

# Attention Seminar

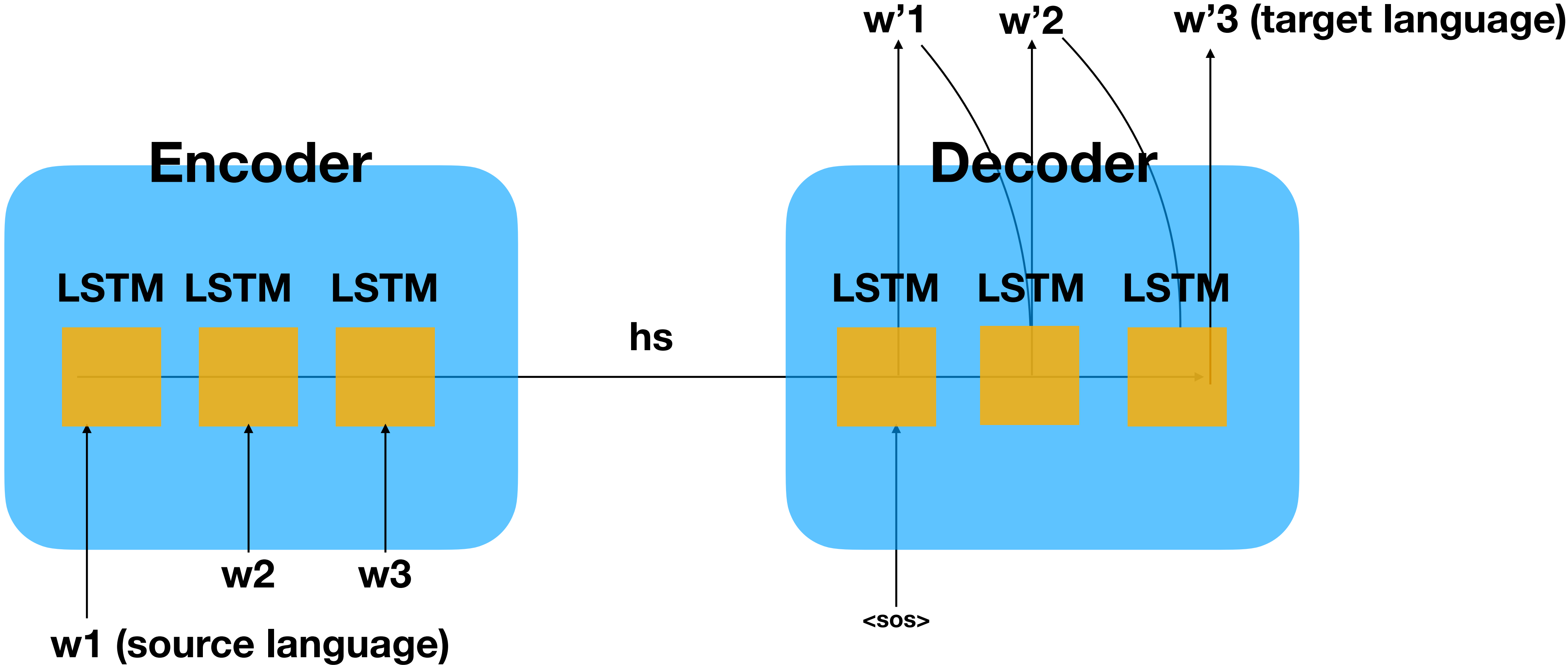
**‘NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE’ (2014)**

**발표자: 정재영**

# Attention

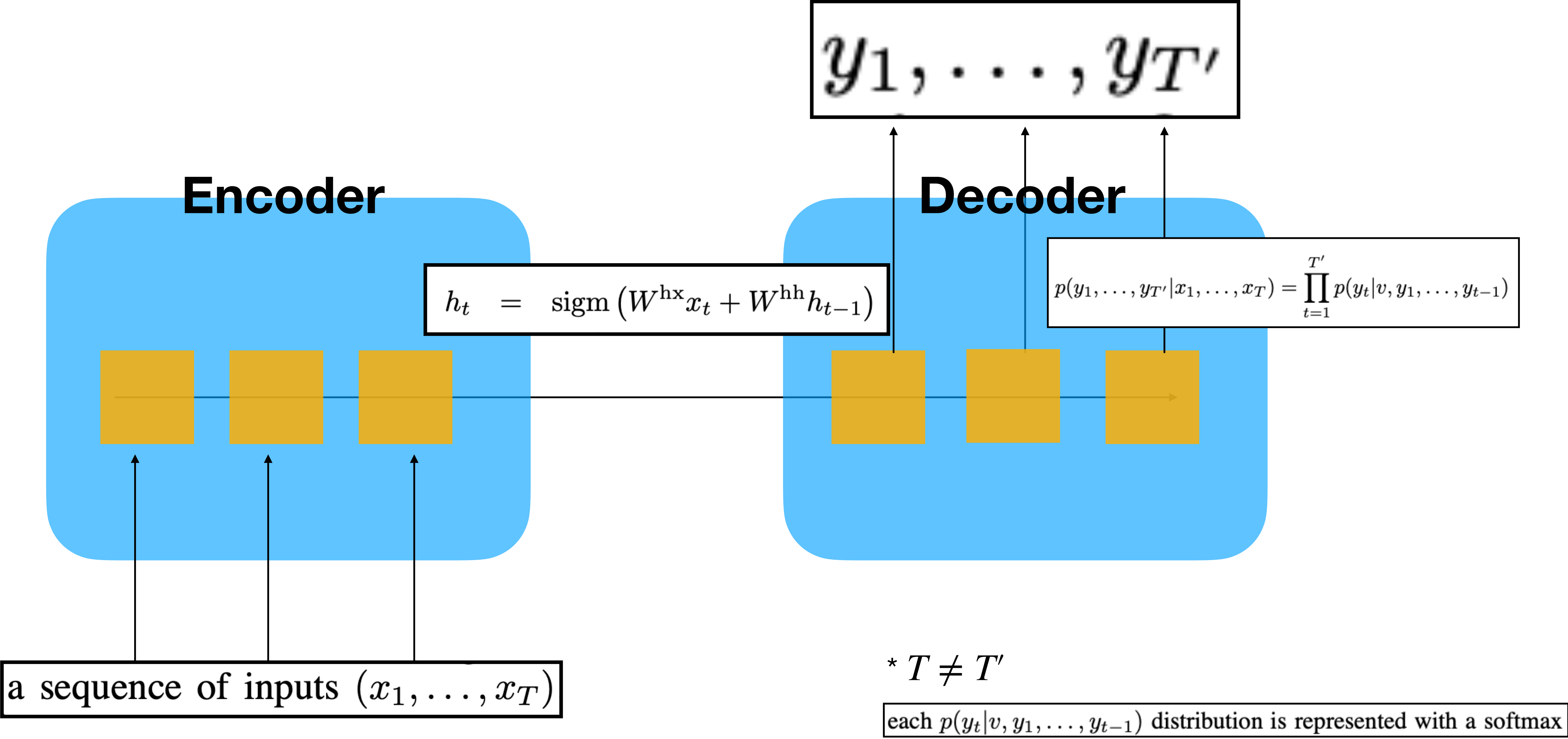
- 1. Seq2seq 소개
- 2. Seq2seq with Attention

seq2seq



# Attention

- 1. Seq2seq 소개
- 2. Seq2seq with Attention



# Attention

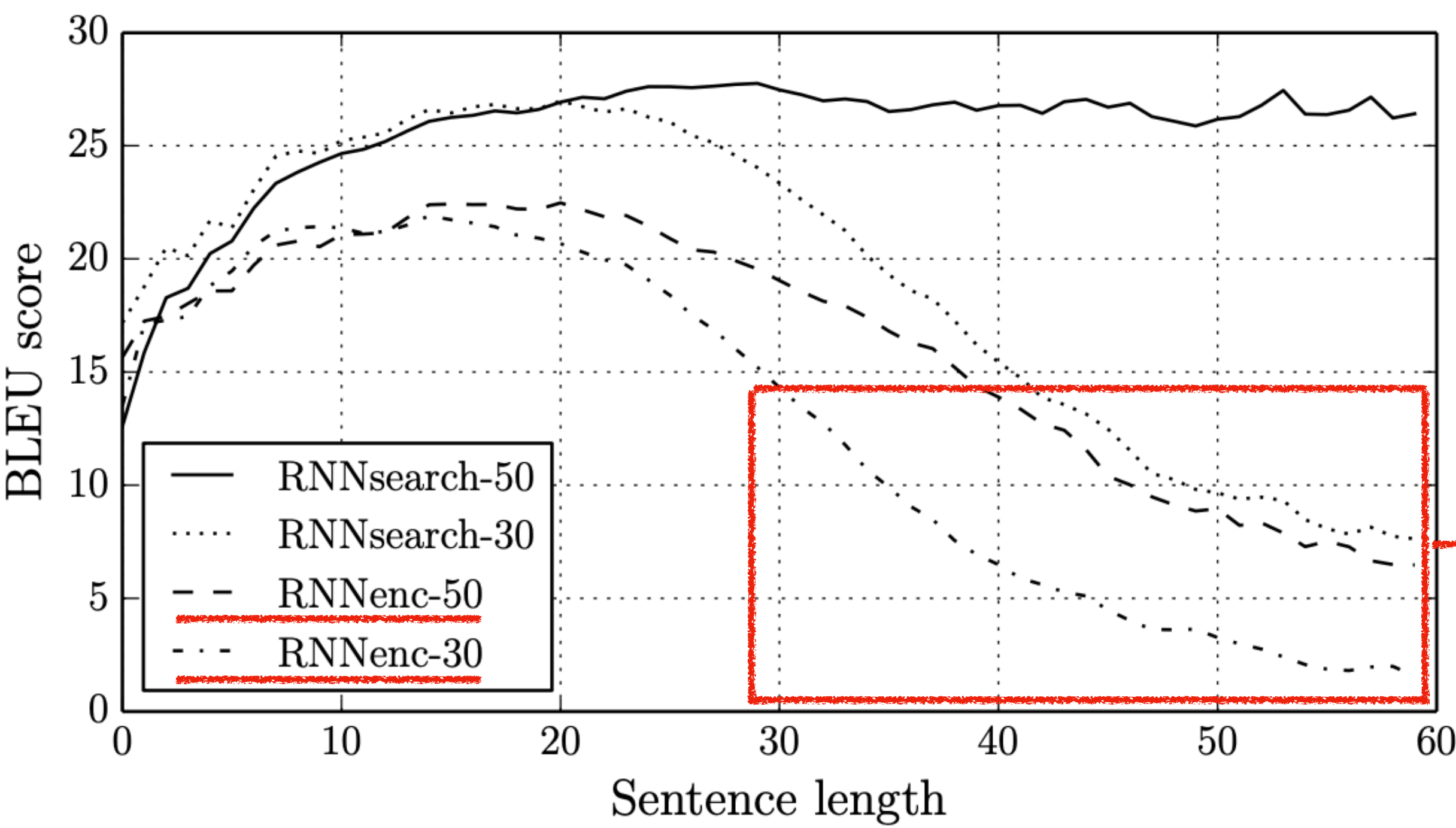
- 1. Seq2seq 소개
- 2. Seq2seq with Attention

## Attention 도입 동기: Vanila seq2seq의 한계

: Vanila seq2seq의 경우, 인코더에서 마지막 은닉 벡터hidden vector는 고정 길이 벡터이다.

➡ 여러 길이의 다양한 문장들에 대해 일정한 성능을 보이지 않는다. ex) 길이가 긴 문장일수록 MT 작업에 대해 낮은 성능을 보인다.

RNNsearch: 어텐션 버전 seq2seq  
RNNenc: Vanila seq2seq  
30: 30 개의 단어가 입력됨



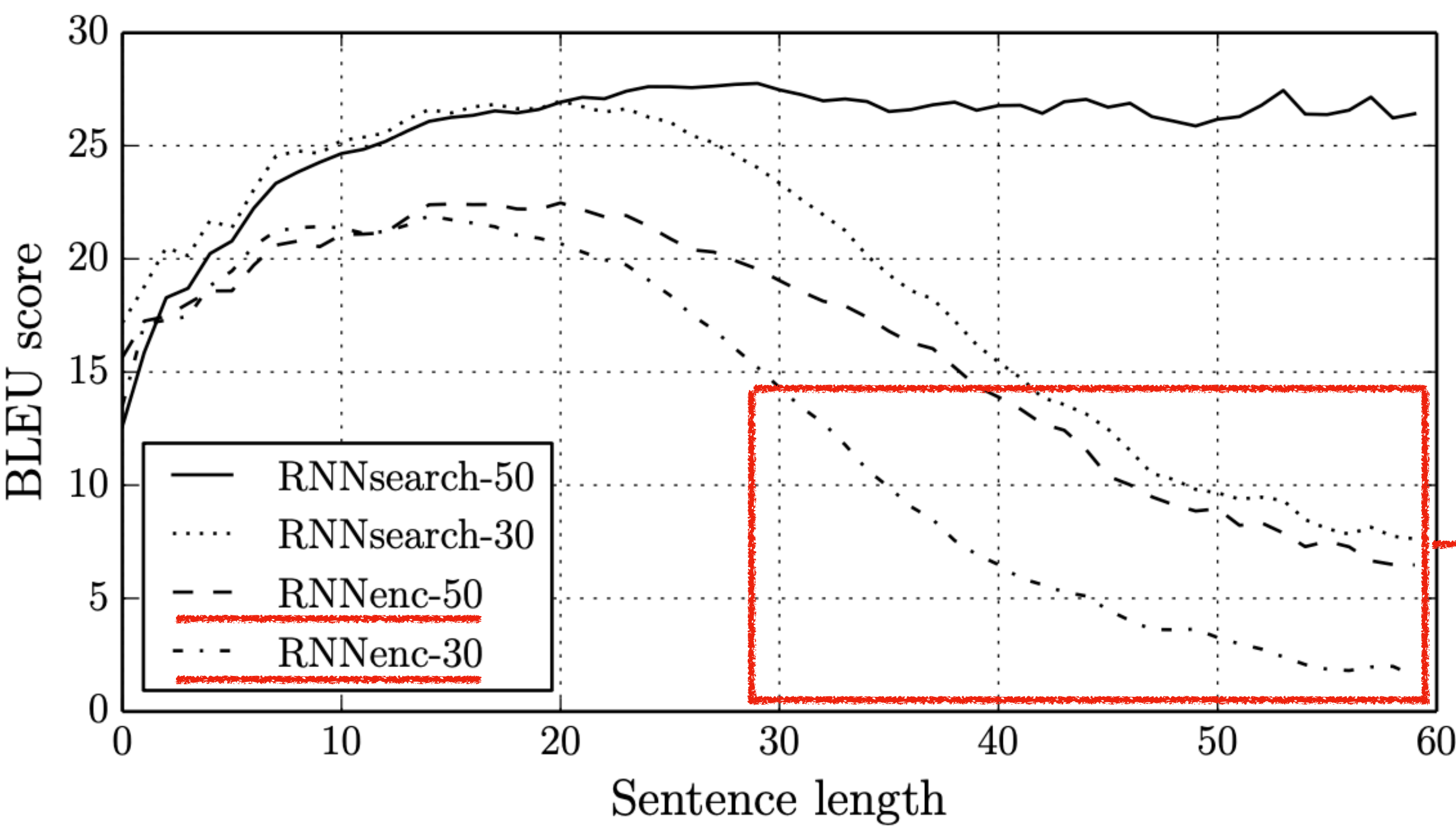
Vanila seq2seq은 입력 문장 길이가 길수록 BLEU 성능 측정 지표가 낮게 나온다.

# Attention

- 1. Seq2seq 소개
- 2. Seq2seq with Attention

## Attention 도입 동기: Vanila seq2seq의 한계 극복 방안

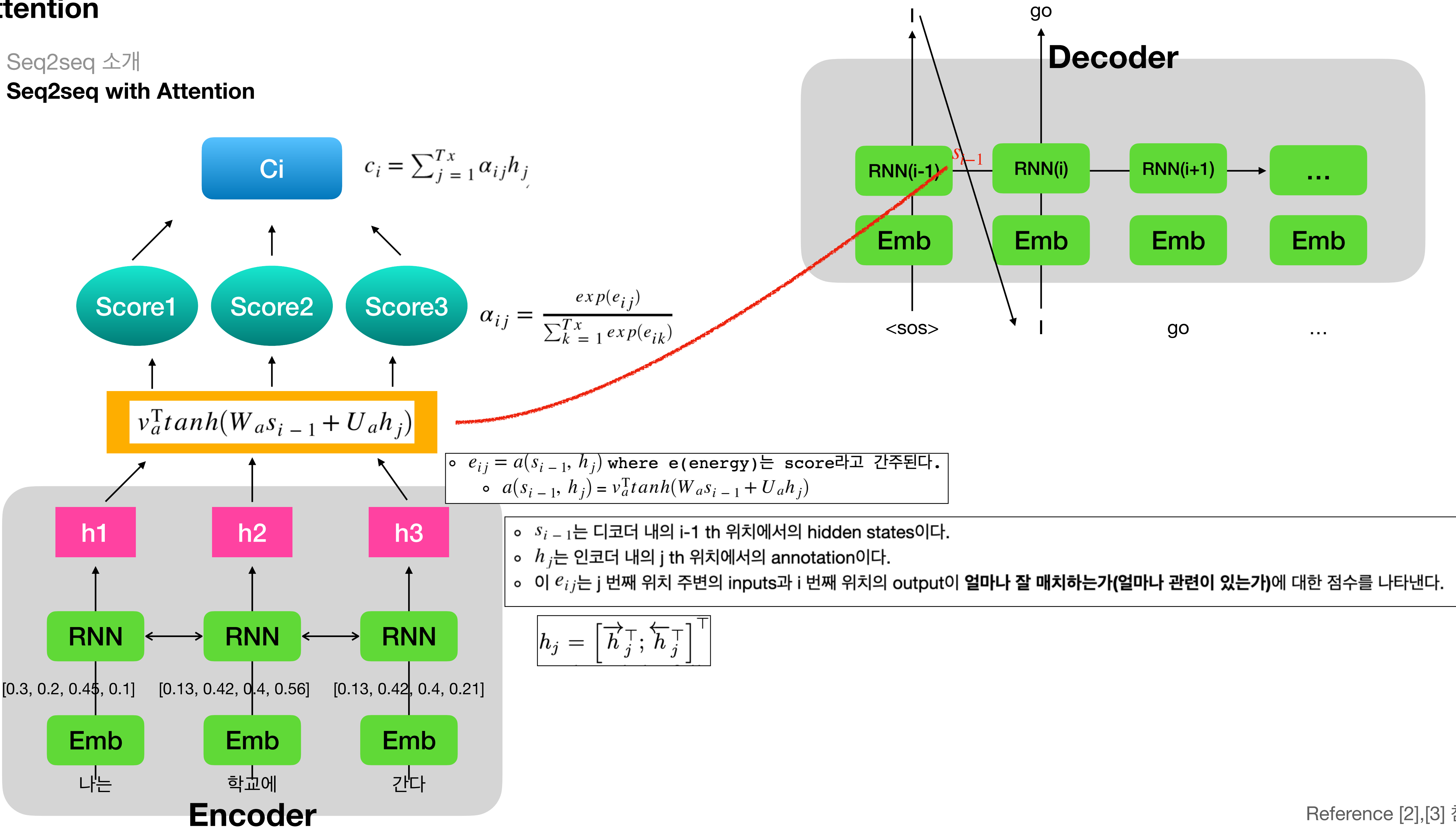
- 어텐션이 첨가된 seq2seq 모델 도입.
- 이것은 디코더 내의 한 타겟 단어를 예측할 때 마다, 인코더 내의 source 문장들 중 어느 입력 단어들과 가장 유관한 지를 탐색
- 이 탐색으로부터 “맥락 벡터context vector”가 만들어진다.



Vanila seq2seq은 입력 문장 길이가 길수록 BLEU 성능 측정 지표가 낮게 나온다.

# Attention

- 1. Seq2seq 소개
- 2. Seq2seq with Attention



# Attention

- 1. Seq2seq 소개
- 2. Seq2seq with Attention

where  $s_i$  is an RNN hidden state for time  $i$ , computed by

$$s_i = f(s_{i-1}, y_{i-1}, c_i).$$



$$p(y_i|y_1, \dots, y_{i-1}, \mathbf{x}) = g(y_{i-1}, s_i, c_i),$$

$$p(\mathbf{y}) = \prod_{t=1}^T p(y_t \mid \{y_1, \dots, y_{t-1}\}, c),$$

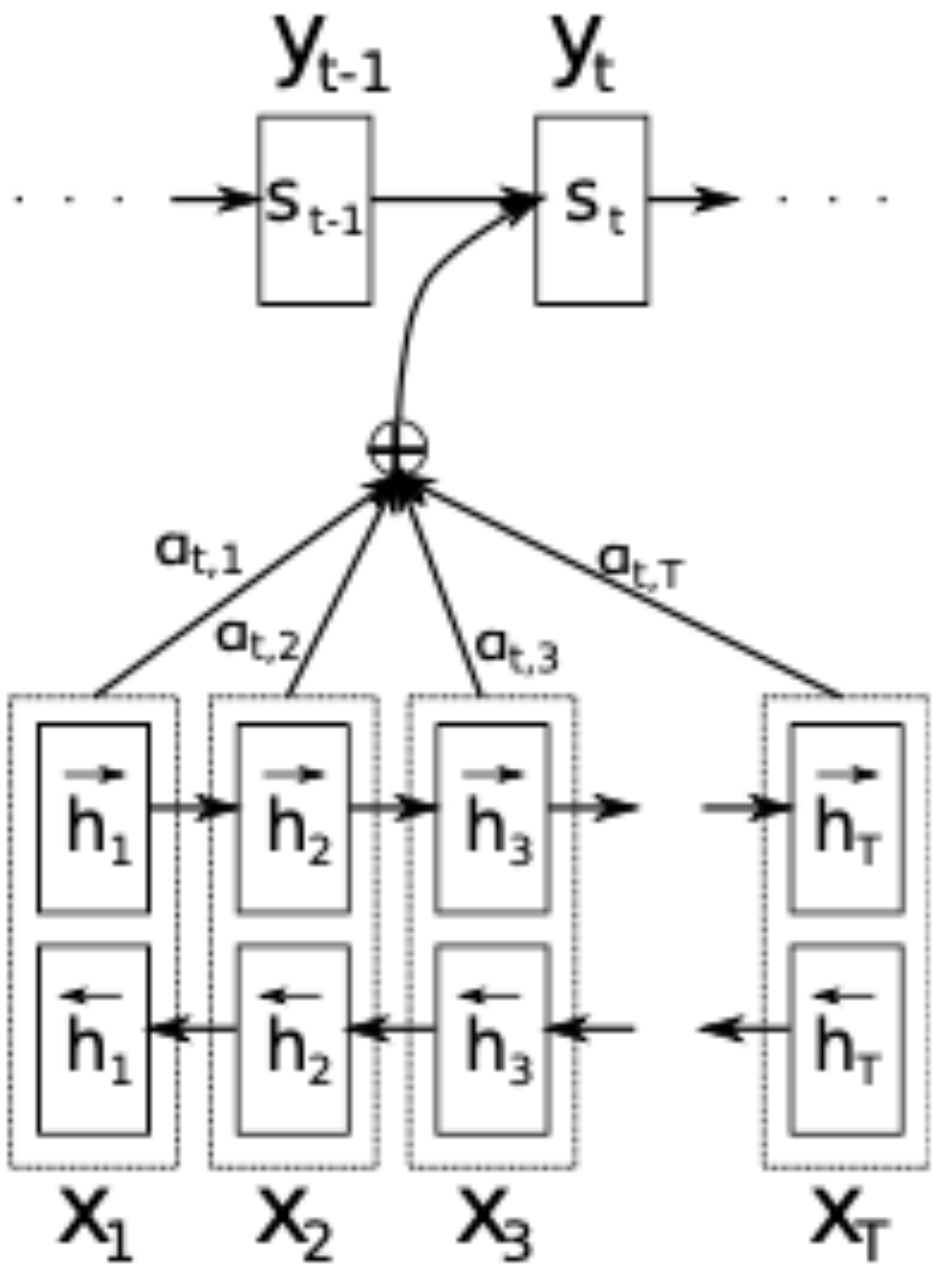
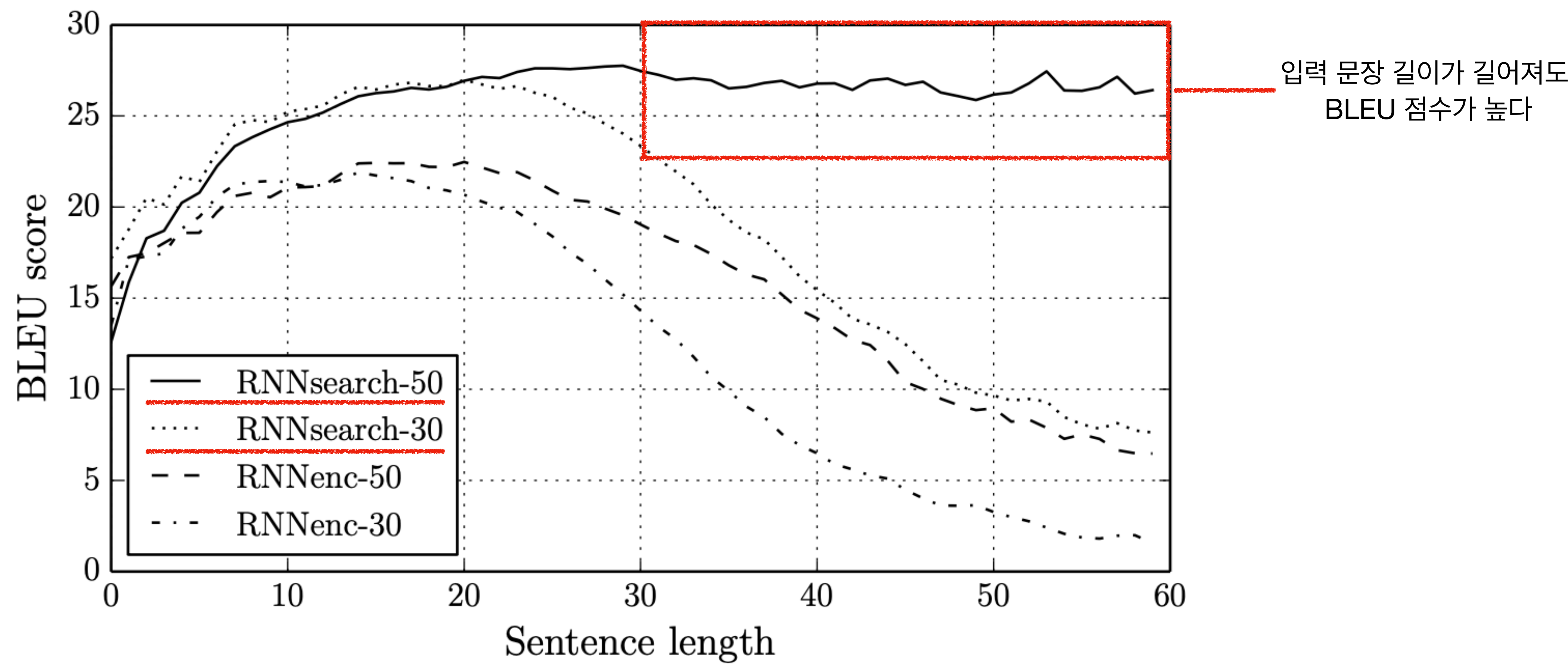


Figure 1: The graphical illustration of the proposed model trying to generate the  $t$ -th target word  $y_t$  given a source sentence  $(x_1, x_2, \dots, x_T)$ .



# Attention

- 1. Seq2seq 소개
- 2. Seq2seq with Attention





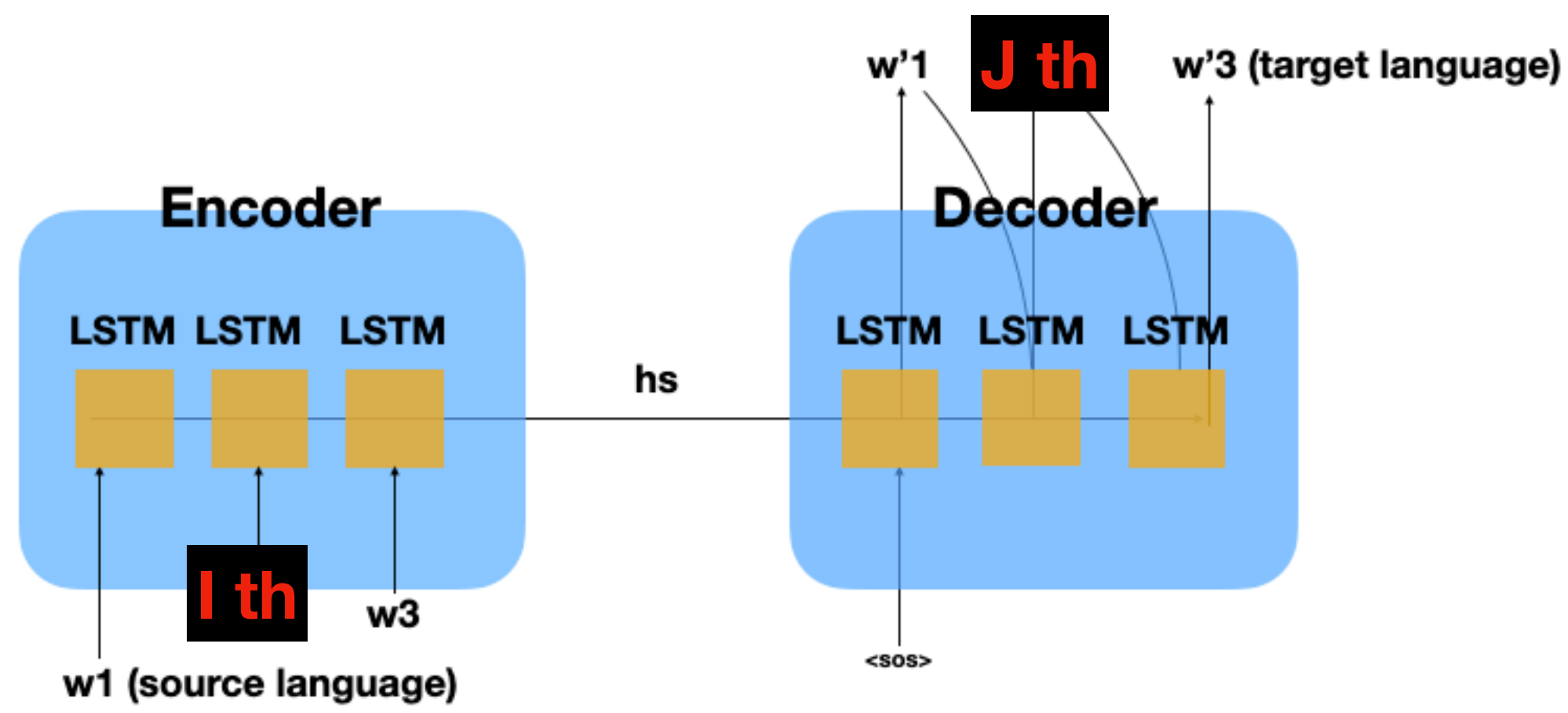
# **Attention 관련 용어 및 개념 설명**

## **1. Soft-alignment VS Hard-alignment**

## **2. Teacher Forcing 교사 강요**

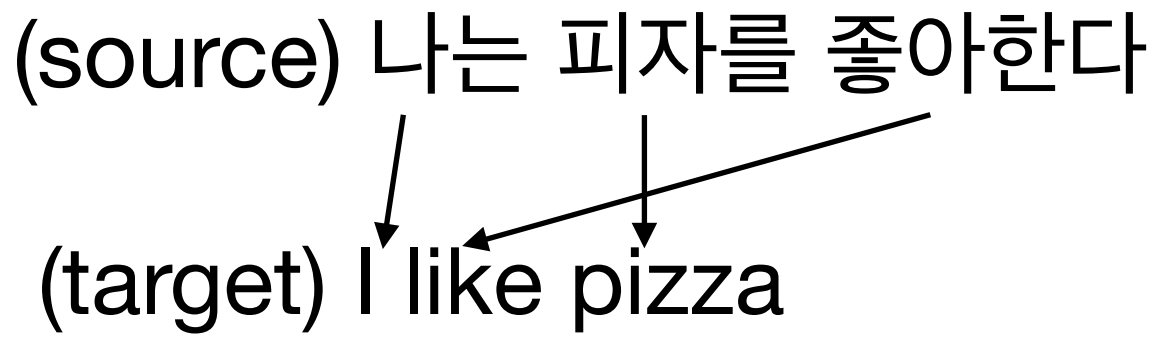
# 1. Soft-alignment VS Hard-alignment

Soft-alignment:



**alignment:** 디코더에서 j th 위치의 입력 단어와 인코더에서 i th 위치 출력이 얼마나 잘 매치되는 지에 대한 점수

**Hard-alignment:** source sentence 내의 각 단어와 target sentence 내의 각 단어를 “하나씩” 정렬(매핑). 이 작업은 사람이 직접 하나씩 매핑해준다는 점에서 비효율적이다.



The strength of the soft-alignment, opposed to a **hard**-alignment, is evident, for instance, from Fig. 3 (d). Consider the source phrase [the man] which was translated into [l' homme]. Any **hard** alignment will map [the] to [l'] and [man] to [homme]. This is not helpful for translation, as one must consider the word following [the] to determine whether it should be translated into [le], [la], [les] or [l']. Our soft-alignment solves this issue naturally by letting the model look at both [the] and

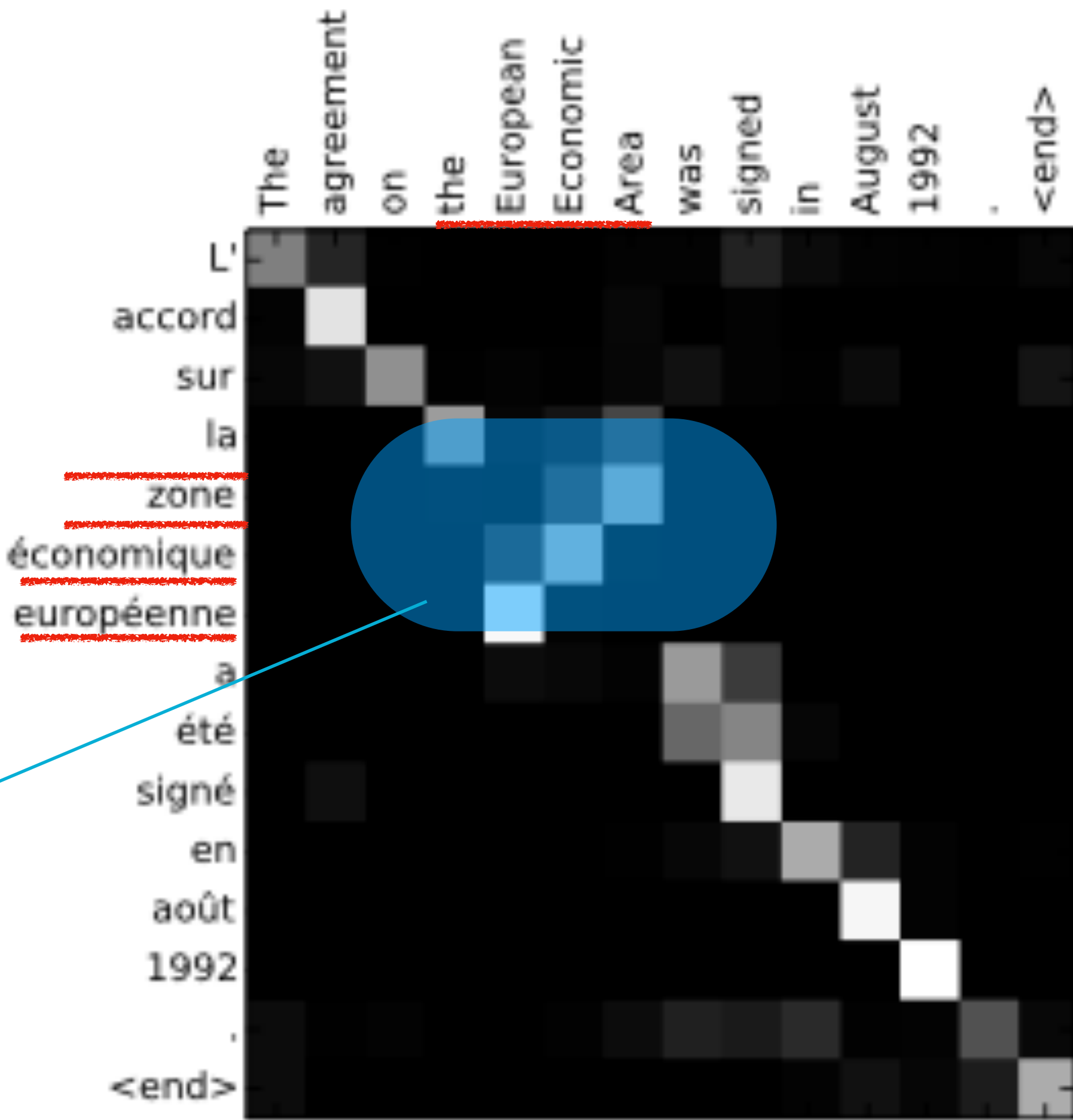
# 1. Soft-alignment VS Hard-alignment

## Soft-alignment (Attention):

- 하나의 타겟 단어가 모든 소스 단어를 확인
- 모델이 학습을 함으로써 타겟 문장과 소스 문장 간의 어순이 다르다는 점도 인지함
- 더 만족스러운 번역 결과를 제공

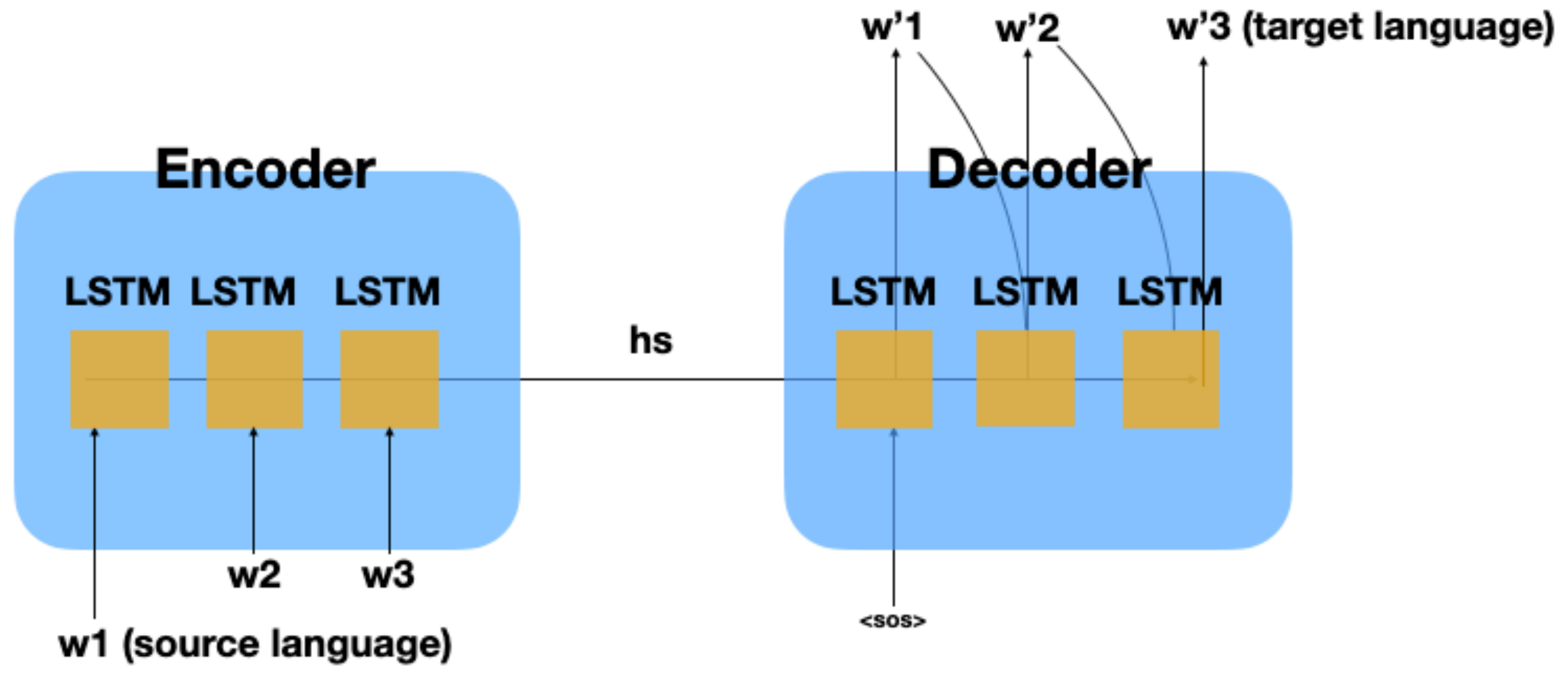
- 영어 -> 프랑스어로 번역하는 과정에서,  $\text{alignment}(\alpha)$ 는 타겟 단어 'zone', 'économique', 'européenne'이 각각 예측될 때 마다 그것과 연관된 소스 단어를 조사함과 동시에 다음에 나오는 소스 단어도 조사한다.

- 이것은 모델이 어구 'the European Economic Area'로부터 어구 'la zone économique européenne'로 — 이 둘의 어순이 다를지라도 — 올바르게 어순의 배열을 정렬해서 예측한다는 것을 보여 준다.



(a)

## 2. Teacher Forcing 교사 강요



기본 작동: seq2seq의 디코더에서 1st LSTM의 출력값(예측된 타겟 단어)은 2nd LSTM의 입력값이 되고, 그 다음도 마지막 출력까지 동일한 기전으로 작동한다.

**하지만 Training 단계에서의 디코더 작동 방식과, Test 단계에서의 디코더 작동 방식은 보통 다르다!**

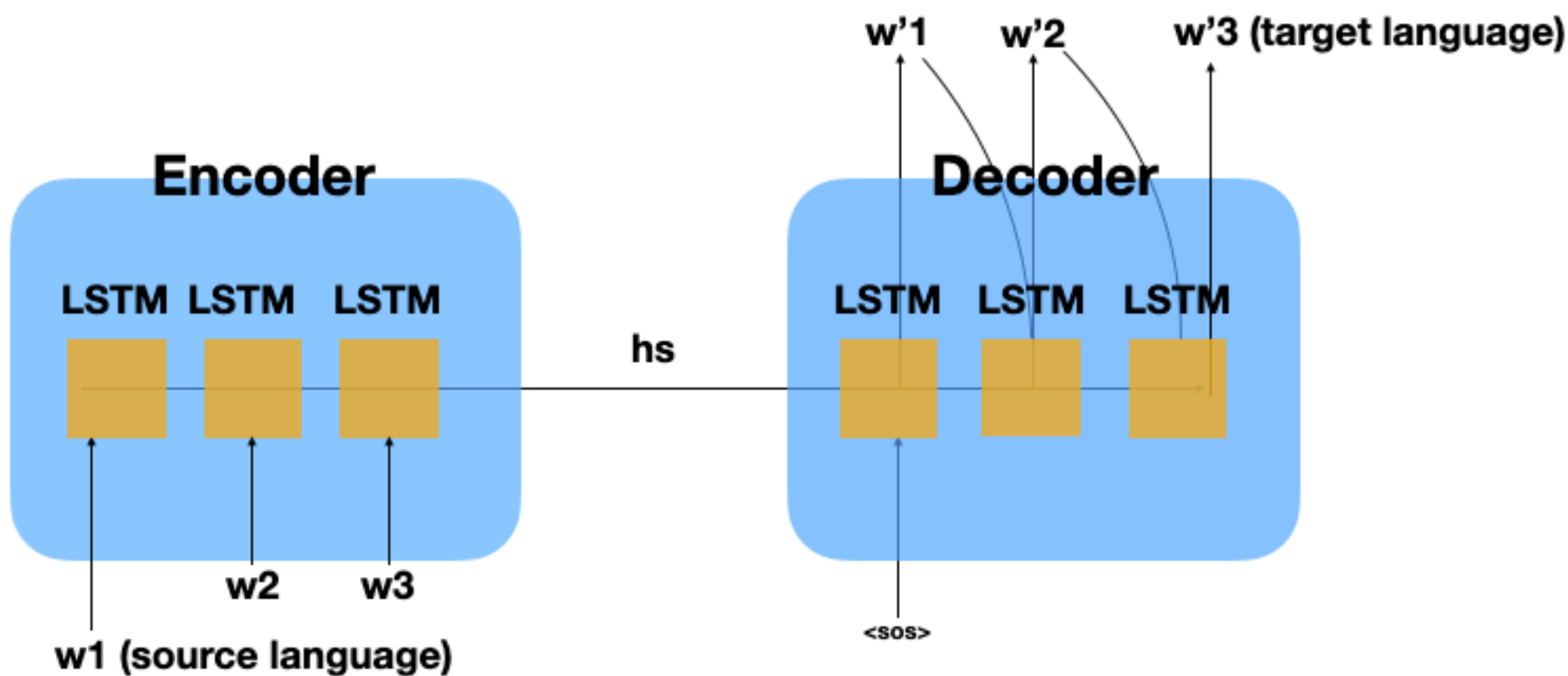
## 2. Teacher Forcing 교사 강요

### Training 단계 (교사 강요 o)

- 교사 강요를 사용 시, 디코더 내의 LSTM은 이전 timestep에서의 예측값을 해당 timestep의 입력 단어로 받지 x
- 훈련 단계에서 모델은 디코더에서 잘못된 예측을 할 수 있기 때문에 디코더 내의 각 timestep의 LSTM에 실제 값(실제 단어)을 입력한다.

### Test 단계 (교사 강요 x)

- 학습이 끝난 뒤 모델을 테스트할 때는 교사 강요가 아니라 위에서 설명한 기본적인 seq2seq 작동 방식을 사용
- 즉, 디코더에서 이전 timestep에서 예측한 단어가 해당 timestep의 입력으로 들어온다.



I left.	Je suis partie.
I lied.	J'ai menti.
I lost.	J'ai perdu.
I paid.	J'ai payé.
I'm 19.	J'ai 19 ans.
I'm OK.	Je vais bien.
I'm OK.	Ça va.

# References

- [1] Ilya Sutskever et al, 'Sequence to Sequence Learning with Neural Networks'
- [2] Dzmitry Bahdanau et al, 'NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE', 2014
- [3] <https://github.com/musicjae/NLP/blob/master/Attention/Attention.ipynb> (발표자 깃허브)
- [4] [https://github.com/musicjae/NLP/tree/master/Attention/seq2seq\\_with\\_attention](https://github.com/musicjae/NLP/tree/master/Attention/seq2seq_with_attention) 코드