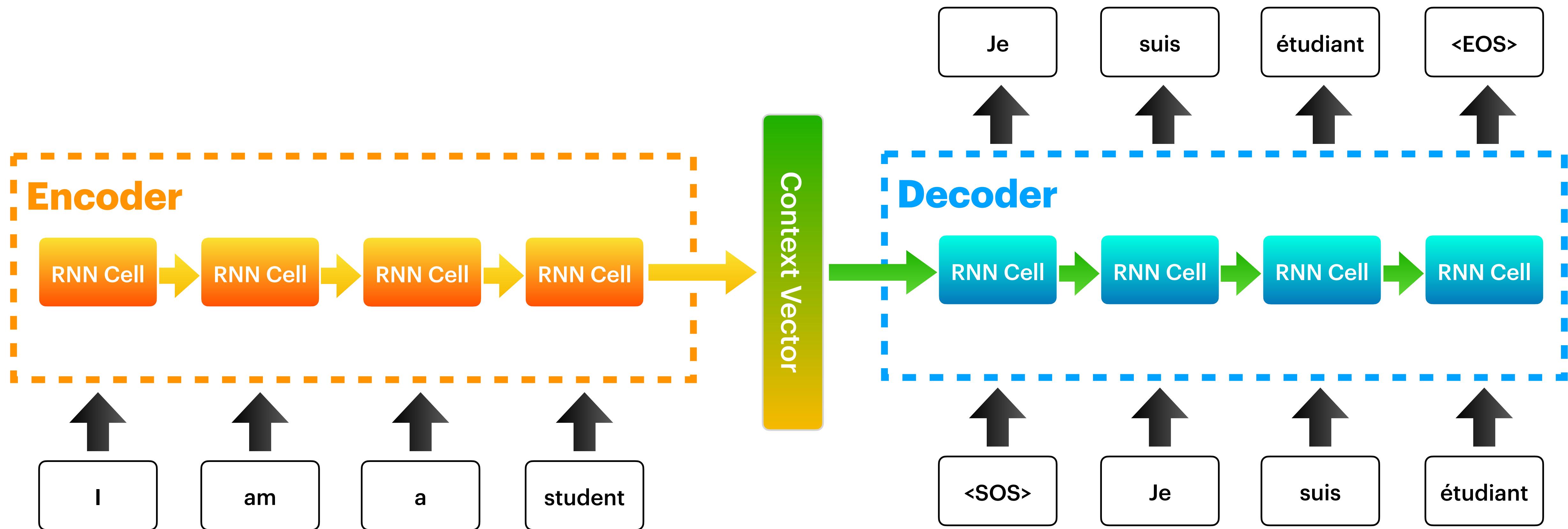
## 2) Recurrent Neural Network



# What Can We Do with RNN?

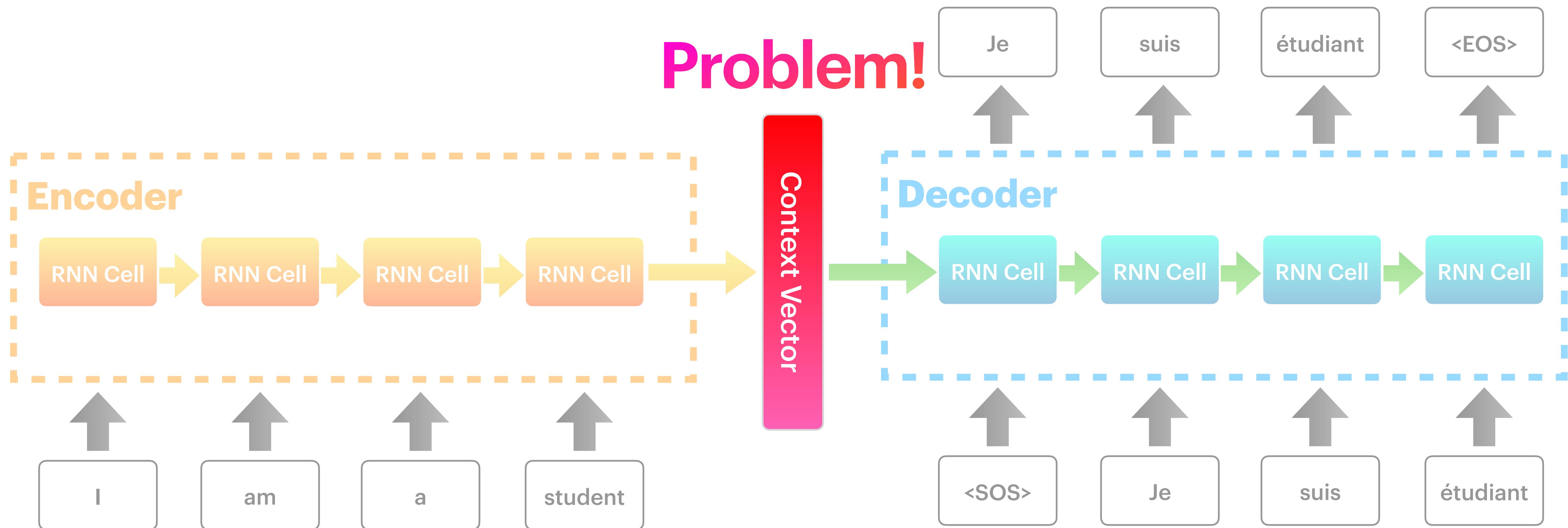
# Background Knowledge

## 3) Sequence to Sequence Learning



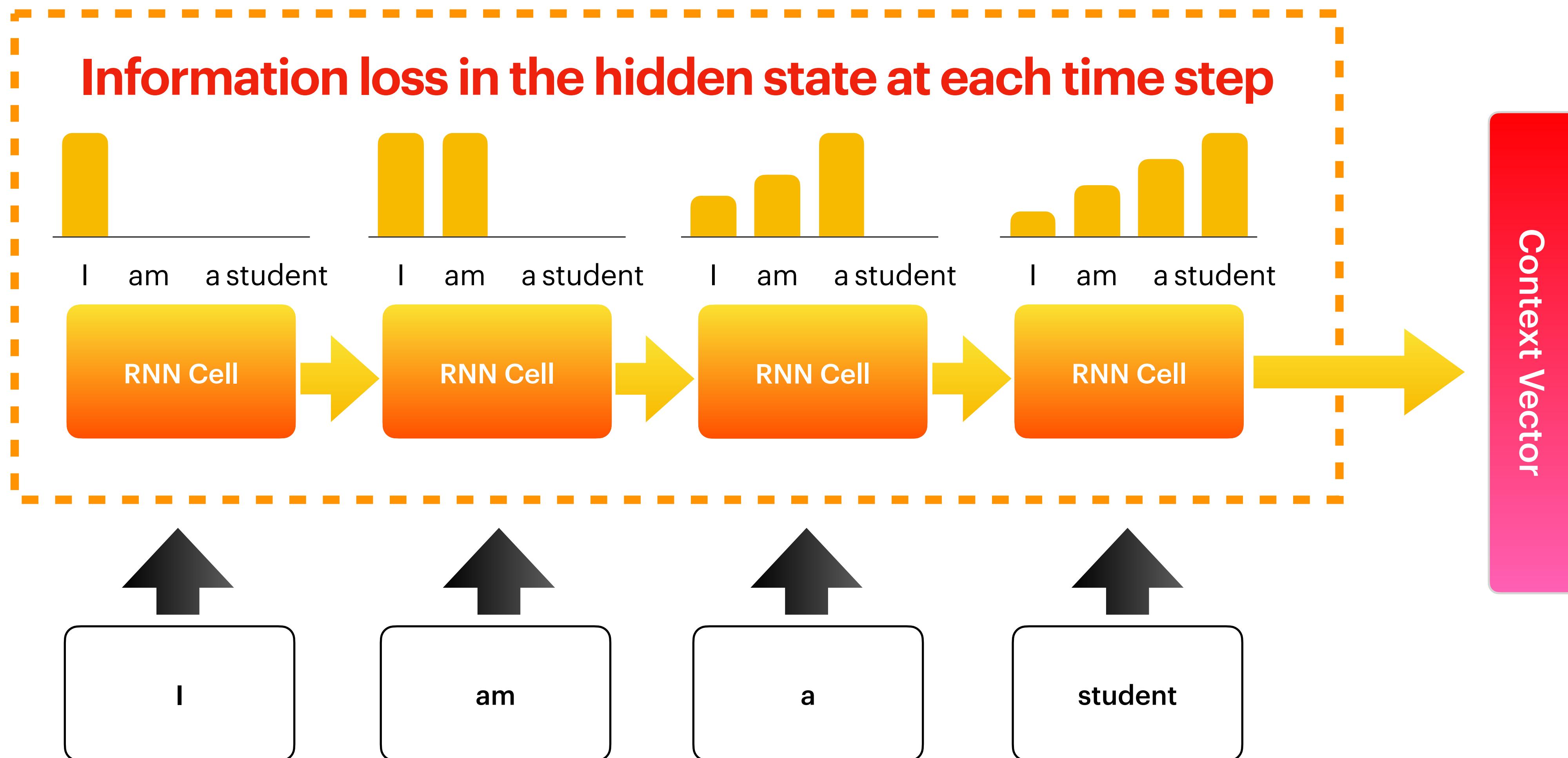
# Background Knowledge

## 3) Sequence to Sequence Learning



# Background Knowledge

## 3) Sequence to Sequence Learning

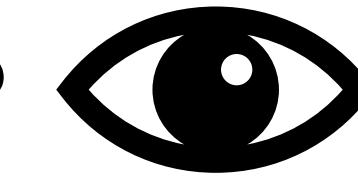


# How Can We Solve This Problem?

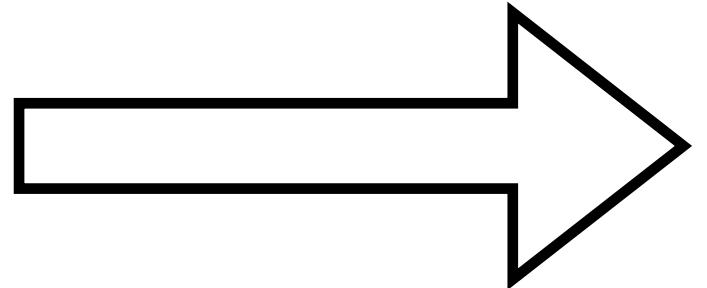
# Background Knowledge

## 4) Attention Mechanism

**How would we translate this sentence?**



**I am a student**

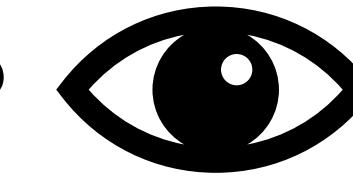


**Je suis étudiant**

# Background Knowledge

## 4) Attention Mechanism

**How would we translate this sentence?**



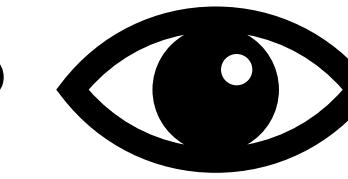
I am a student

Je suis étudiant

# Background Knowledge

## 4) Attention Mechanism

**How would we translate this sentence?**

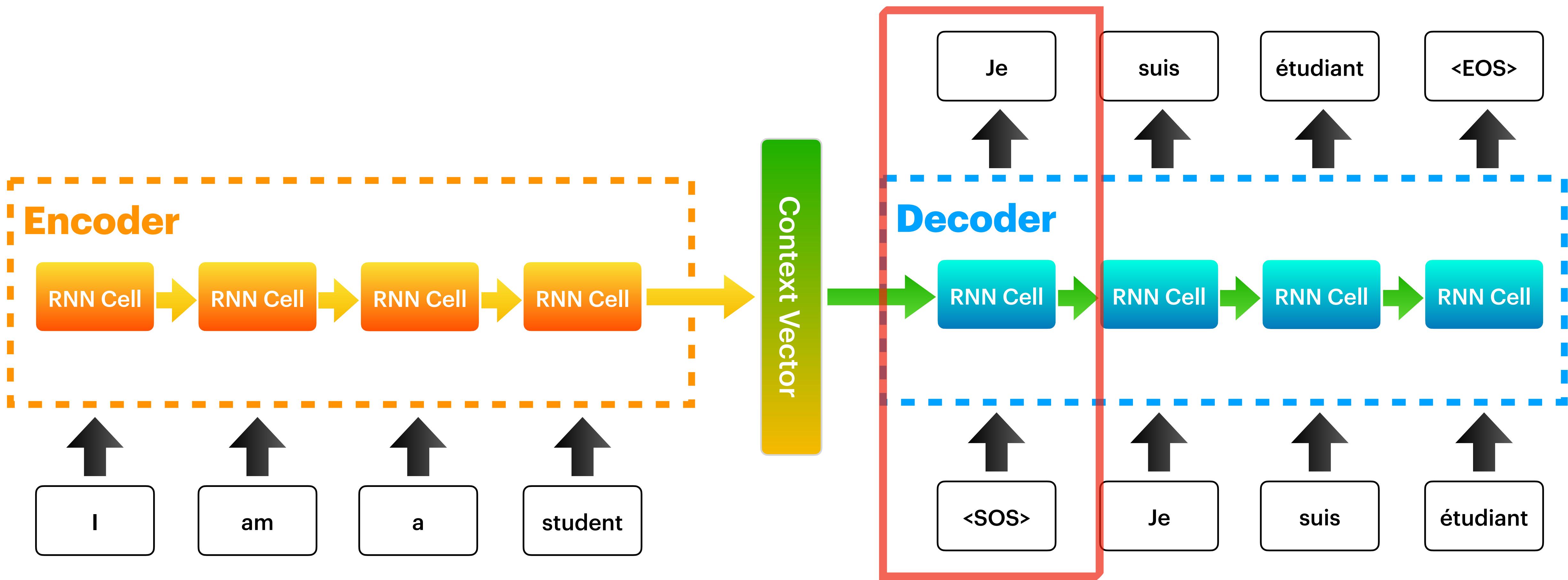


I am a student

Je suis étudiant

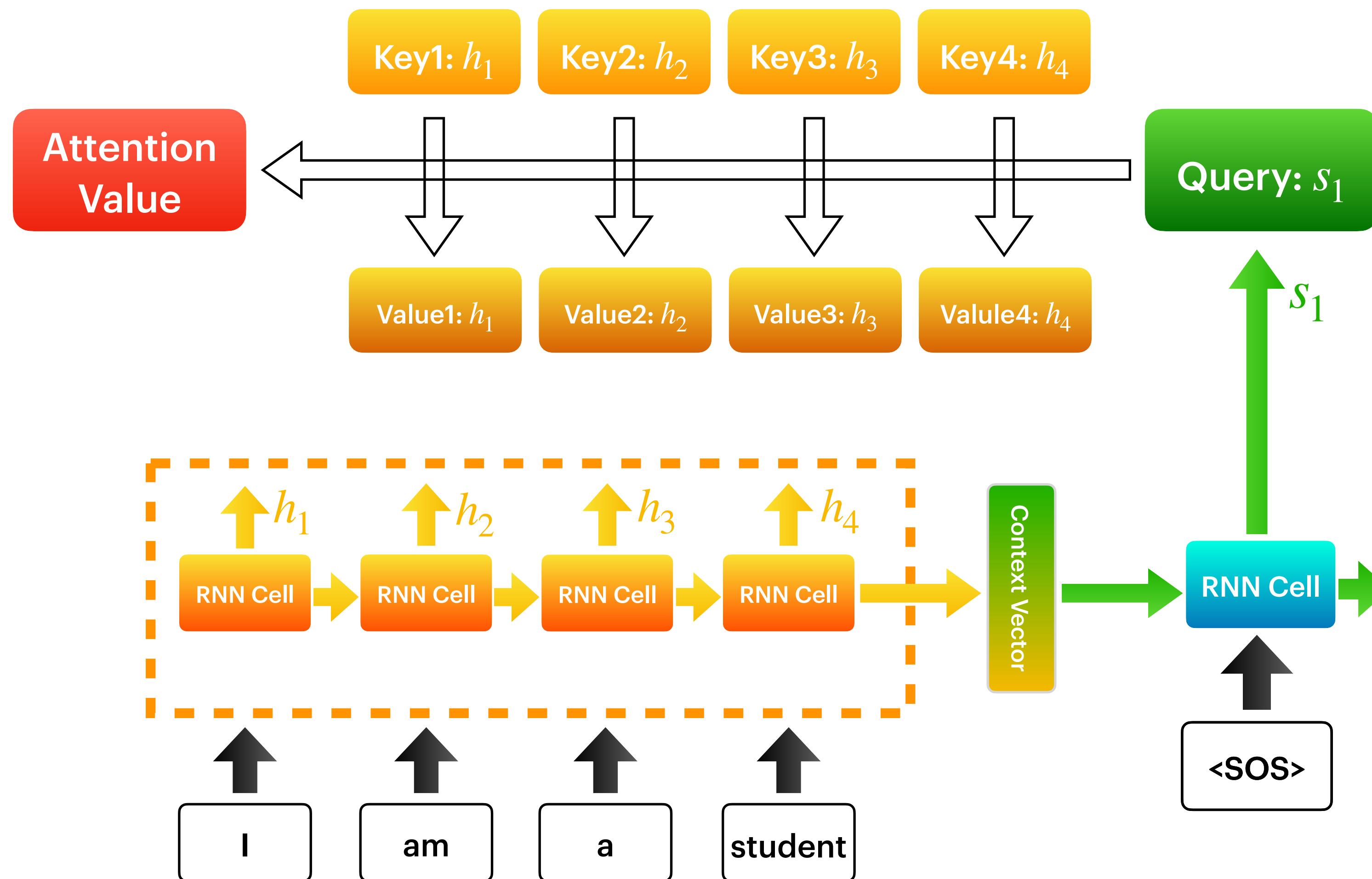
# Background Knowledge

## 4) Attention Mechanism



# Background Knowledge

## 4) Attention Mechanism



# Background Knowledge

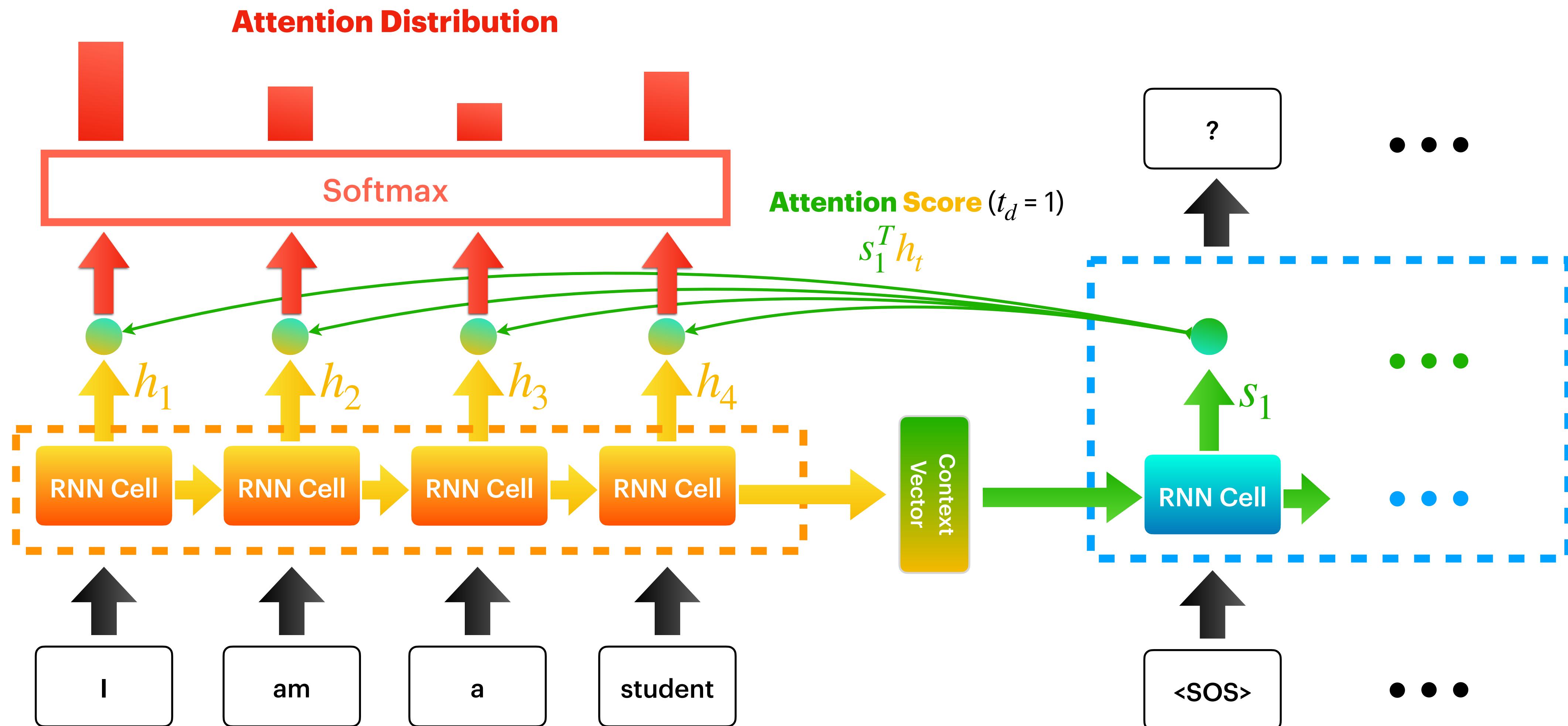
## 4) Attention Mechanism

$$\text{Attention}(Q, K, V) \quad \equiv$$

Attention Value

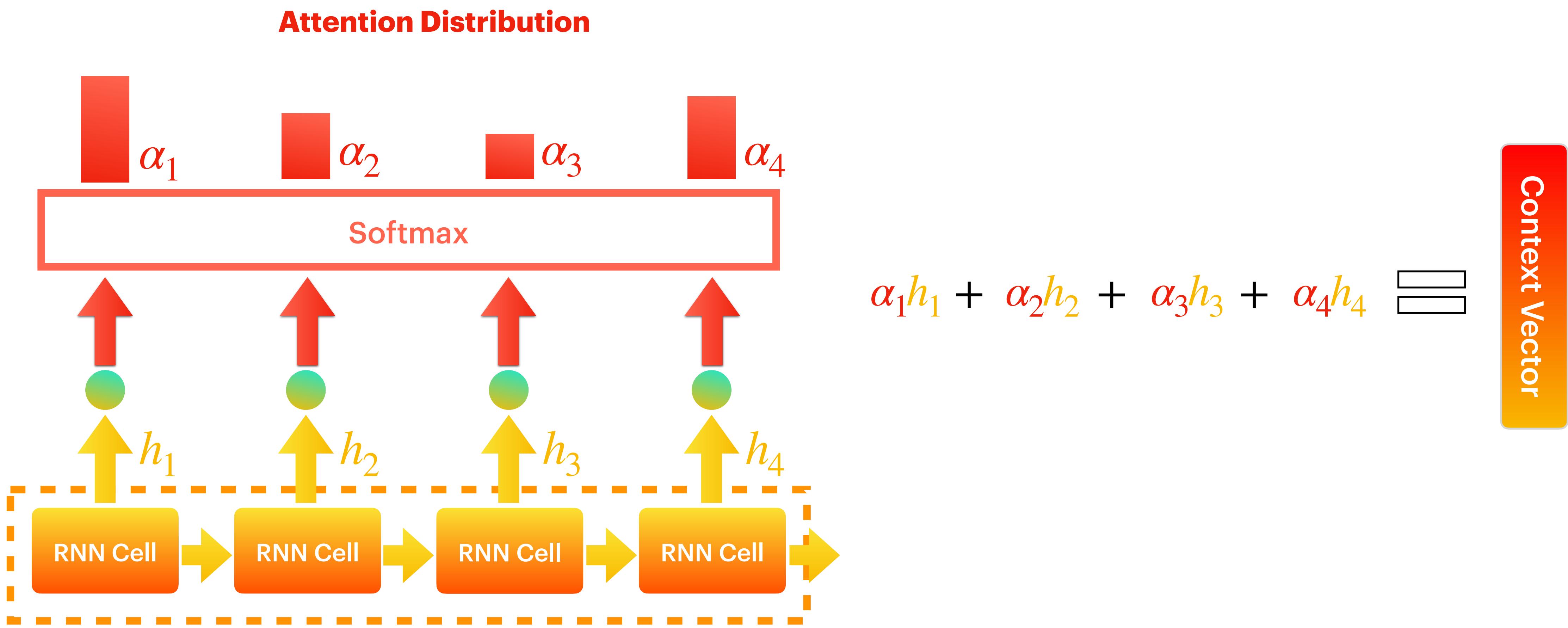
# Background Knowledge

## 4) Attention Mechanism



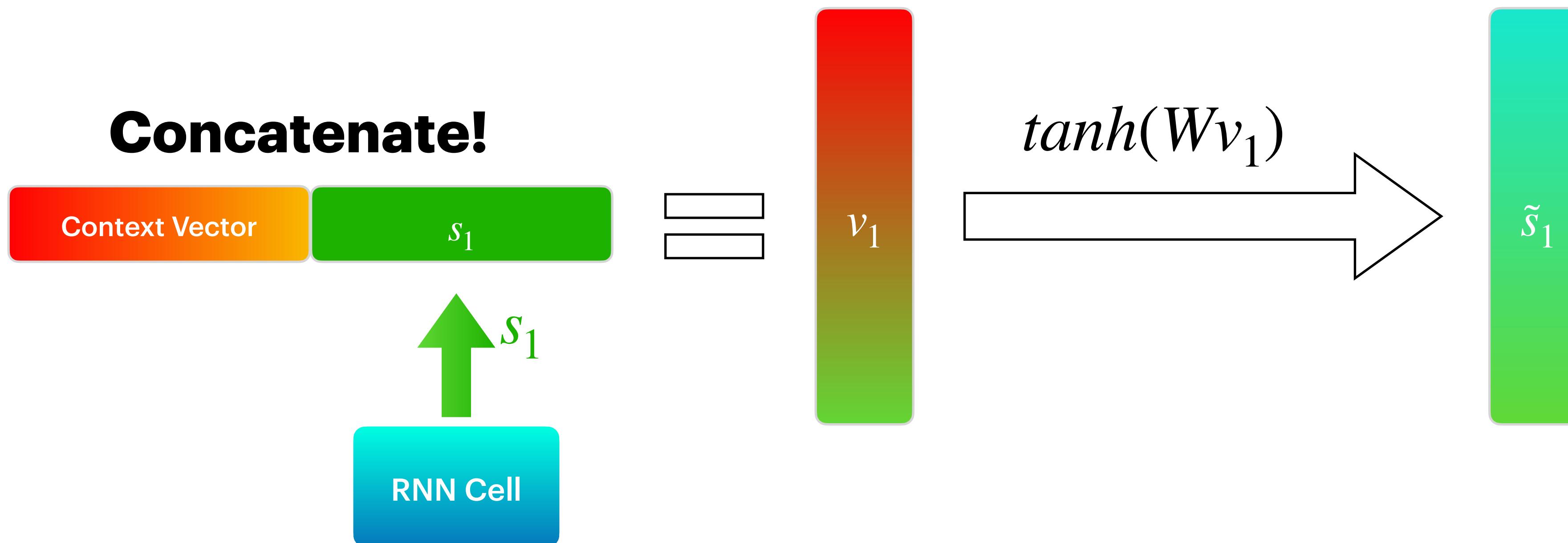
# Background Knowledge

## 4) Attention Mechanism



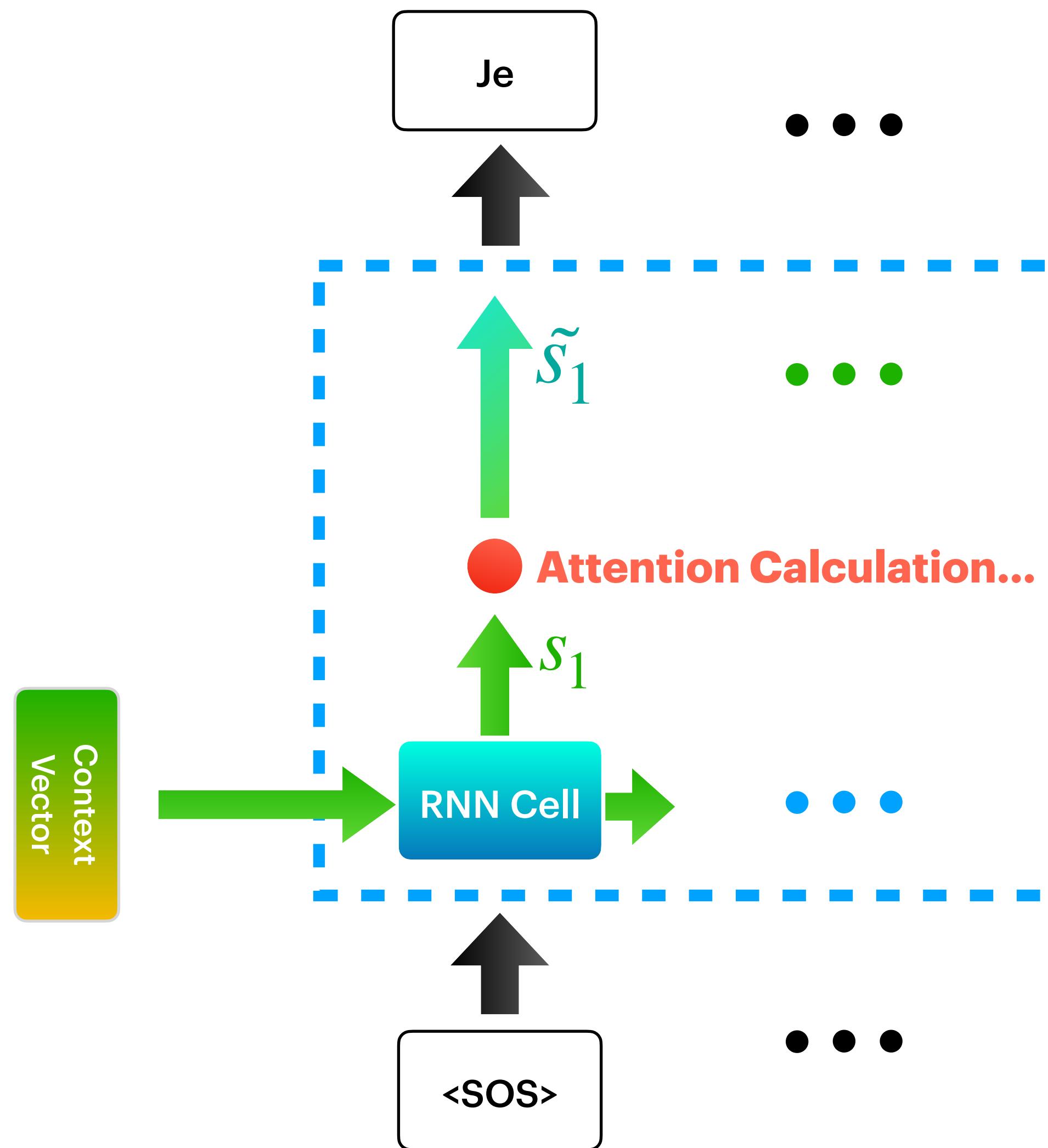
# Background Knowledge

## 4) Attention Mechanism



# Background Knowledge

## 4) Attention Mechanism



# In Short...

- **RNN**
  - Utilizes information from the previous time step
- **Sequence to Sequence**
  - Process its input “sequentially”
  - The context vector tends to lose early input information
- **Attention Mechanism**
  - References every encoder hidden state at each decoding step
  - Calculates relevance of each input word to the decoder’s hidden states



# References

- 1) <https://wikidocs.net/book/2155>
- 2) <https://arxiv.org/abs/1409.3215>

