





# Traditional Language Models



J'ai vu le match de foot



# Traditional Language Models

		Sequence	$P(\text{Sequence})$
  J'ai vu le match de foot 		I saw the game of soccer	4.5 e-5
		<u>I saw the soccer game</u>	<u>6.0 e-5</u>
		I saw the soccer match	4.6 e-5
		Saw I the game of soccer	2.6 e-9

# N-grams

$$P(w_2|w_1) = \frac{\text{count}(w_1, w_2)}{\text{count}(w_1)} \longrightarrow \text{Bigrams}$$

$$P(w_3|w_1, w_2) = \frac{\text{count}(w_1, w_2, w_3)}{\text{count}(w_1, w_2)} \longrightarrow \text{Trigrams}$$

# N-grams

$$P(w_2|w_1) = \frac{\text{count}(w_1, w_2)}{\text{count}(w_1)} \longrightarrow \text{Bigrams}$$

$$P(w_3|w_1, w_2) = \frac{\text{count}(w_1, w_2, w_3)}{\text{count}(w_1, w_2)} \longrightarrow \text{Trigrams}$$

$$P(w_1, w_2, w_3) = P(w_1) \times P(w_2|w_1) \times P(w_3|w_2)$$

- Large N-grams to capture dependencies between distant words
- Need a lot of space and RAM

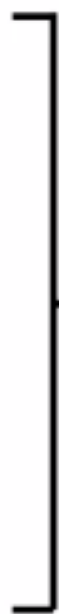
# Advantages of RNNs

Nour was supposed to study with me. I called her but she did not \_\_\_\_\_

# Advantages of RNNs

Nour was supposed to study with me. I called her but she did not have

want  
respond  
choose  
want  
have  
ask  
attempt  
answer  
know



Similar probabilities with trigram

# Advantages of RNNs

Nour was supposed to study with me. I called her but she did not answer

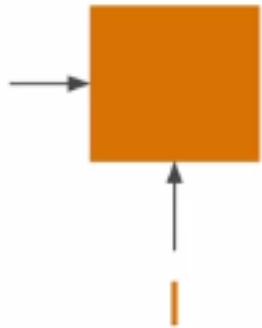
want  
respond  
choose  
want  
have  
ask  
attempt  
answer  
know

RNNs look at every previous word

Similar probabilities with trigram

# RNNs Basic Structure

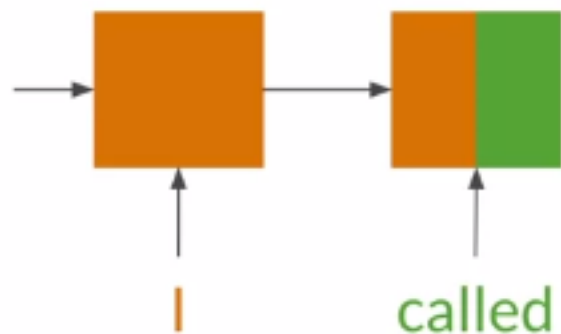
I called her but she did not \_\_\_\_\_





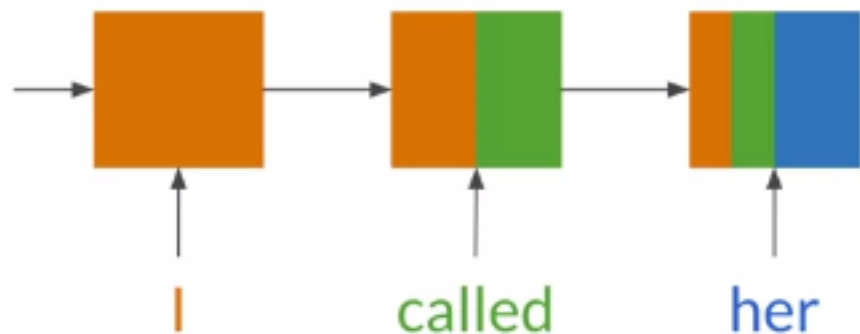
# RNNs Basic Structure

I called her but she did not \_\_\_\_\_



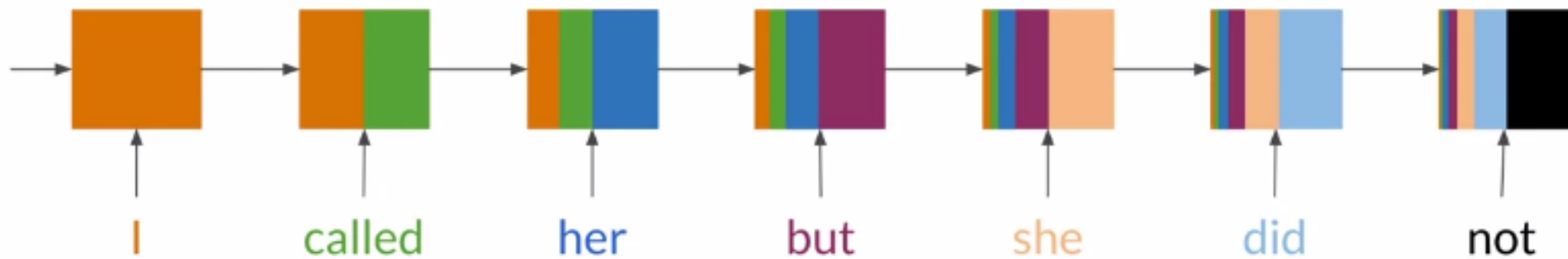
# RNNs Basic Structure

I called her but she did not \_\_\_\_\_



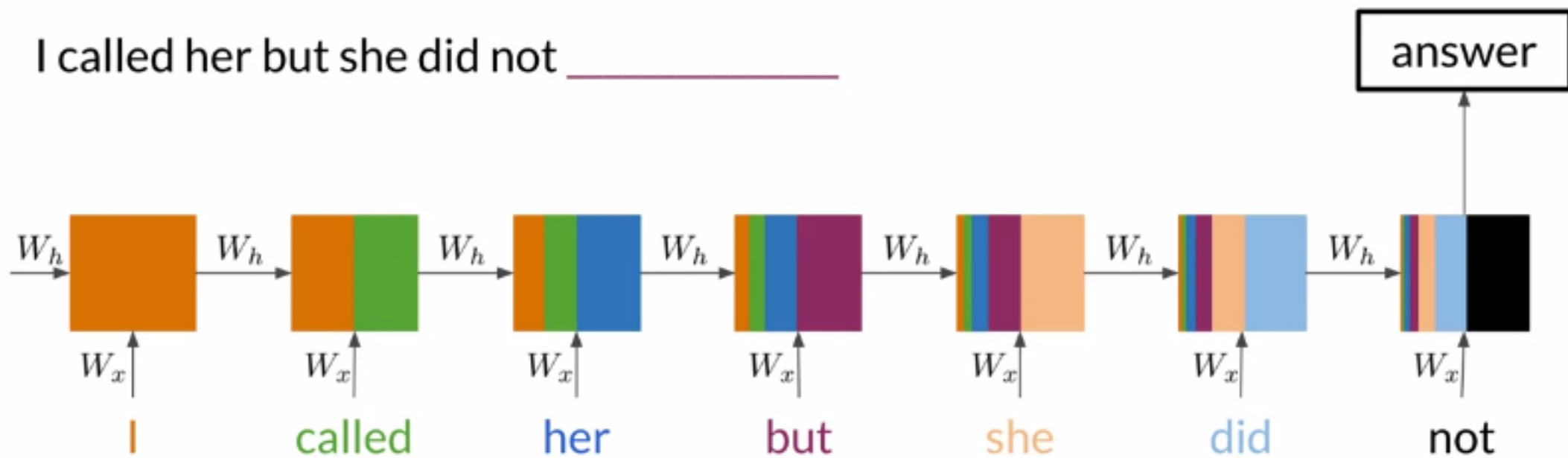
# RNNs Basic Structure

I called her but she did not \_\_\_\_\_



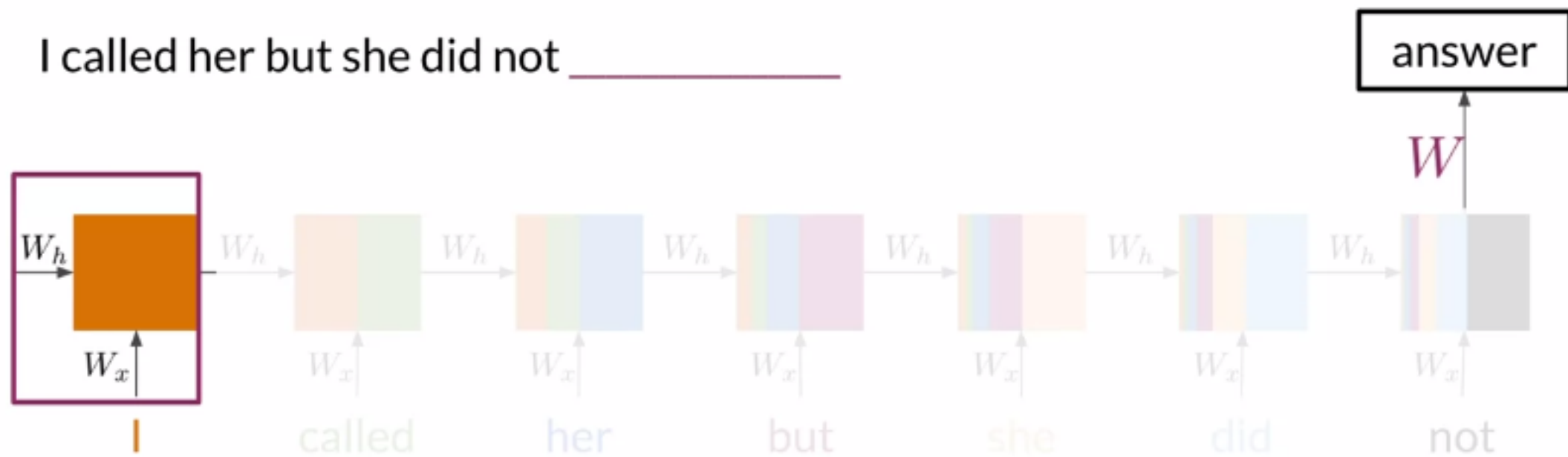
# RNNs Basic Structure

I called her but she did not \_\_\_\_\_



# RNNs Basic Structure

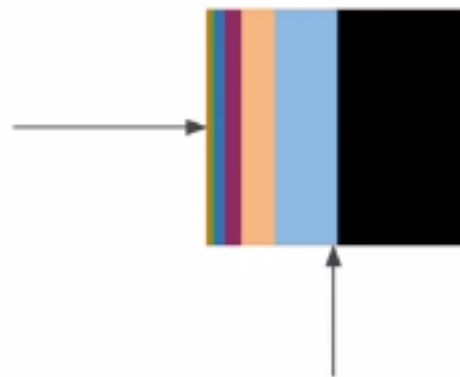
I called her but she did not \_\_\_\_\_



Learnable parameters

# Summary

- RNNs model relationships among distant words
- In RNNs a lot of computations share parameters

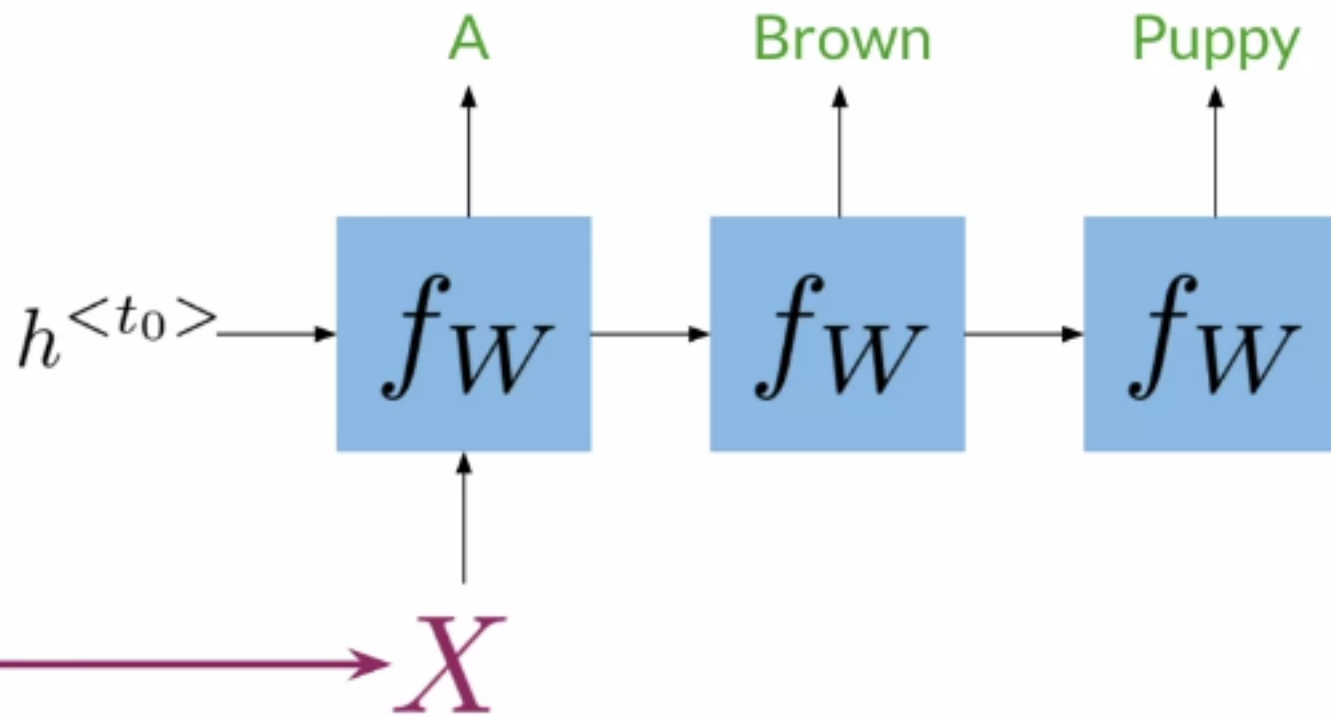


# One to One



# One to Many

Caption  
generation





# Many to One

Sentiment  
analysis

Positive

Tweet:

I

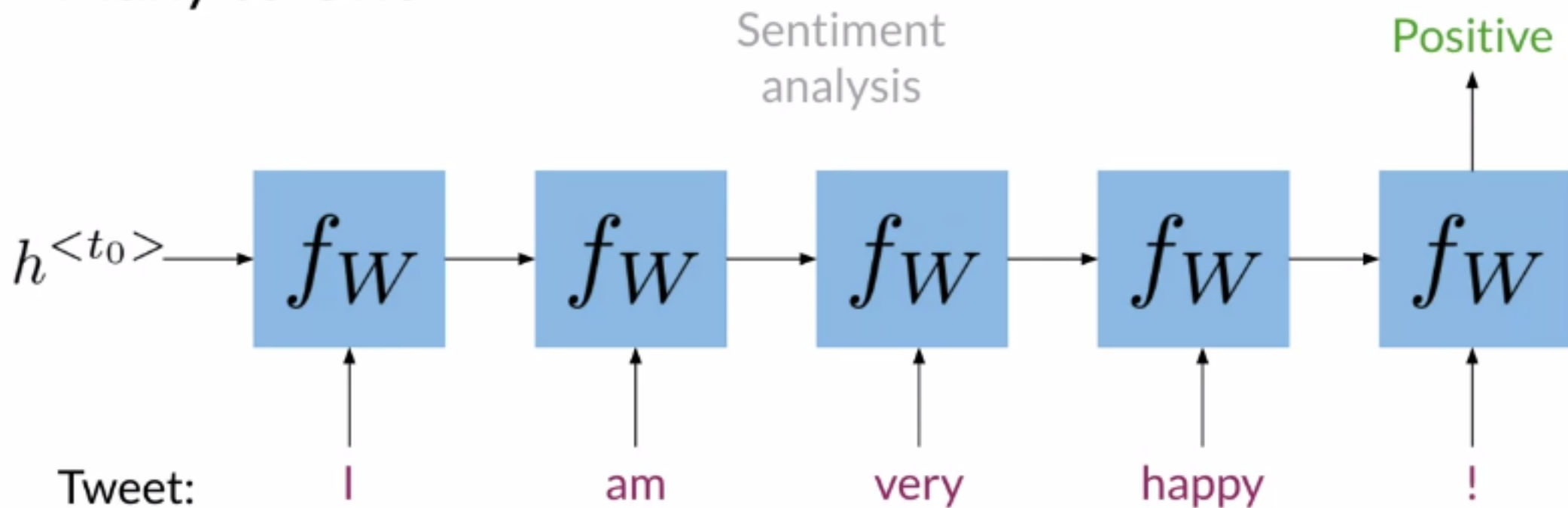
am

very

happy

!

## Many to One



## Many to Many



I

am

hungry

J'

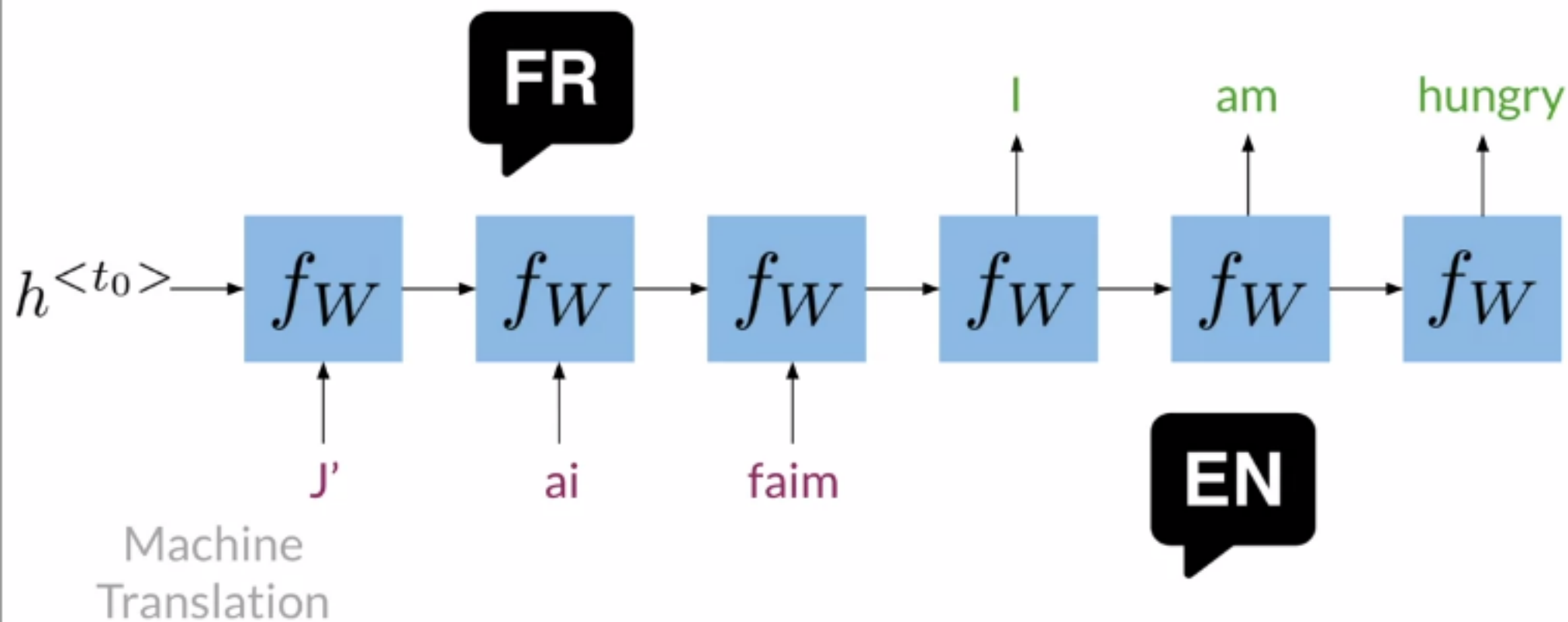
ai

faim

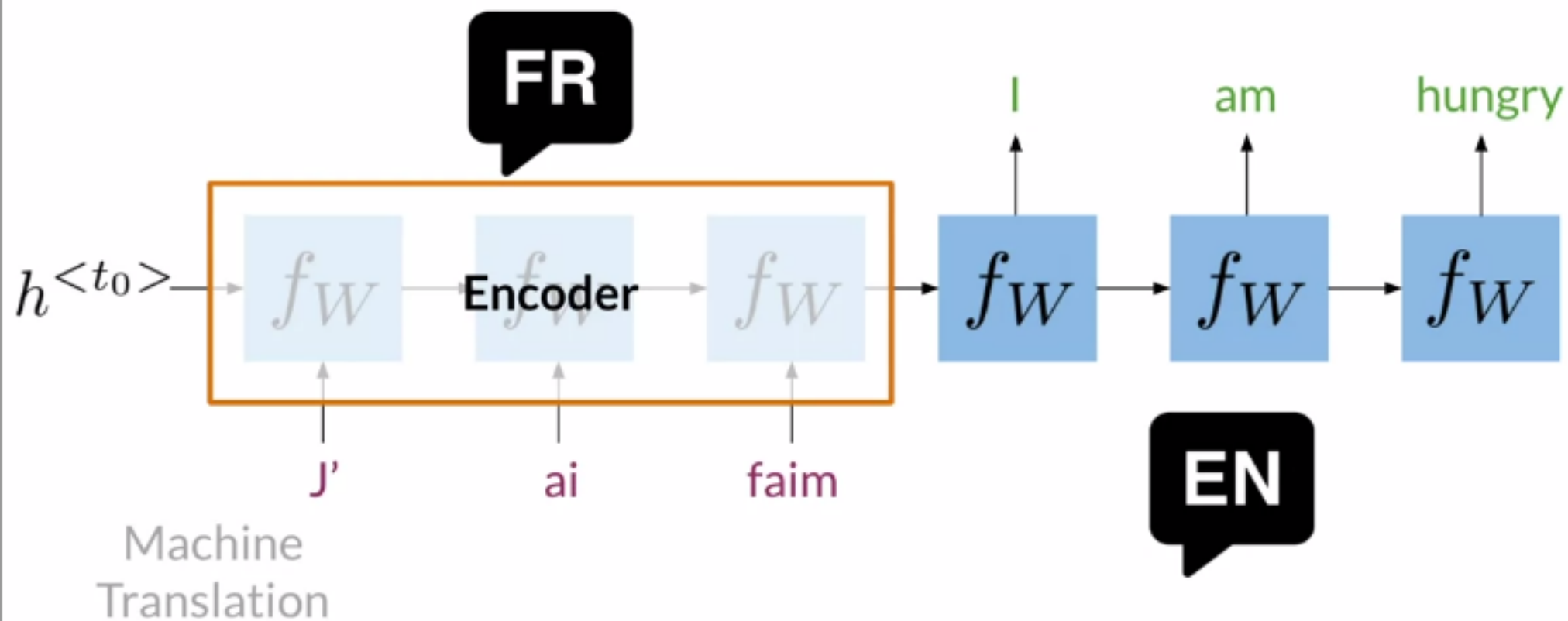


Machine  
Translation

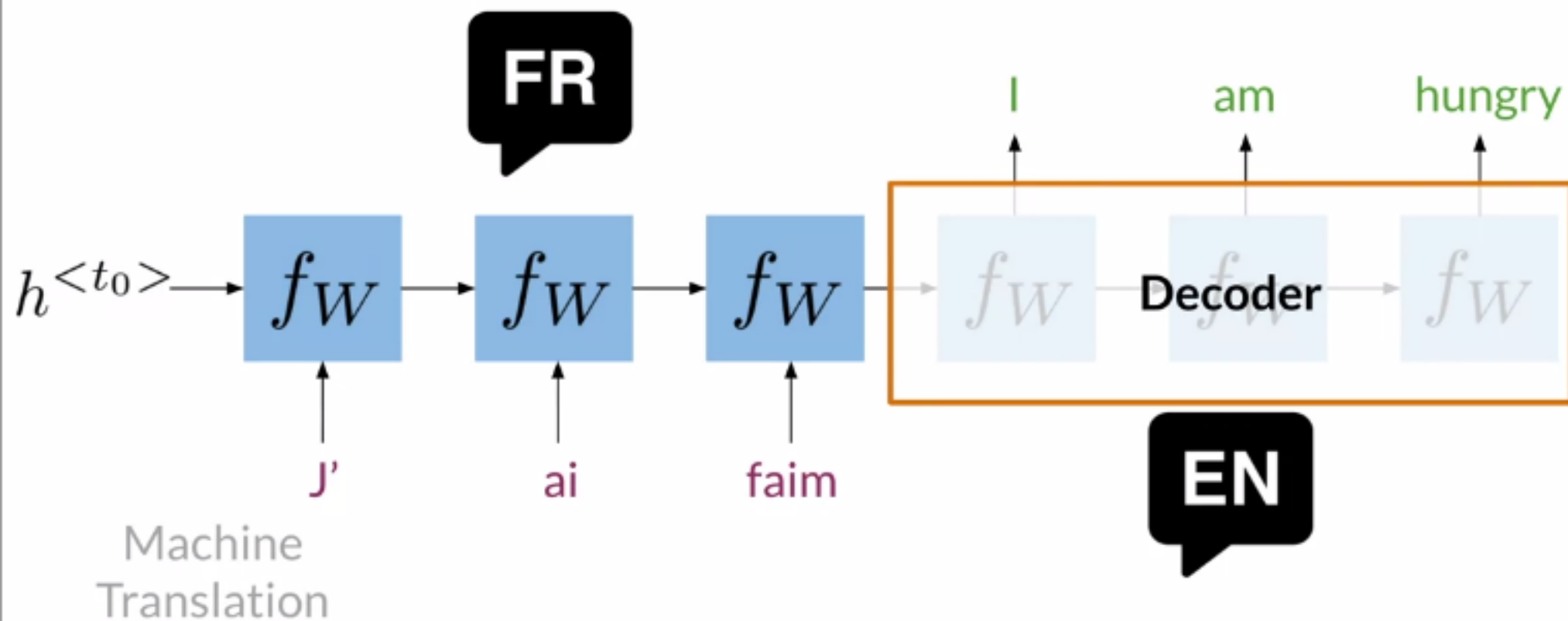
## Many to Many



## Many to Many



## Many to Many

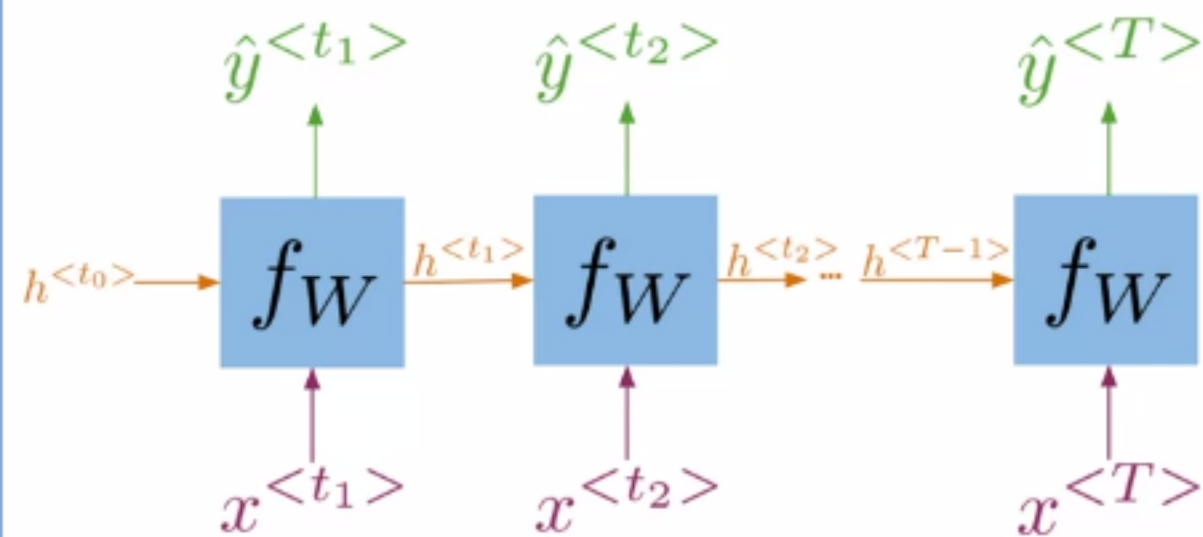


# Summary

- RNNs can be implemented for a variety of NLP tasks
- Applications include Machine translation and caption generation

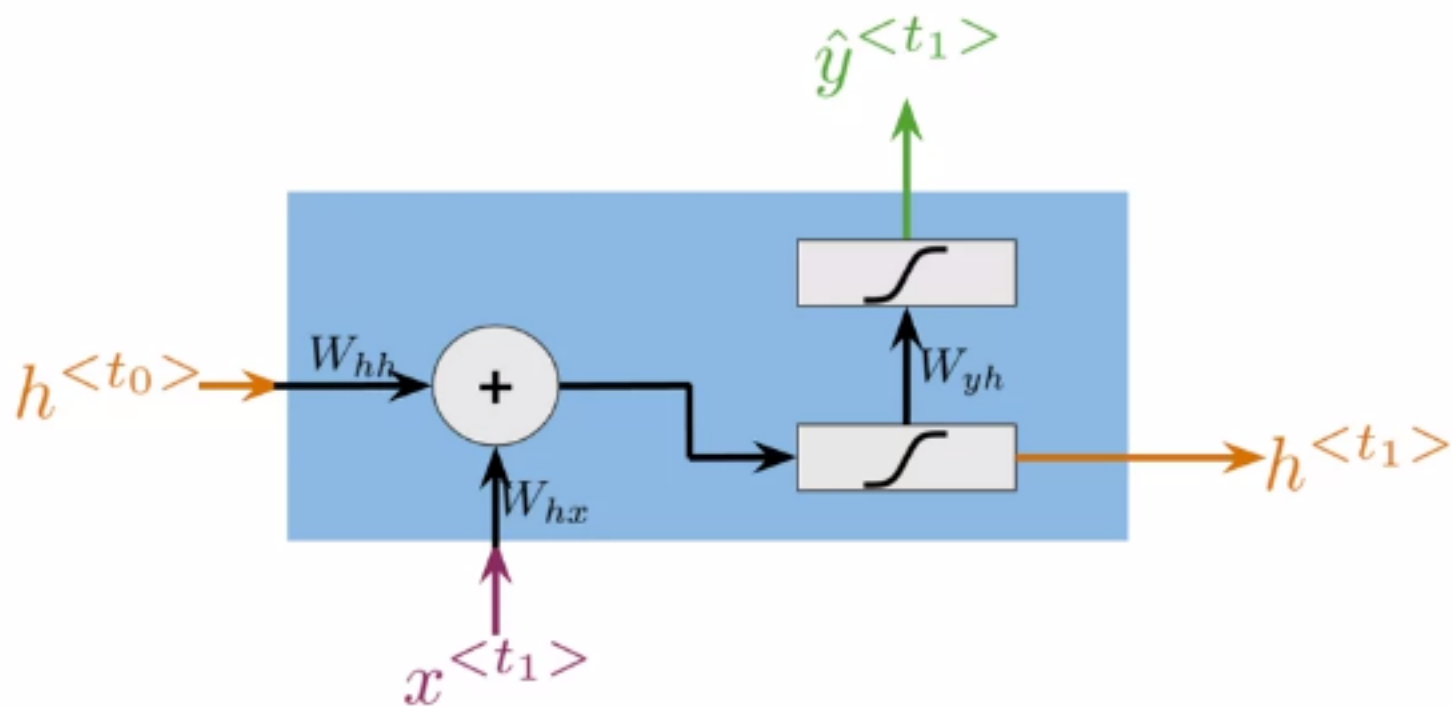
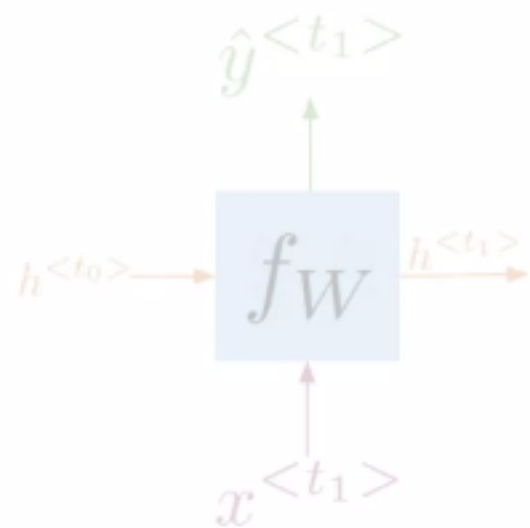


# A Vanilla RNN





# A Vanilla RNN



$$h^{<t>} = g(W_{hh}h^{<t-1>} \oplus W_{hx}x^{<t>} + b_h)$$

$$\hat{y}^{<t>} = g(W_{yh}h^{<t>} + b_y)$$

# Summary

- Hidden states propagate information through time
- Basic recurrent units have two inputs at each time:  $h^{<t-1>}$ ,  $x^{<t>}$

