

# CƠ SỞ LÝ THUYẾT NEURAL MACHINE TRANSLATION

Dịch máy từ tiếng Việt sang tiếng Anh

Bộ môn : TRÍ TUỆ NHÂN TẠO  
GVHD: Đoàn Văn Thắng

# MỤC LỤC

## Nội dung trình bày

**01**

**Neural Machine  
Translation là  
gì?**

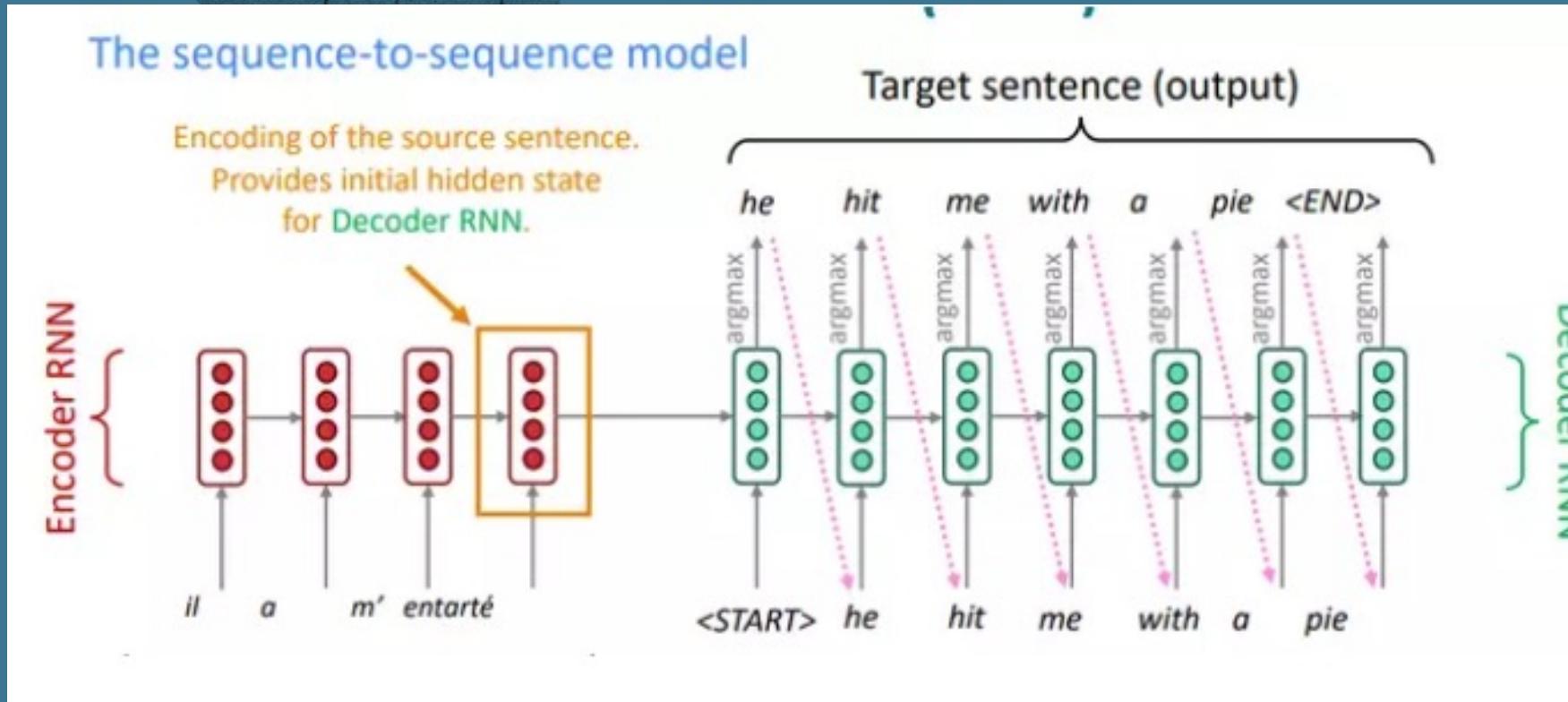
**02**

**Seq2Seq**

**03**

**Một số cách  
đánh giá một  
mô hình  
Machine  
Translation**

# NEURAL MACHINE TRANSLATION LÀ GÌ?



**Neural Machine Translation(NMT) là**  
**một cách để cho máy tính thực hiện**  
**dịch ngôn ngữ với một end-to-end**  
**neural network.**

**Kiến trúc neural network này được gọi là**  
**sequence-to-sequence model(aka**  
**seq2seq).** (Ảnh minh họa như hình bên)

# SEQ2SEQ

**S**eq2Seq là một mô hình được đề xuất từ năm 2014 cho việc dịch tiếng Anh-tiếng Pháp. Seq2Seq cũng là một Conditional Language Model. NMT-S seq2Seq tính toán  $P(y|x)$  một cách trực tiếp:

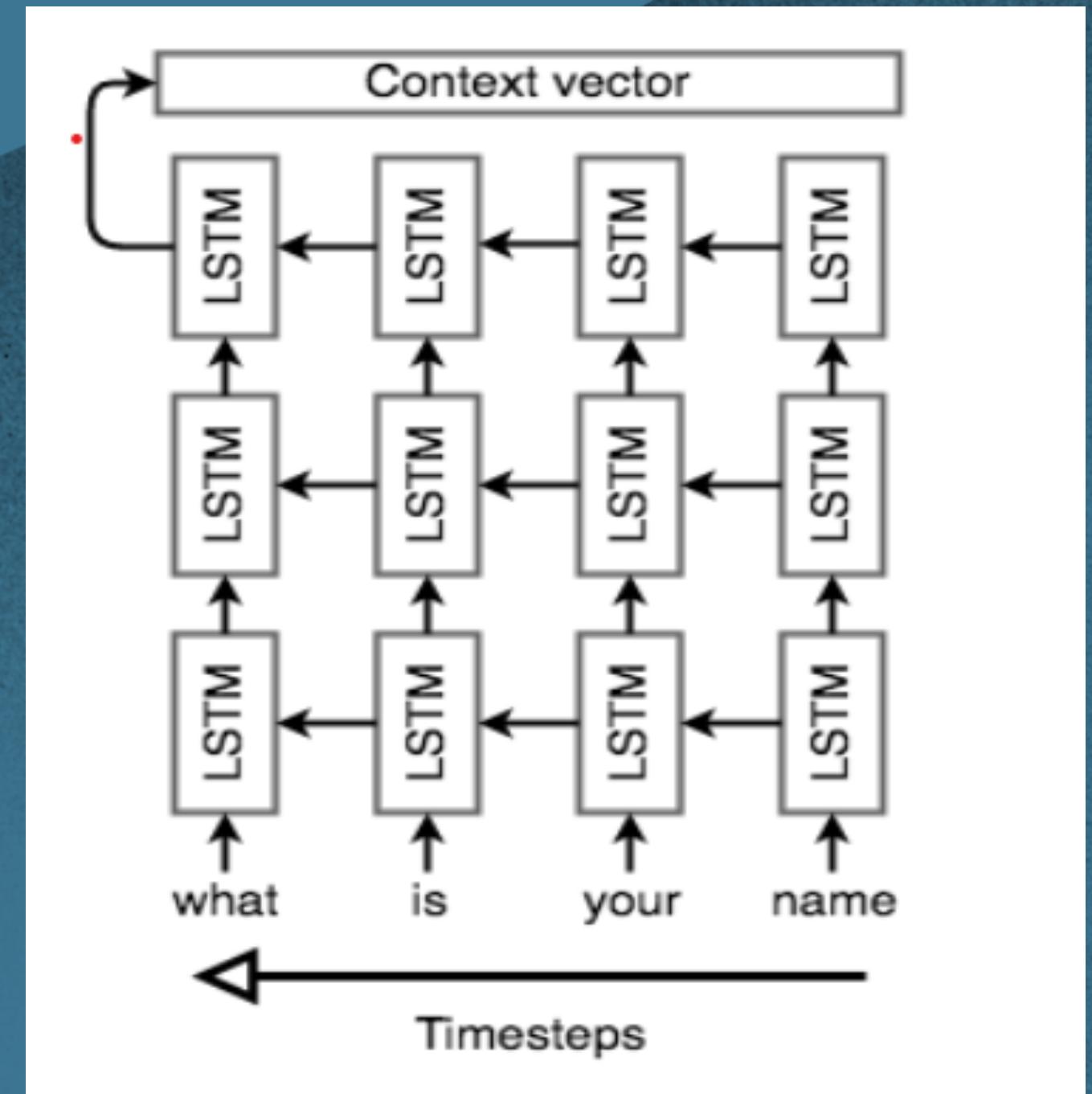
$$P(y|x) = P(y_1|x)P(y_2|y_1, x)P(y_3|y_1, y_2, x)...P(y_T|y_1...y_{T-1}, x)$$

với  $T$  là số bước thời gian.

# KIẾN TRÚC SEQ2SEQ-ENCODER

**Nhiệm vụ của Encoder là đọc câu nguồn cho mô hình Seq2Seq và tạo context-vector C có kích thước cố định. Để làm được điều này, Encoder sử dụng một mạng RNN(thường là LSTM) để đọc từng token tại một bước thời gian t. Hidden-state cuối cùng sẽ trở thành C.**

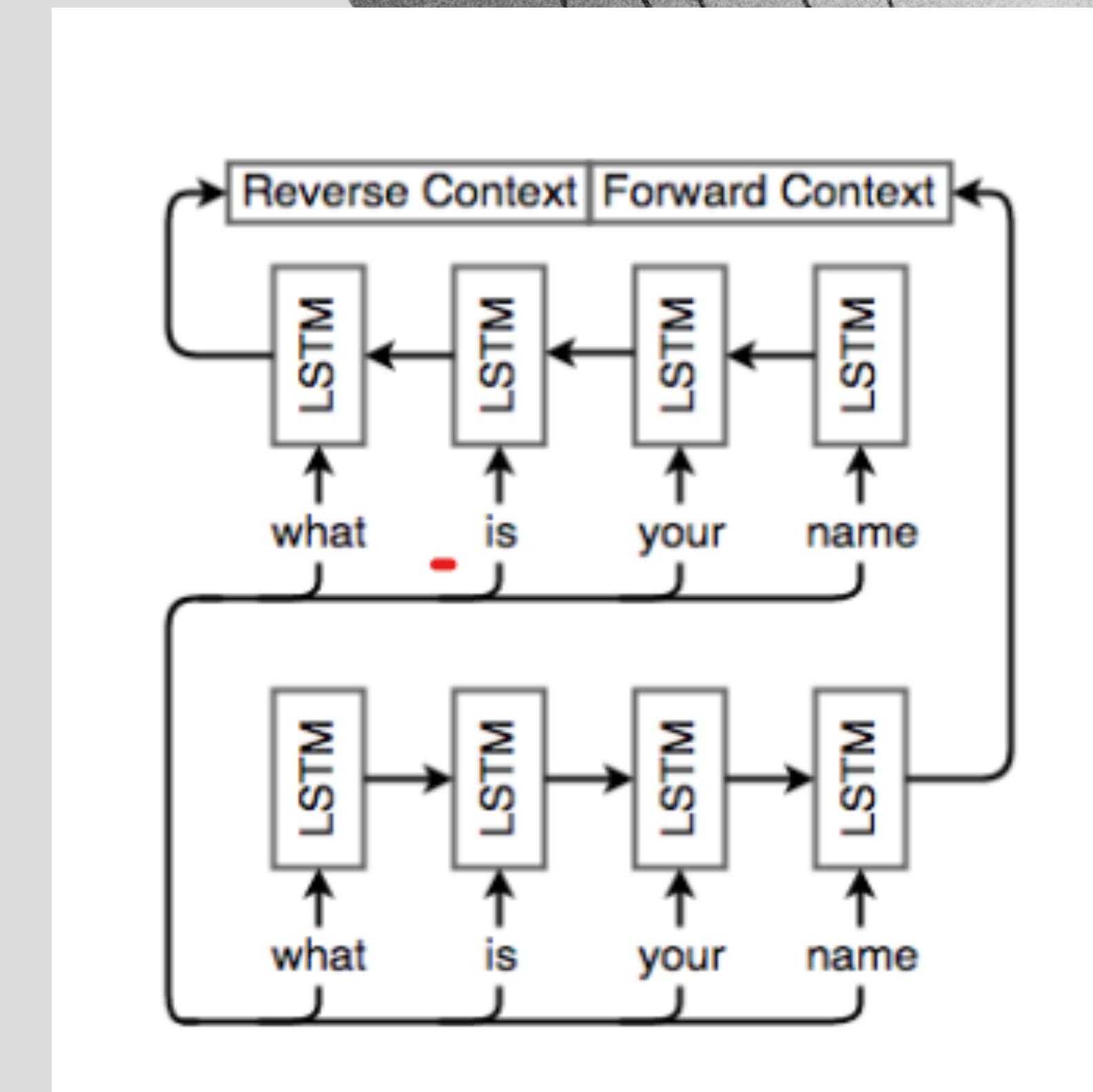
Sau đây là hình ảnh minh họa về encoder sử dụng LSTM:



# BIDIRECTIONAL RNNs

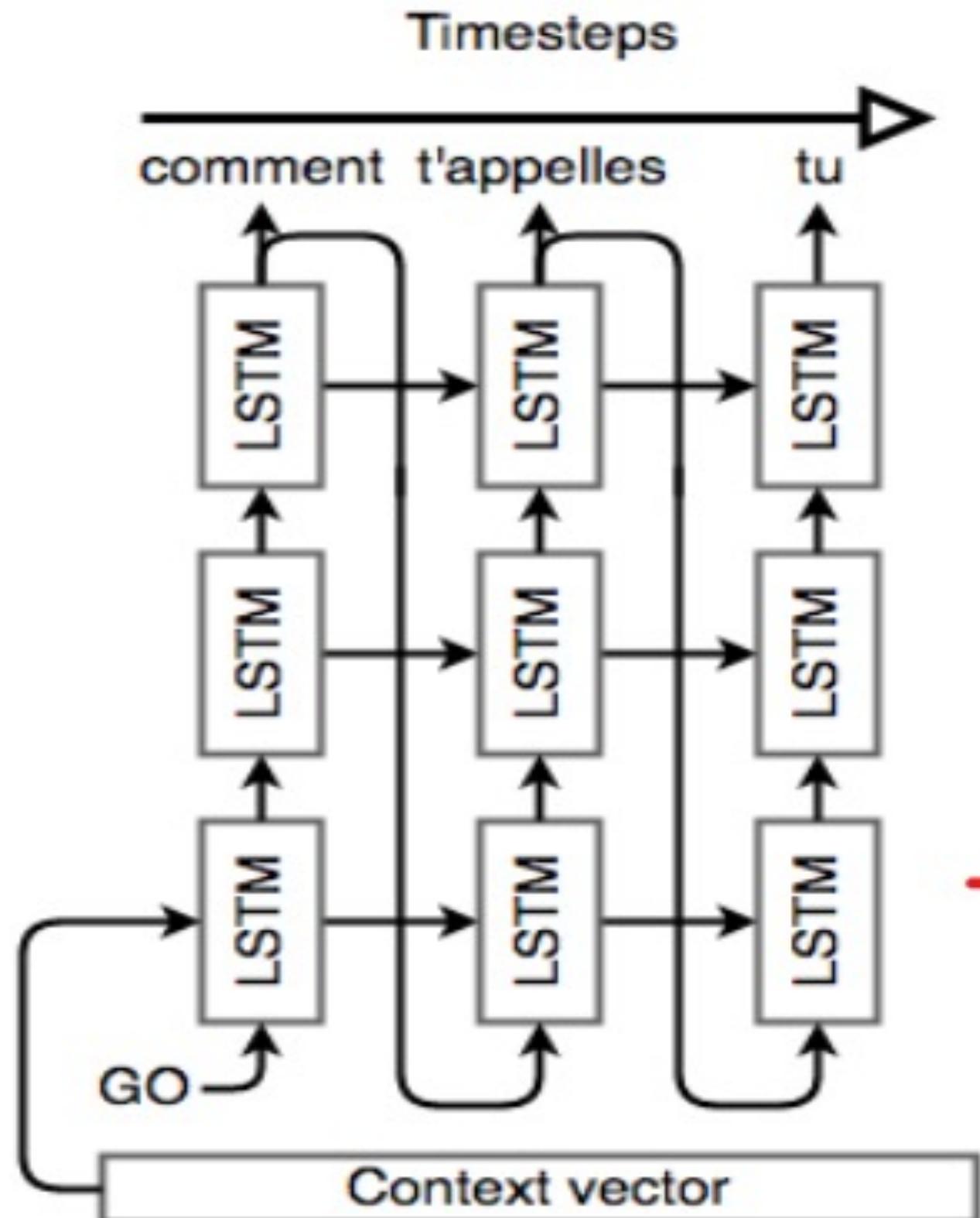
Để giúp model nắm bắt tốt sự phụ thuộc trong câu nguồn, ta sử dụng Bidirectional RNNs.

Minh họa cho việc encoder sử dụng Bidirectional RNN:



# KIẾN TRÚC SEQ2SEQ-DECODER

Decoder cũng là một mạng LSTM nhưng sử dụng phức tạp hơn encoder một chút, để tạo ra các từ ở mỗi bước thời gian. Để làm được điều đó, ta sử dụng multi-layer và khởi tạo hidden-state của layer đầu tiên bằng việc sử dụng context-vector của encoder.

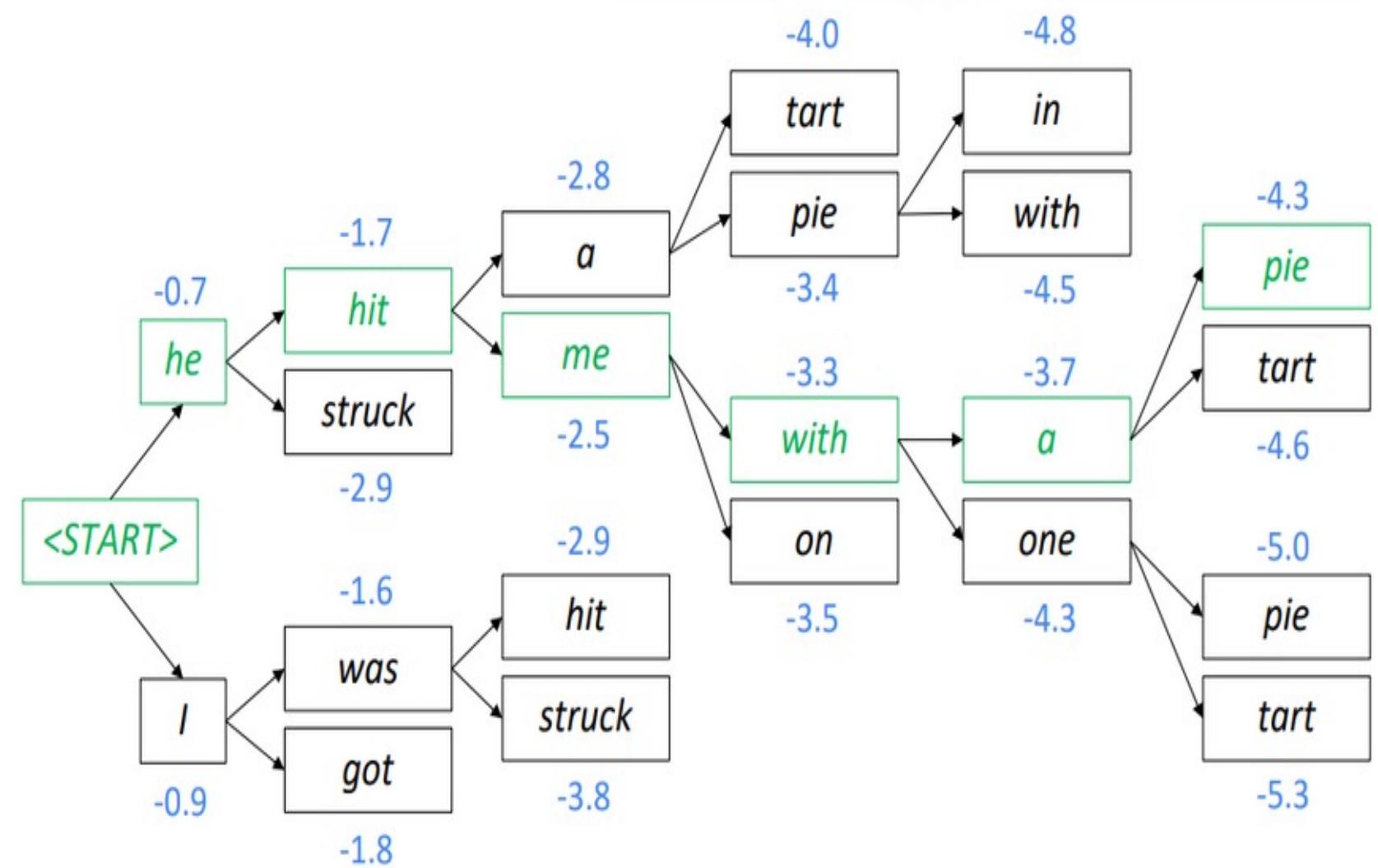


# CÁCH CHỌN ĐẦU RA CHO SEQ2SEQ-NMT

Vì không gian tìm kiếm có thể rất lớn, ta phải thu nhỏ kích thước của nó

Chúng em chọn phương pháp Beam Search:  
Ý tưởng của beam search là sẽ chọn lấy K xác suất lớn nhất của một softmax trong một bước thời gian và chỉ xem K sự lựa chọn cho một bước thời gian, cụ thể được giải thích trong hình sau

Beam size =  $k = 2$ . Blue numbers =  $\text{score}(y_1, \dots, y_t) = \sum_{i=1}^t \log P_{\text{LM}}(y_i | y_1, \dots, y_{i-1}, x)$



# MỘT SỐ CÁCH ĐÁNH GIÁ MỘT MÔ HÌNH MACHINE TRANSLATION

## • Bilingual Evaluation

**Understudy(BLEU):** Trong năm 2002, cách nhà nghiên cứu của IBM đã phát triển e Bilingual Evaluation Understudy (BLEU). Đến nay nhiều biến thể của BLEU đã được sinh ra, BLEU là một trong các phương pháp tin cậy nhất trong Machine Translation

- **Human Evaluation:** Phương pháp đầu tiên và có thể ít gây ngạc nhiên nhất là để mọi người tự đánh giá tính đúng đắn, đầy đủ và thông thạo của hệ thống dịch. Giống như Turing Test, nếu bạn có thể đánh lừa con người rằng không thể phân biệt bản dịch do con người tạo ra với bản dịch hệ thống của bạn, mô hình của bạn đã vượt qua bài kiểm tra để trông giống như một câu ngoài đời thực.

THANK FOR WATCHING