

# INF721

2023/2



# Aprendizado em Redes Neurais Profundas

## A20: Atenção

# Logística

## Avisos

- ▶ Entrega da PF: Proposta de Problema nesta quarta-feira (18/10)!

## Última aula

- ▶ Estudo de casos de CNNs
- ▶ CNNs clássicas (LeNet-5, AlexNet, VGG-16)
- ▶ ResNet
- ▶ Inception Network

# Plano de Aula

- ▶ Tradução Automática
  - ▶ Formalização
  - ▶ Decodificação
    - ▶ Amostragem
    - ▶ Busca Gulosa
    - ▶ Beam Search
- ▶ Modelo de Atenção em RNNs

# Tradução Automática

## Conjunto de dados

Pares de sentenças de um idioma origem (x) para um idioma destino (y)

### Inglês

### Português

---

See you!

Nos vemos!

---

The book is on the table.

O livro está em cima da mesa.

---

Lucas is visiting Chile in January.

Lucas irá visitar o Chile em Janeiro.

---

Lucas is visiting Chile in January.

Em Janeiro, Lucas irá visitar o Chile.

---

....

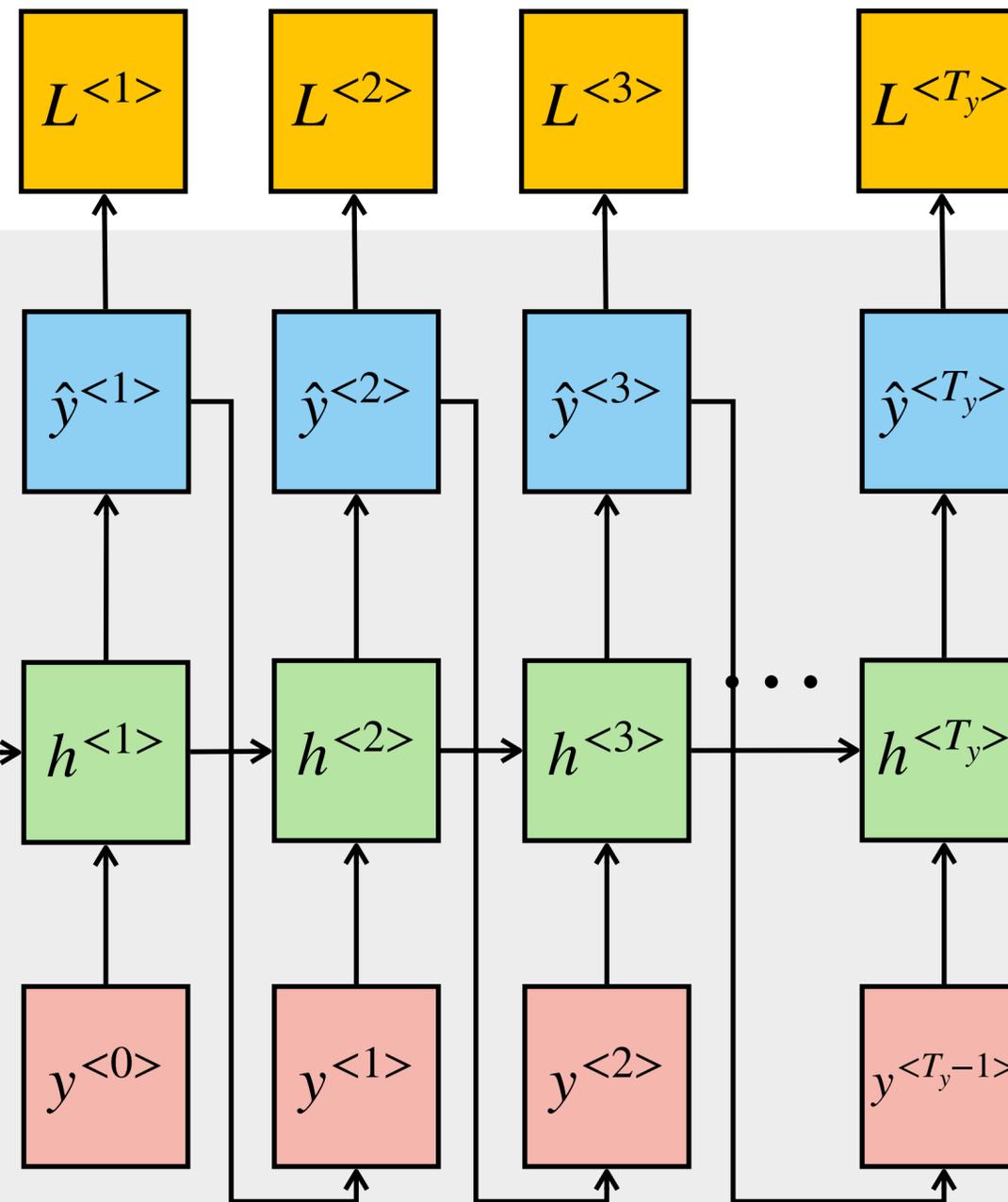
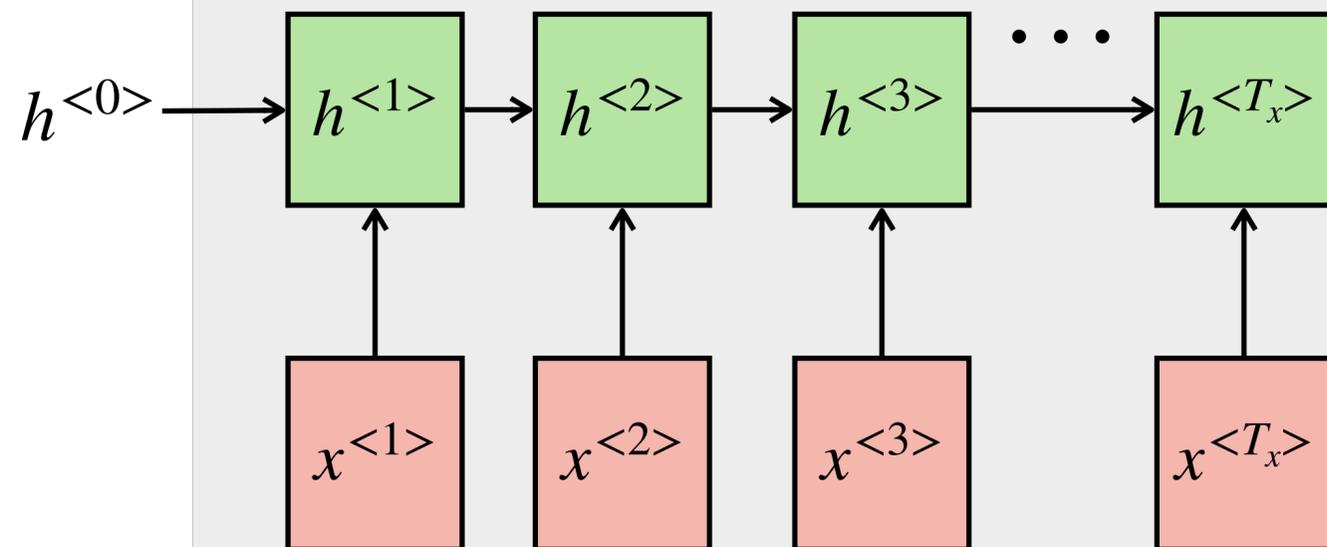
....

---

# Sequência para Sequência

Encoder

$$W_h^{[E]} \quad W_x^{[E]}$$



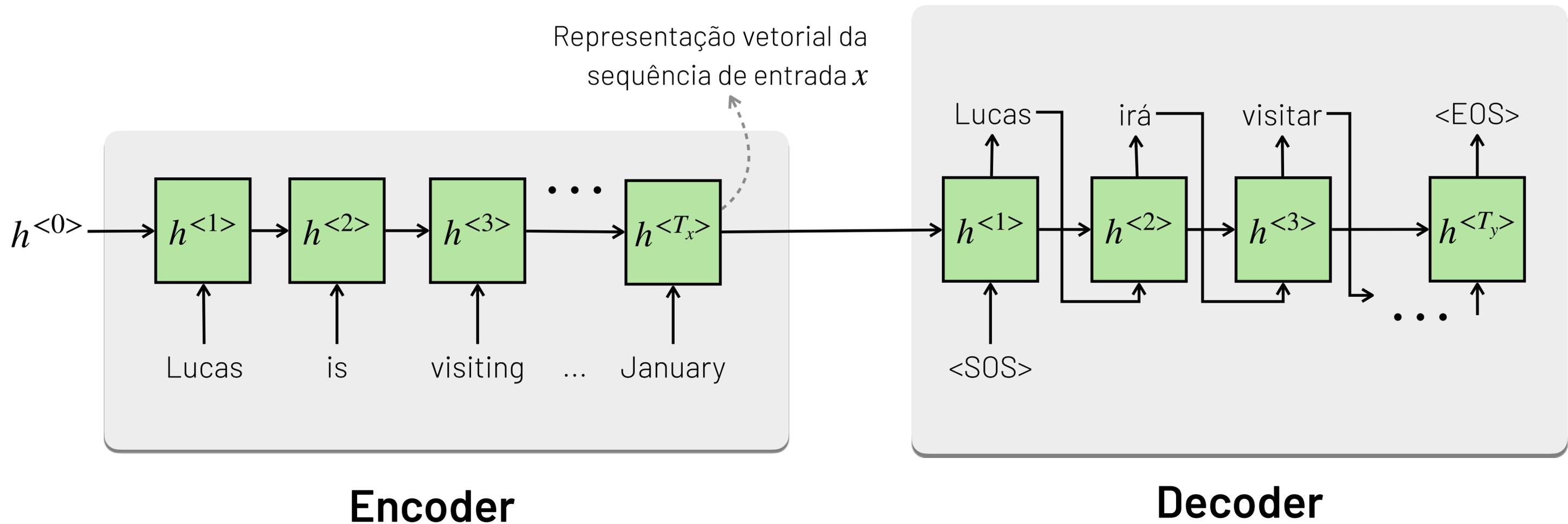
Decoder

$$L = \sum_{t=1}^{T_y} L^{<t>}$$

$$W_y^{[D]} \\ W_h^{[D]} \quad W_x^{[D]}$$

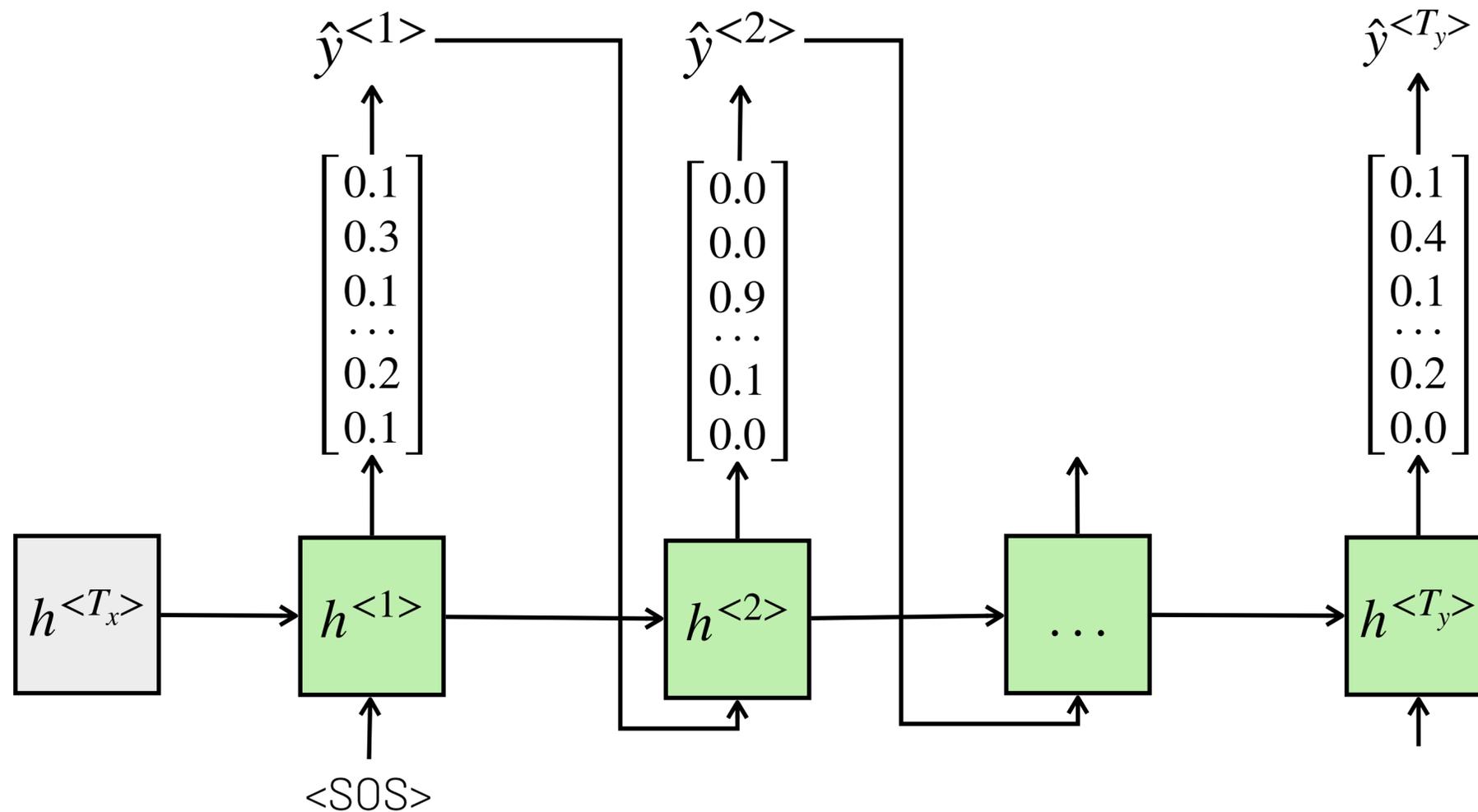
# Tradução Automática

Formalmente, um modelo seq2seq estima:  $P(y^{<1>}, \dots, y^{<T_y>} | x^{<1>}, \dots, x^{<T_x>})$



# Decodificação

Procedimento para gerar elemento-a-elemento uma sequência  $\hat{y}$  a partir do estado escondido  $h^{<T_x>}$  que maximize  $P(y^{<1>}, \dots, y^{<T_y>} | x)$ .

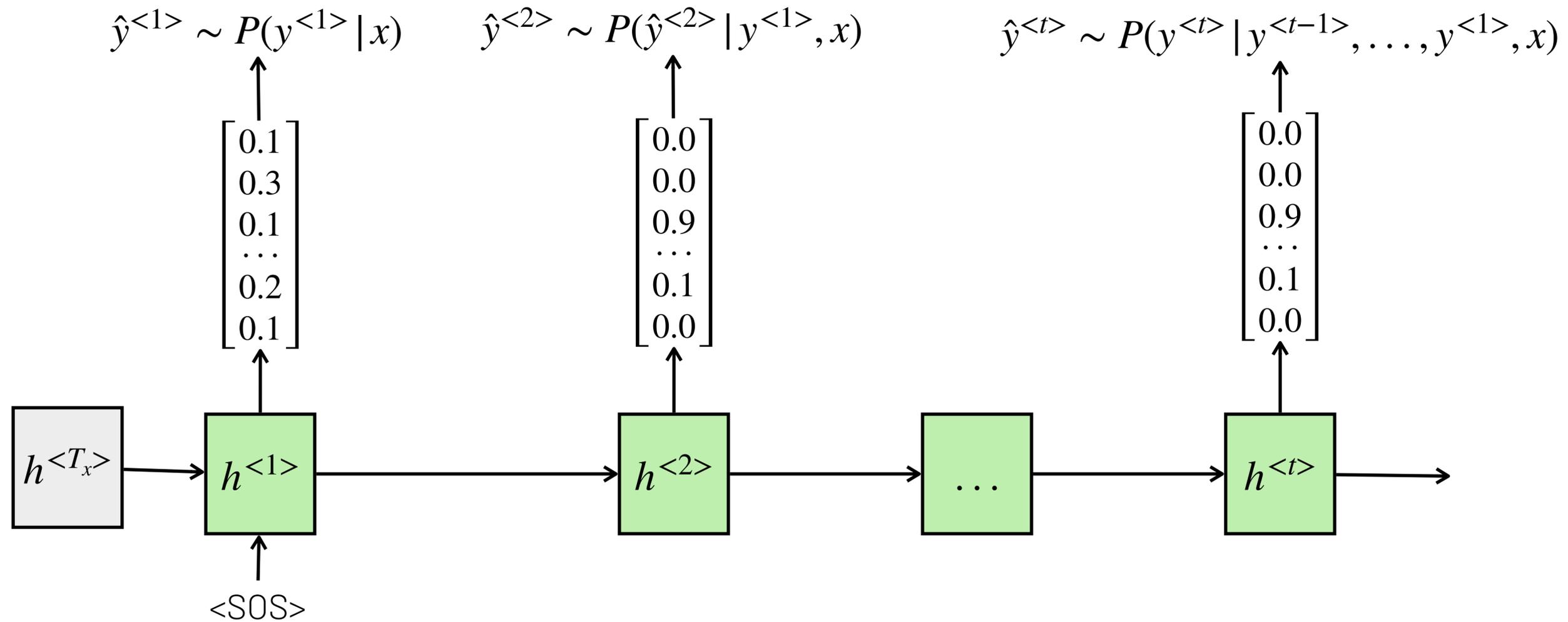


Técnicas mais comuns:

- ▶ **Amostragem**
- ▶ **Busca Gulosa**
- ▶ **Beam Search**

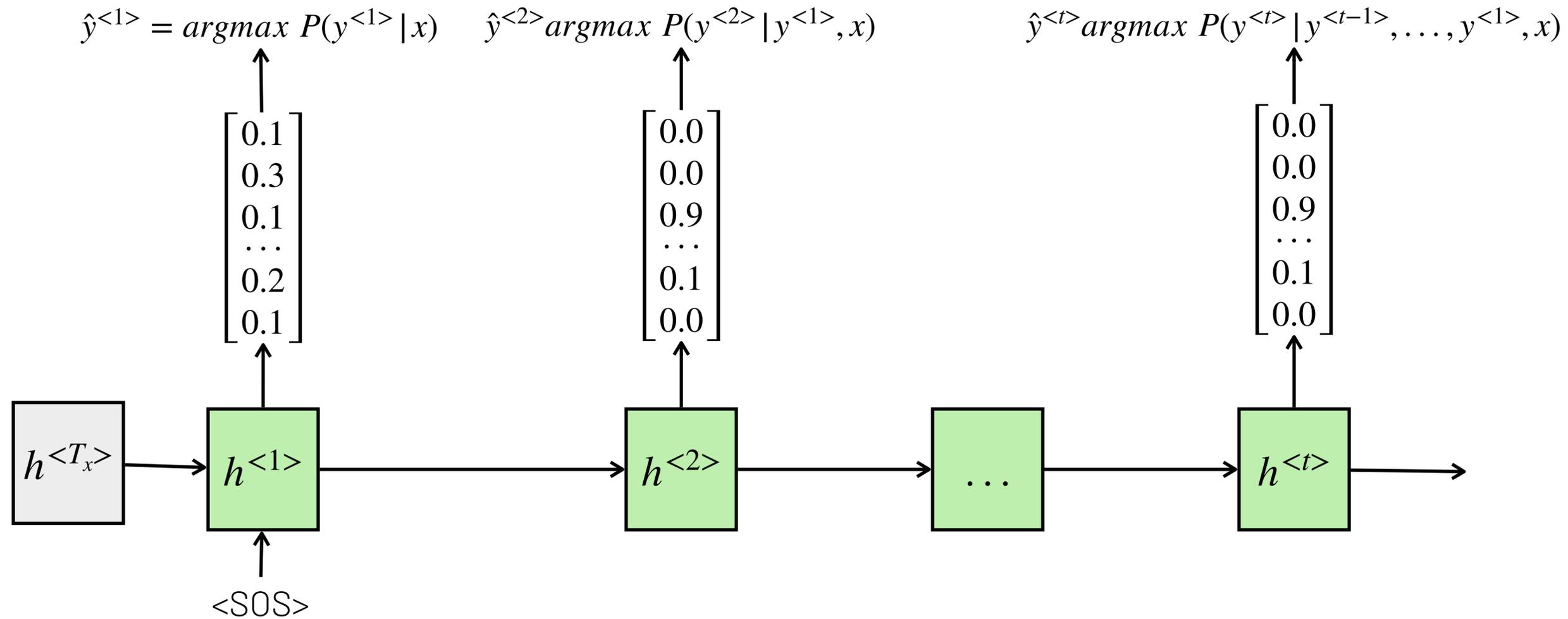
# Amostragem

Amostrar palavra  $\hat{y}^{<t>}$  de acordo com a distribuição  $P(y^{<t>} | x, y^{<t-1>}, \dots, y^{<1>})$



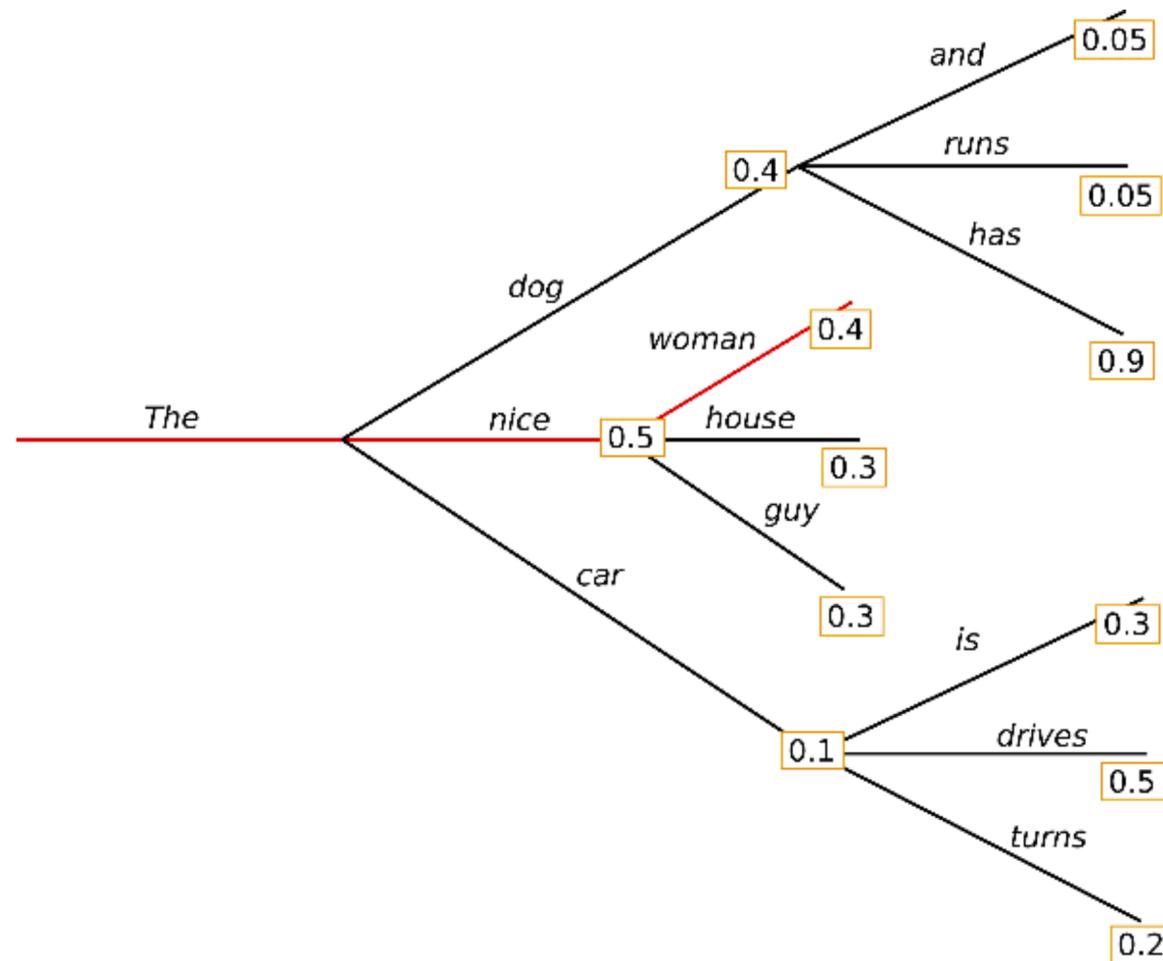
# Busca Gulosa

Selecionar a palavra  $\hat{y}^{<t>}$  com maior probabilidade em  $P(y^{<t>} | x, y^{<t-1>}, \dots, y^{<1>})$



# Busca Gulosa

Começando com a palavra "The", a busca gulosa seleciona a palavra "nice" seguida de "woman", gerando a sentença "The nice woman" com probabilidade  $0.5 \times 0.4 = 0.2$

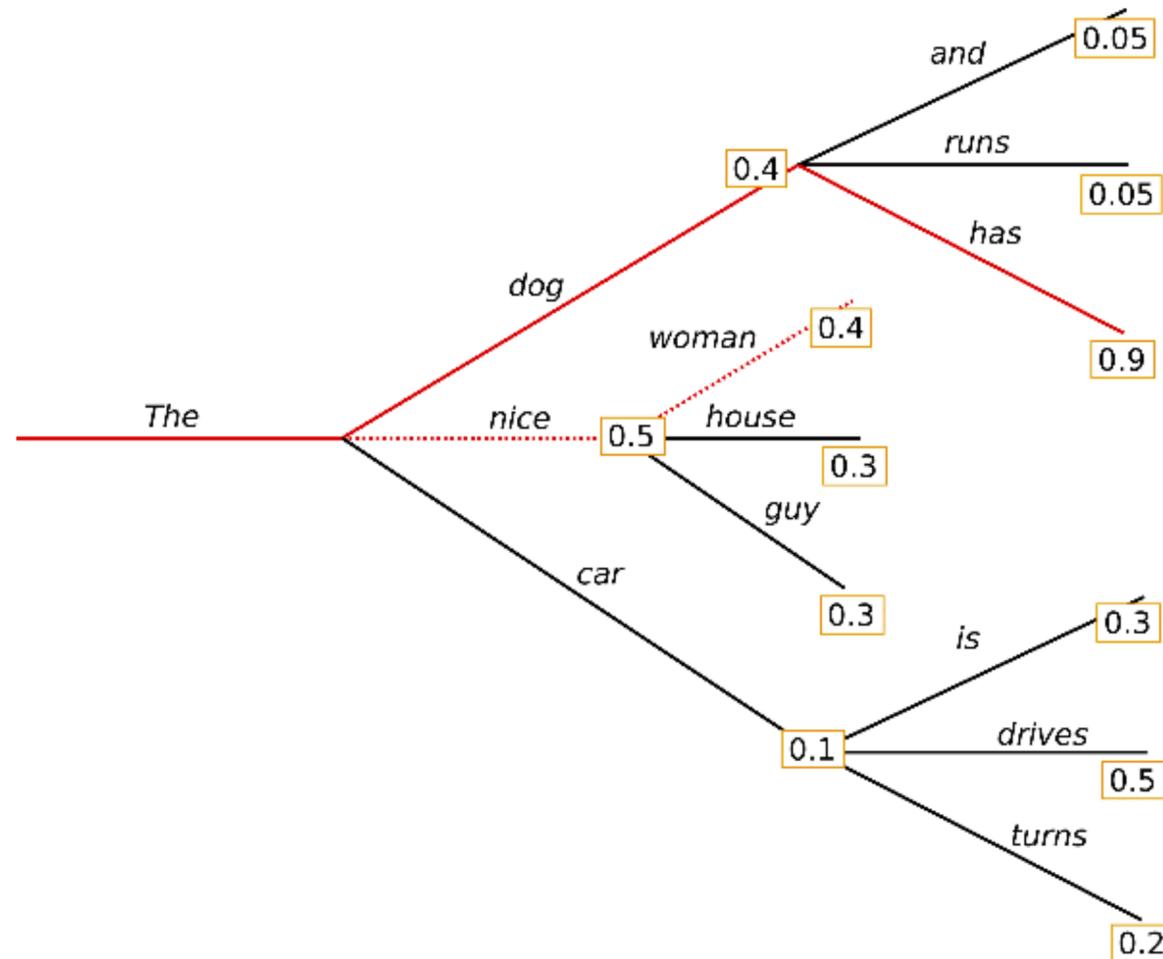


## Problema!

A sentença "The dog has" tem probabilidade  $0.5 \times 0.9 = 0.36$  maior do que  $0.2$  mas foi desconsiderada pela busca gulosa

# Beam Search

Algoritmo de busca local para encontrar a sequência que maximiza e probabilidade  $P(y^{<1>}, \dots, y^{<T_y>} | x)$ . Mantém as  $b$  melhores soluções a cada iteração.

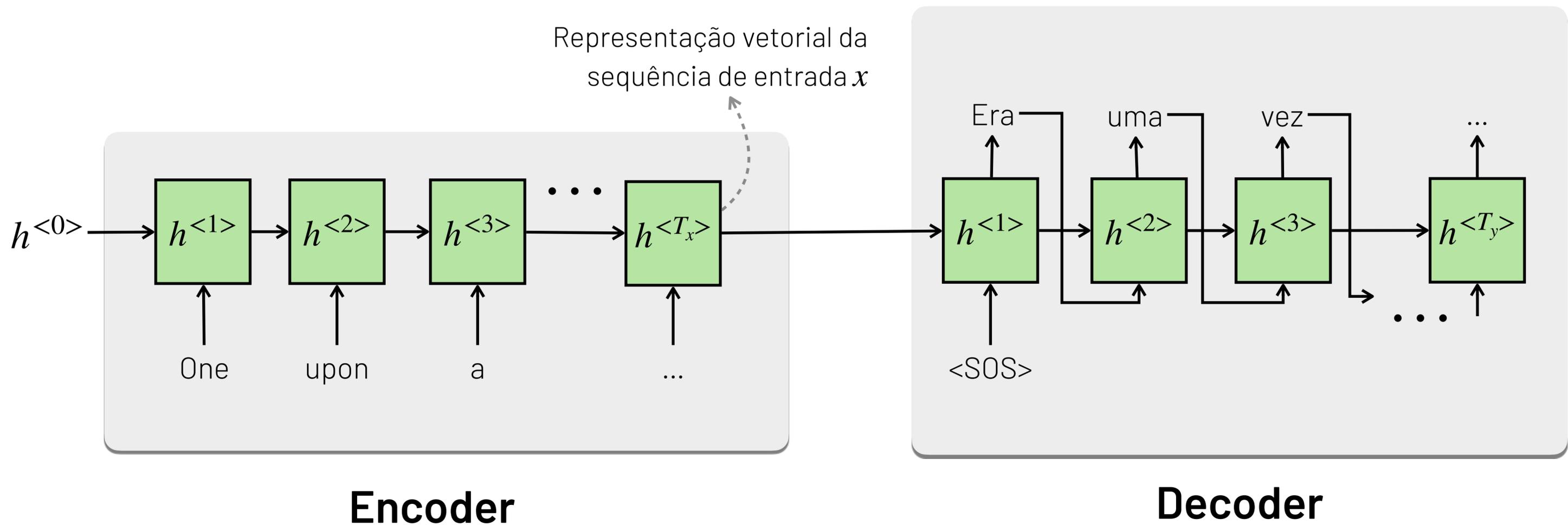


## Solução

Com  $b = 2$ , a sentença "The dog has" (solução ótima) foi escolhida.

# Tradução de sentenças longas

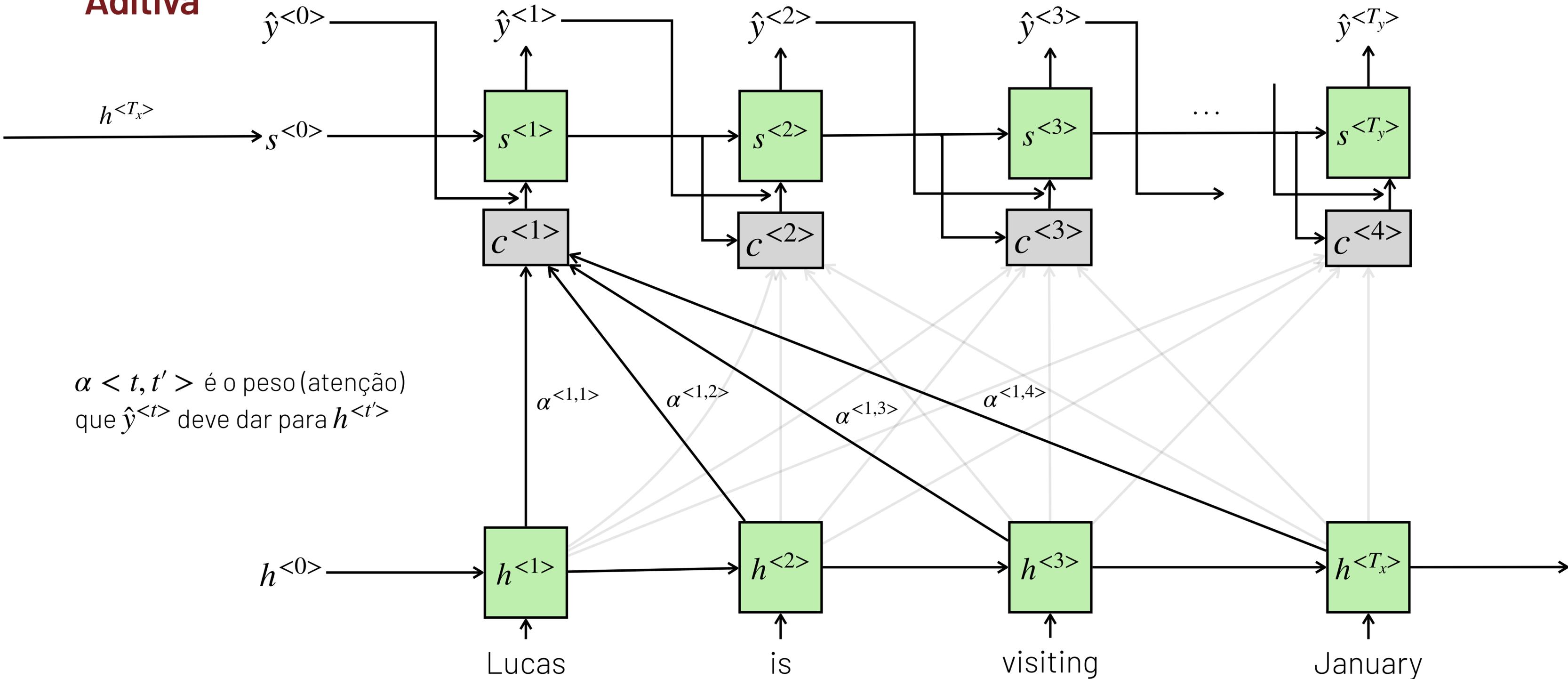
Para sentenças muito longas, é difícil modelar  $x$  em um único vetor  $h^{<T_x>}$



# Atenção!

Considerar todos os estados  $h^{<t>}$  ponderados por pesos  $\alpha^{<t,t'>}$  para calcular  $\hat{y}^{<t>}$

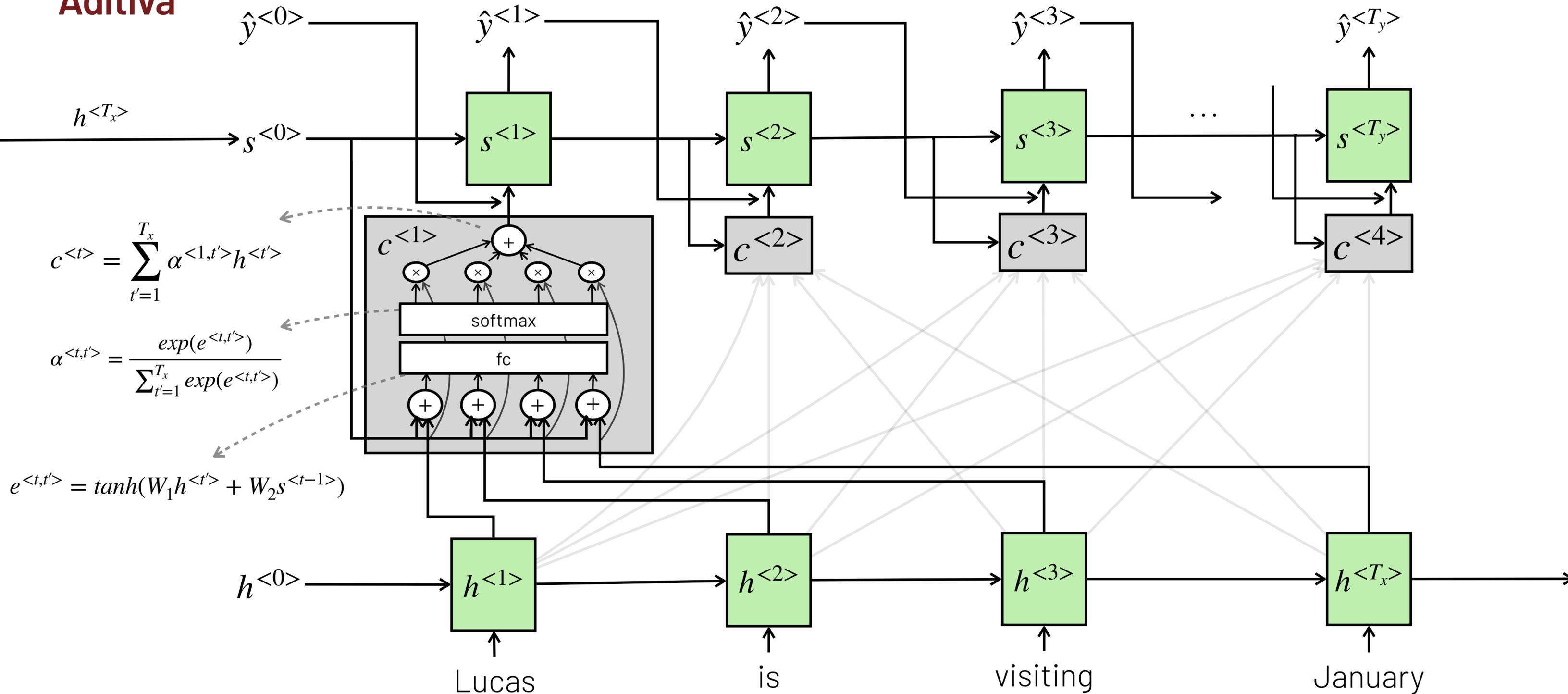
## Aditiva



# Atenção!

Considerar todos os estados  $h^{<t>}$  ponderados por pesos  $\alpha^{<t,t'>}$  para calcular  $\hat{y}^{<t>}$

## Aditiva



# Próxima aula

## **A21:** Transformers

Transformers para problemas de aprendizado com sequências.