

# Cơ Chế Attention Trong Mô Hình Seq2Seq

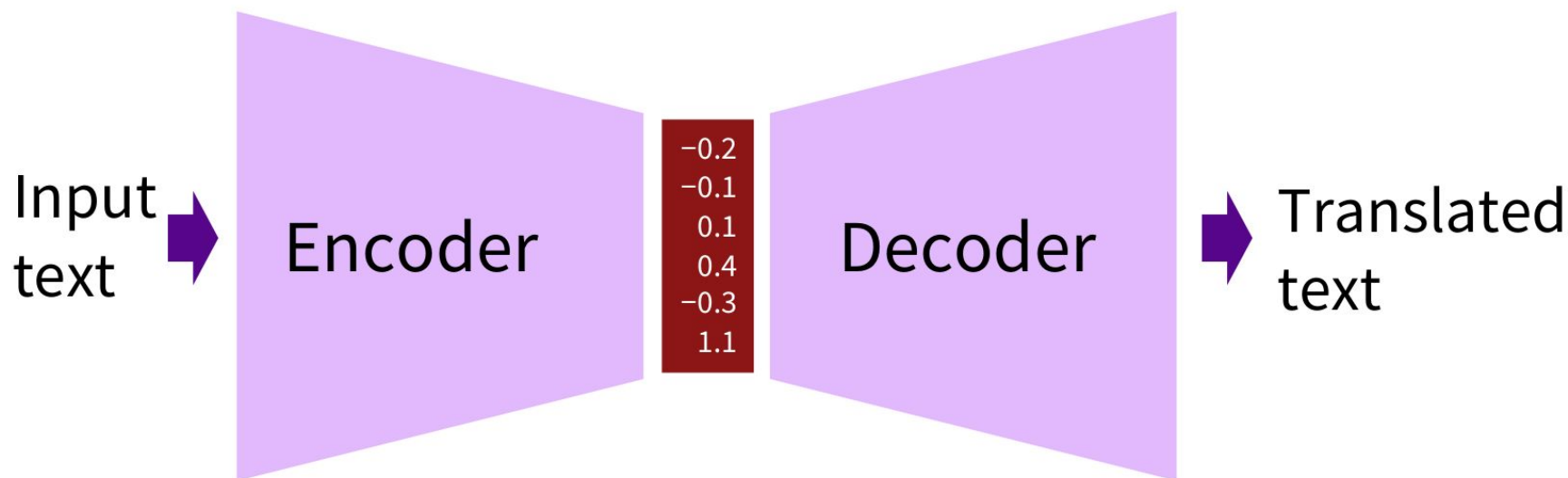


VietAI Teaching Team

# Nội dung

1. Kiến trúc Encoder - Decoder
2. Cơ chế Attention
3. Beam Search

# 1 Kiến trúc Encoder - Decoder



## 2 Cơ chế Attention - Ý tưởng

Thay vì để encoder nén toàn bộ thông tin, ta cho decoder được quan sát toàn bộ output của encoder.

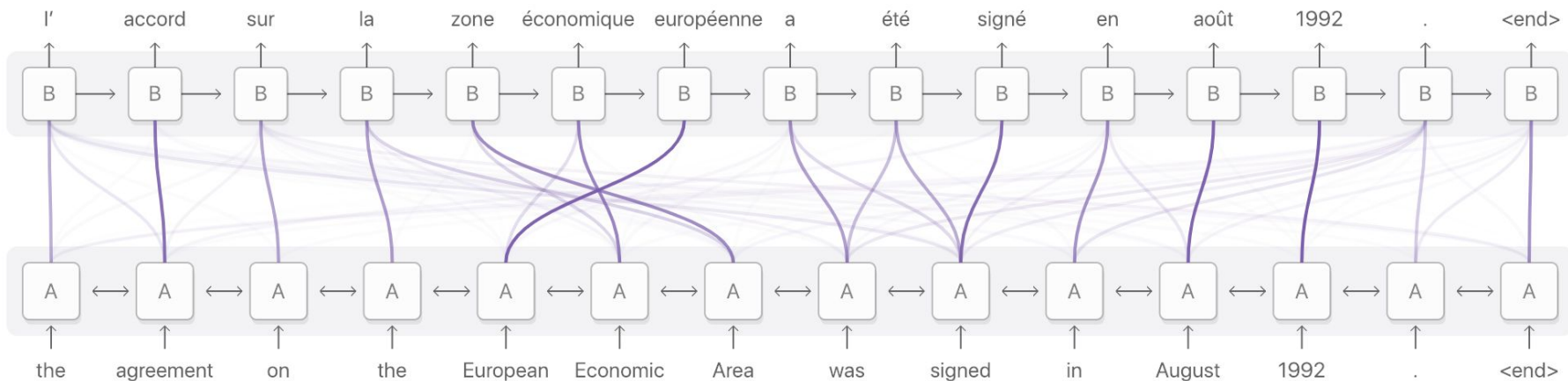
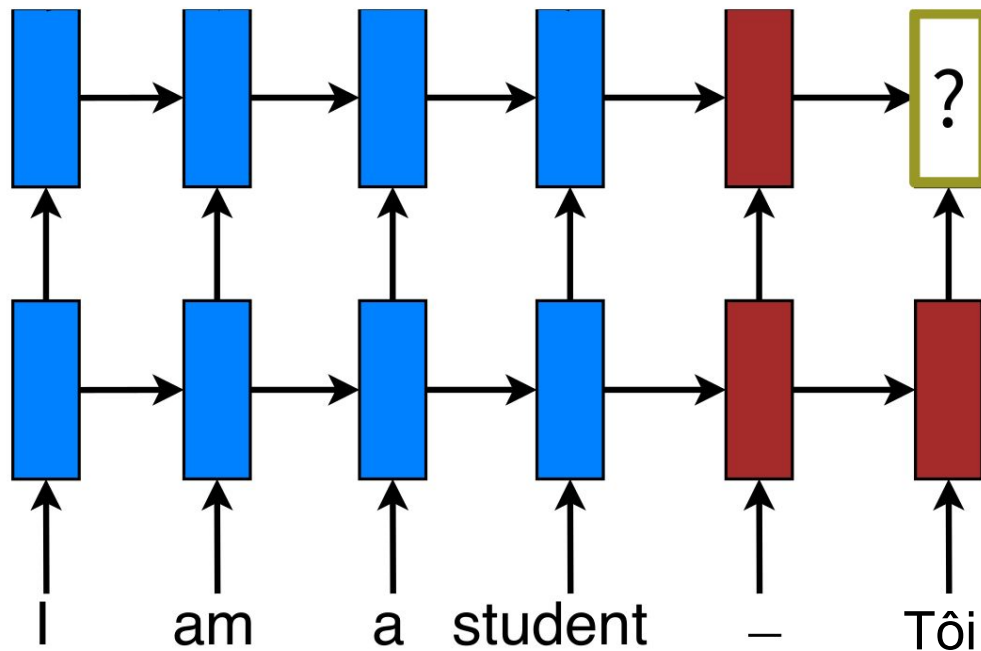
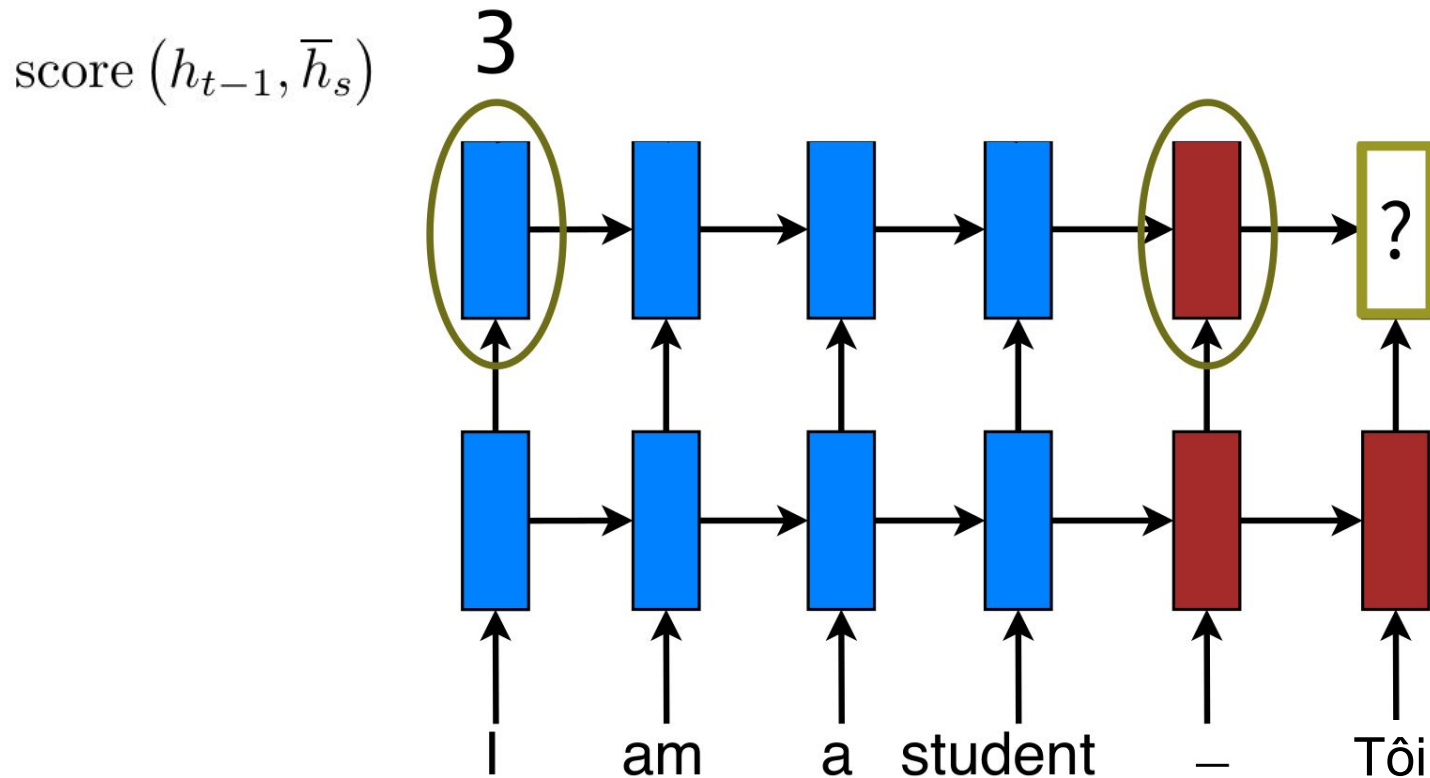


Diagram derived from Fig. 3 of [Bahdanau, et al. 2014](#)

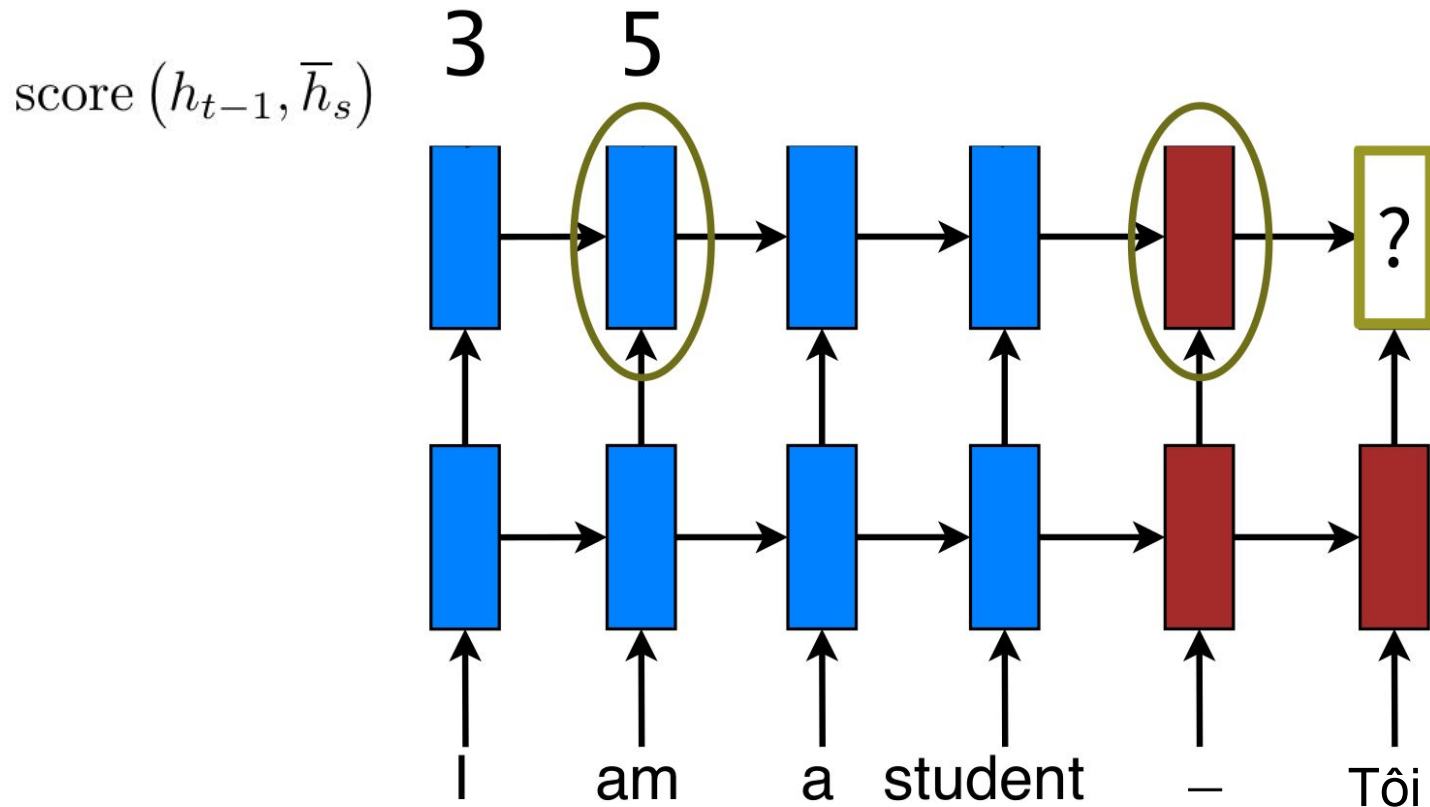
## 2 Cơ chế Attention



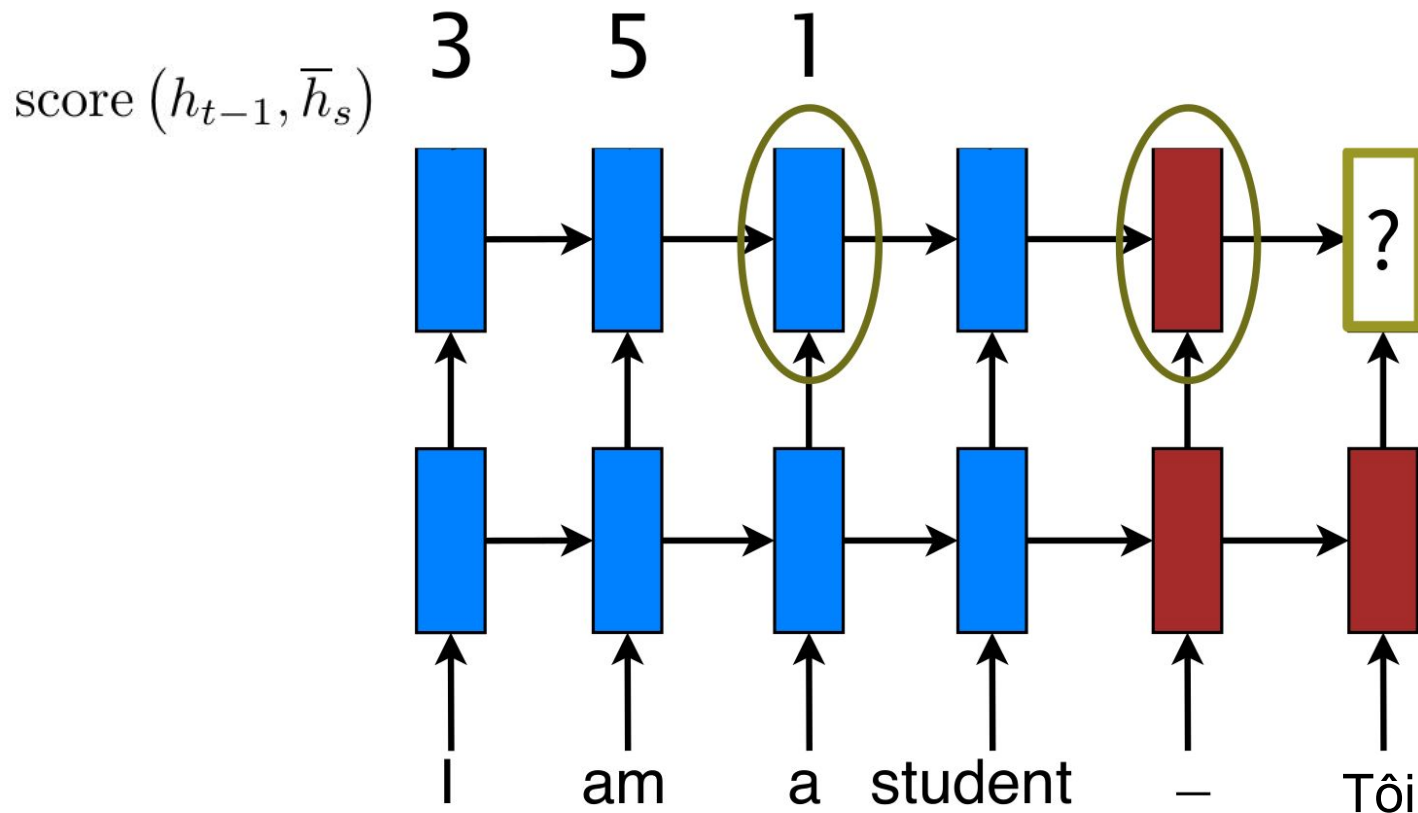
## 2 Cơ chế Attention



## 2 Cơ chế Attention

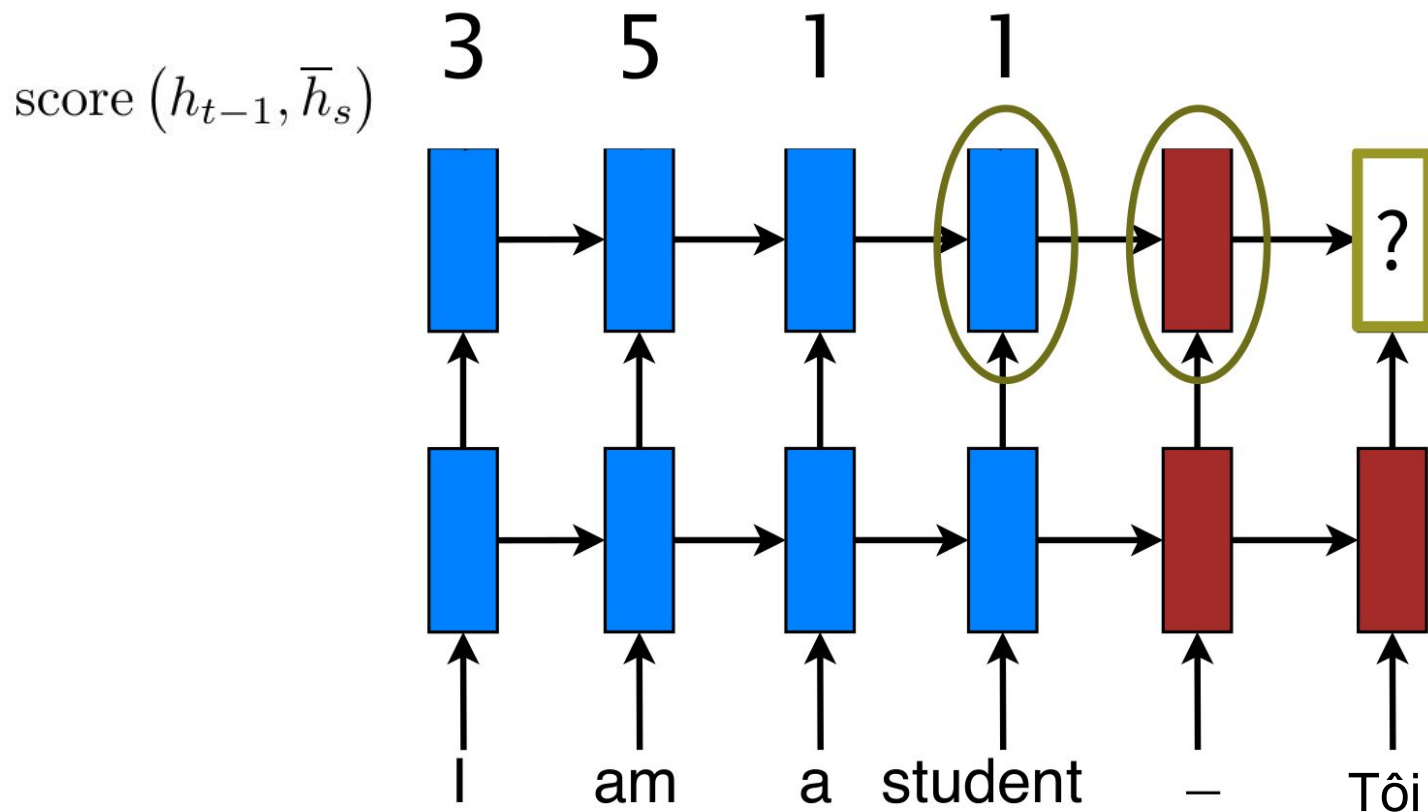


## 2 Cơ chế Attention



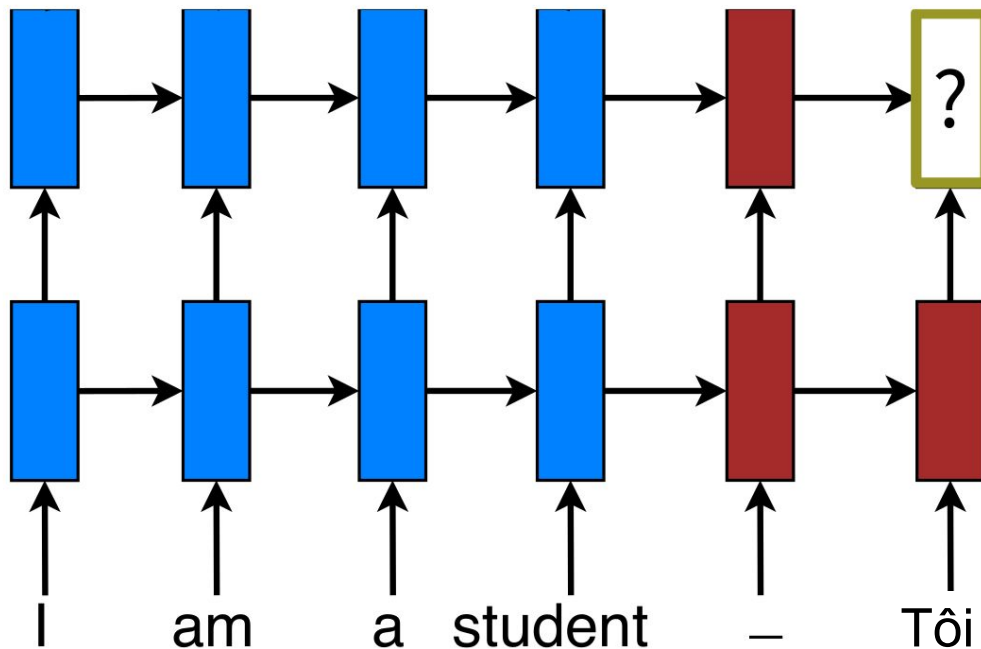


## 2 Cơ chế Attention



## 2 Cơ chế Attention

score  $(h_{t-1}, \bar{h}_s)$     0.3   0.5   0.1   0.1



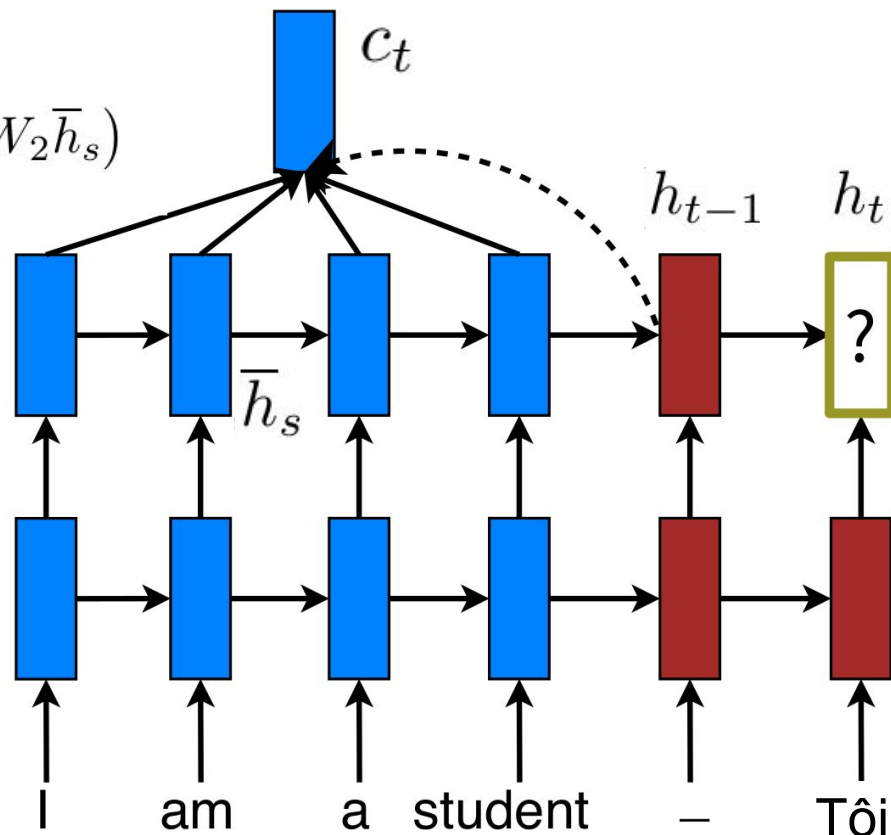
## 2 Cơ chế Attention

$$\text{score}(h_{t-1}, \bar{h}_s) = v_a^\top \tanh(W_1 h_{t-1} + W_2 \bar{h}_s)$$

$$\alpha_{ts} = \frac{\exp(\text{score}(h_{t-1}, \bar{h}_s))}{\sum_{s'} \exp(\text{score}(h_{t-1}, \bar{h}_{s'}))}$$

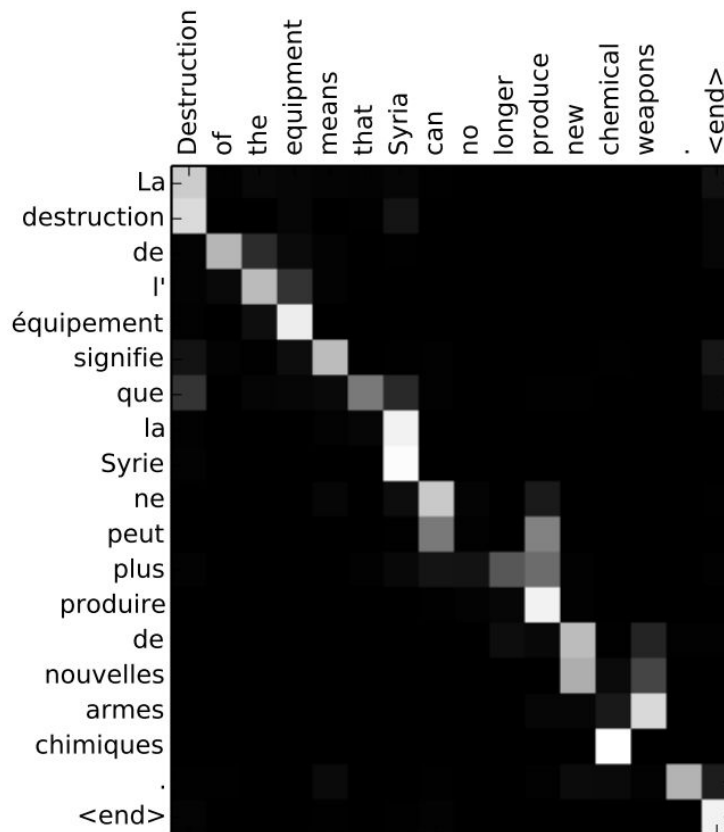
$$c_t = \sum_s \alpha_{ts} \bar{h}_s$$

$$h_t = f(h_{t-1}, y_t, c_t)$$

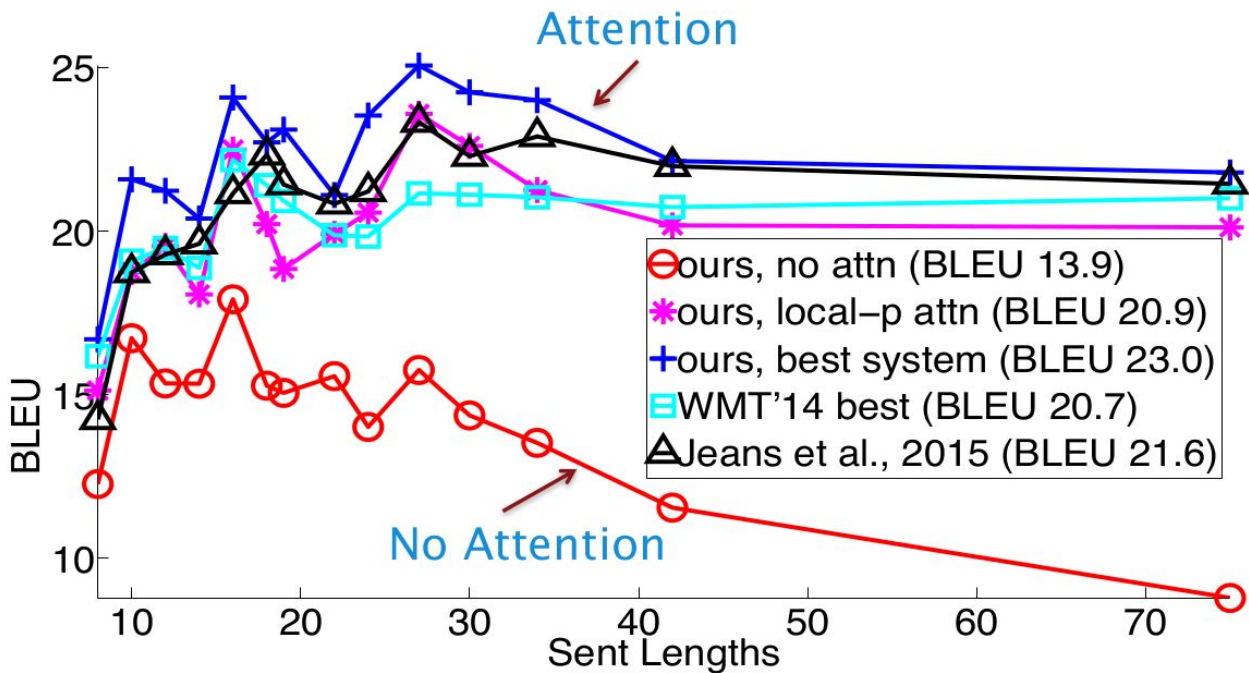


## 2 Cơ chế Attention

Cơ chế Attention tự động xây dựng sự tương ứng (alignment) giữa các từ tiếp theo phải sinh ra trong câu đích và các từ trong câu nguồn từ đó giúp việc dịch chính xác hơn, đặc biệt với các câu dài.

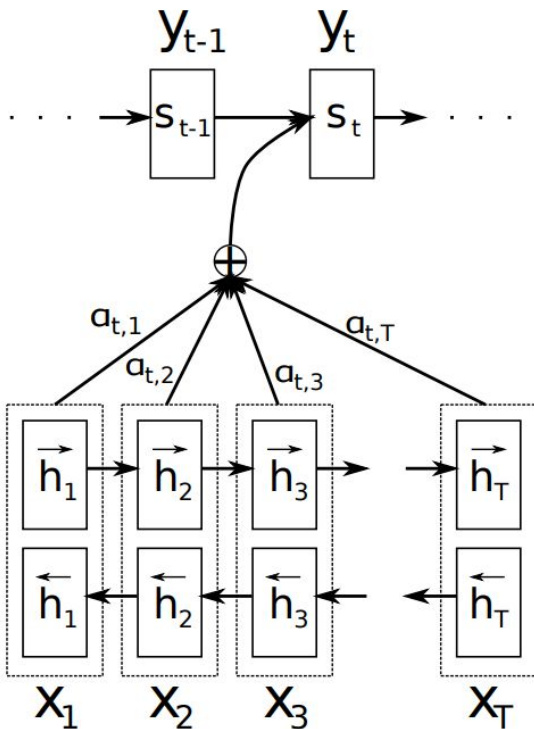


## 2 Cơ chế Attention

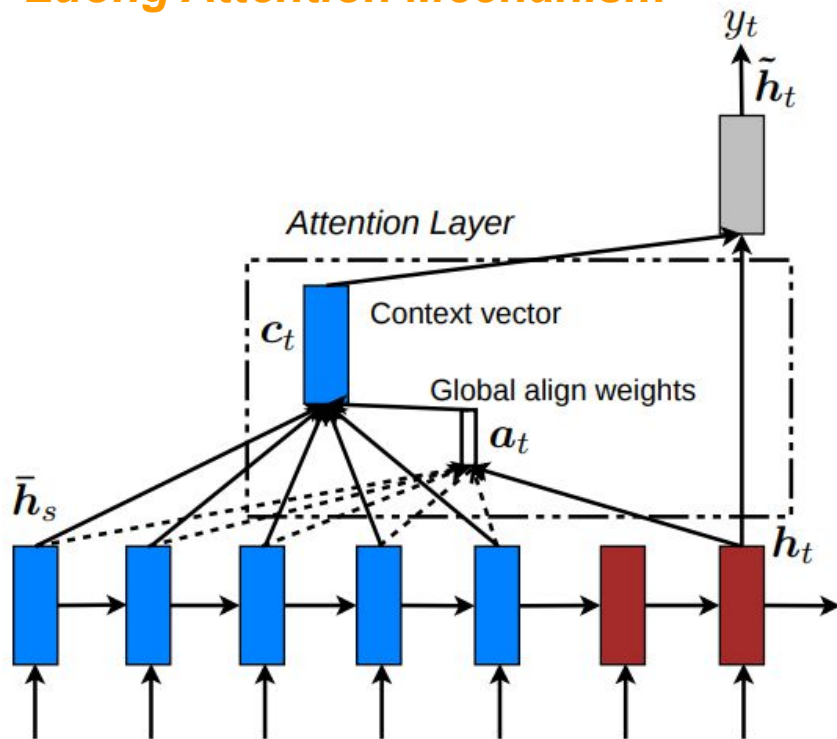


## 2 Attention Models

### Bahdanau Attention Mechanism

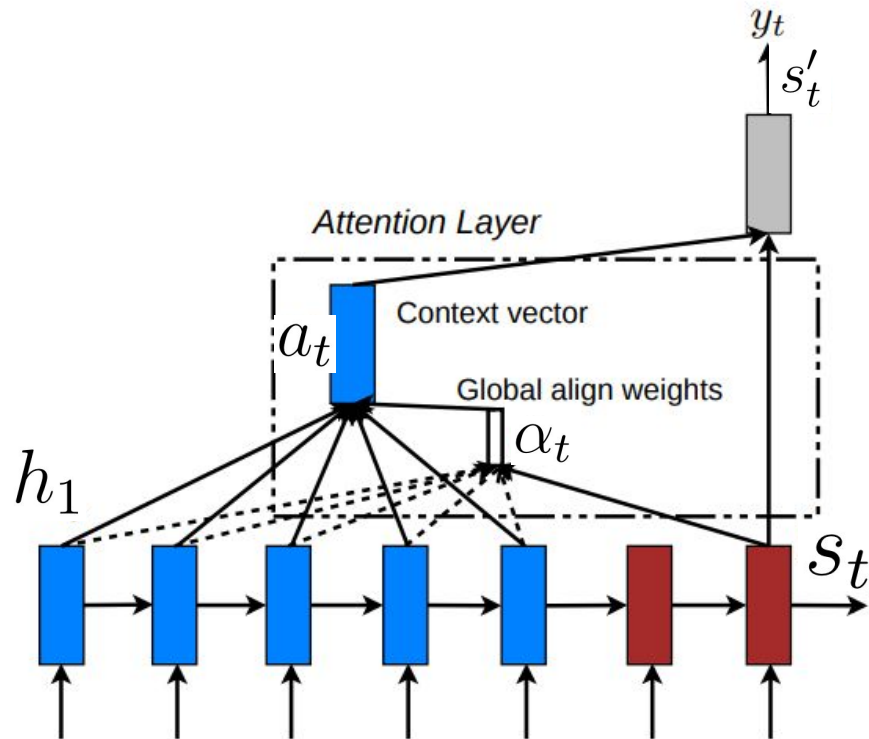
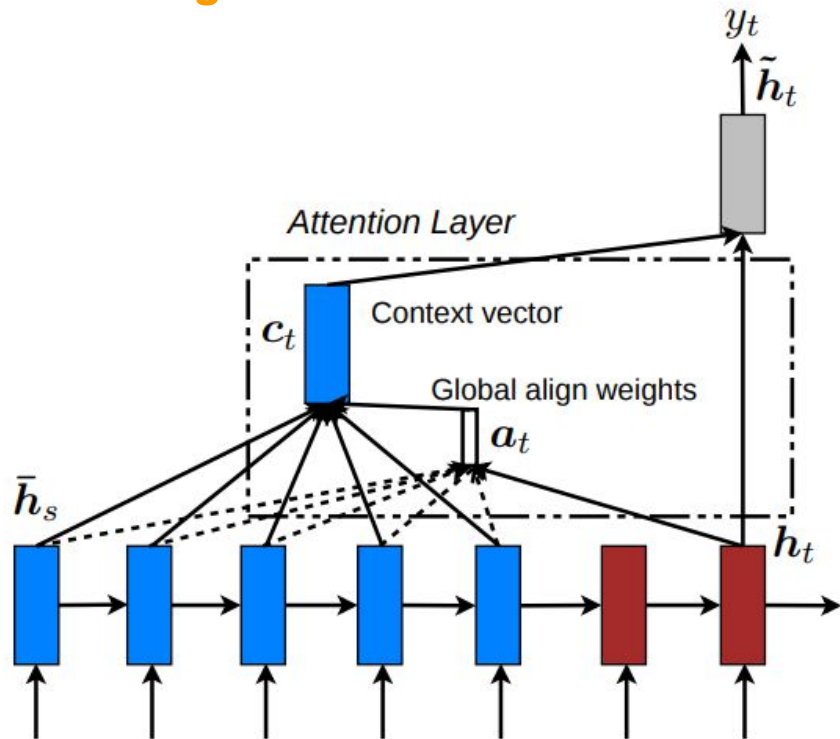


### Luong Attention Mechanism



## 2 Attention Models

### Luong Attention Mechanism

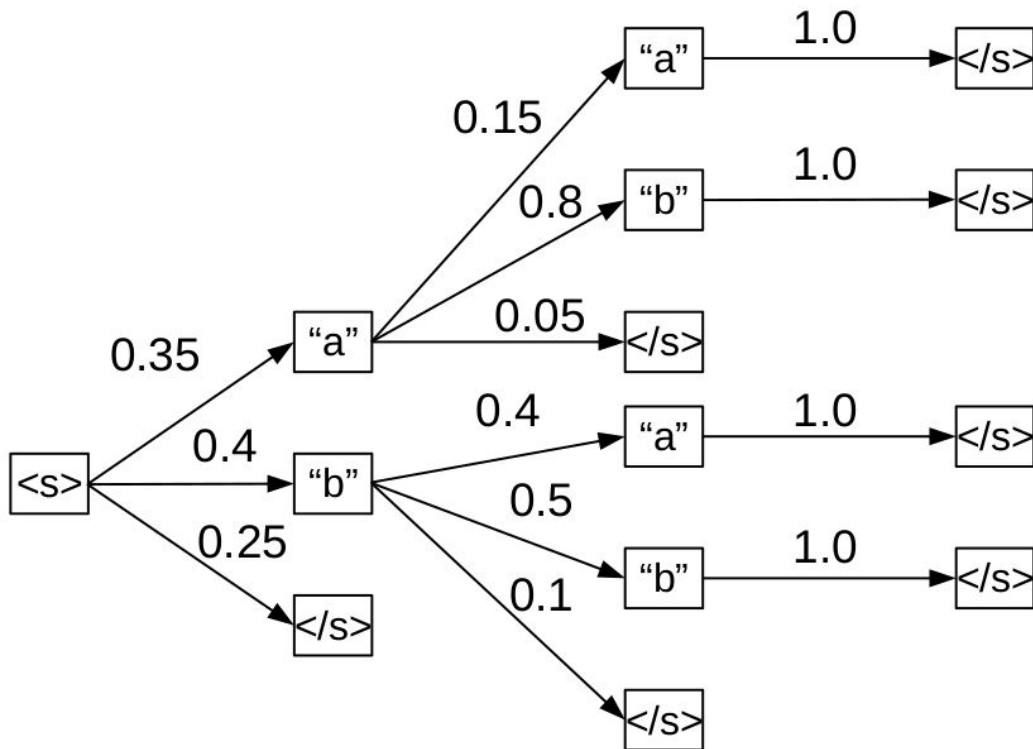


### 3 Beam Search

- Để tìm được chuỗi  $y$  cho xác suất  $P(y_1 y_2 \dots y_N | x)$  cao nhất, ta phải tìm thông qua các xác suất  $P(y_i | y_1 \dots y_{i-1}, x)$
- Có nhiều cách để tìm  $y$ :
  - **Random Sampling (Acestral Sampling)**: tại bước thứ  $i$ , chọn ngẫu nhiên  $y_i$  theo phân phối xác suất được cho từ mạng decoder.
  - **Greedy Search**: tại bước thứ  $i$ , chọn  $y_i$  là từ cho giá trị  $P(y_i | y_1 \dots y_{i-1}, x)$  cao nhất.
  - **Beam Search**: tại bước thứ  $i$ , chỉ giữ lại  $k$  chuỗi con độ dài  $i$  cho xác suất  $P(y_1 y_2 \dots y_i)$  ( $k$  được gọi là *beam size* hay *beam width*) cao nhất để thực hiện việc tìm kiếm tại bước  $i + 1$  tiếp theo.



### 3 Beam Search



- Theo hình bên, chuỗi "a b </s>" cho ta xác suất cao nhất.
- Nếu áp dụng Greedy Search sẽ cho ta chuỗi "b b </s>".
- Nếu áp dụng Beam Search với  $k = 2$  sẽ cho ta chuỗi: "a b </s>"

# Tài liệu tham khảo

1. CS224N, Stanford University
2. Bahdanau et al., [Neural machine translation by jointly learning to align and translate](#) - ICLR15.
3. Sutskever et al., [Sequence to sequence learning with neural networks](#) - NIPS14.
4. Luong et al., [Effective approaches to attention-based neural machine translation](#) - EMNLP15.
5. [Neural Machine Translation \(seq2seq\) Tutorial with TensorFlow](#)