

Seq2Seq (Sequence to Sequence)

자연어처리 텍스트마이닝

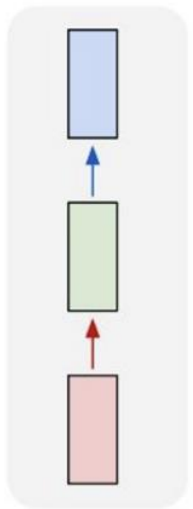
Seq2Seq

(Sequence to Sequence)

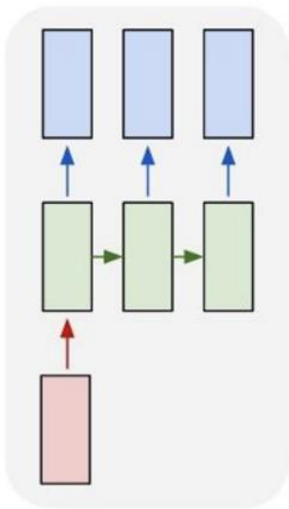
<https://arxiv.org/pdf/1409.3215.pdf>

RNN 활용

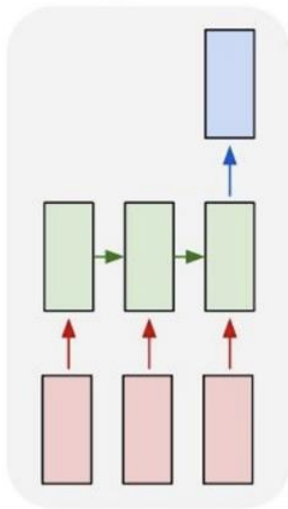
one to one



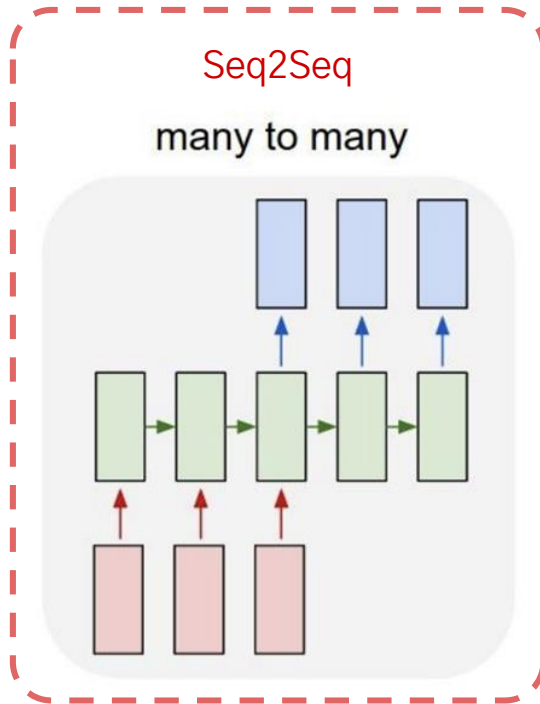
one to many



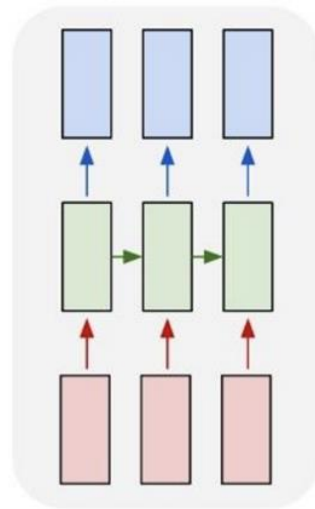
many to one



many to many



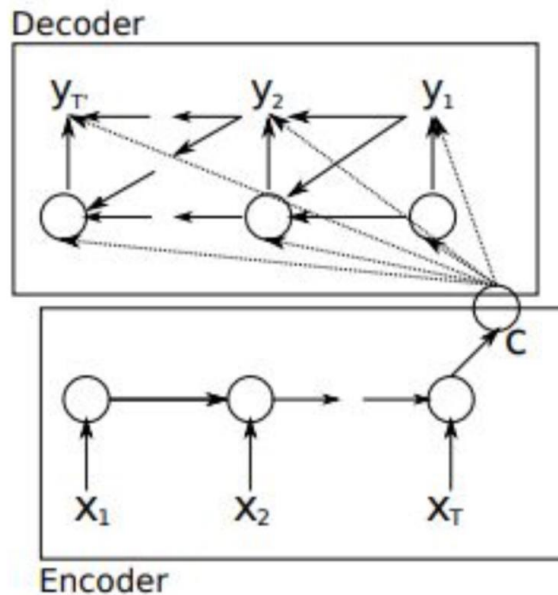
many to many



http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture10.pdf

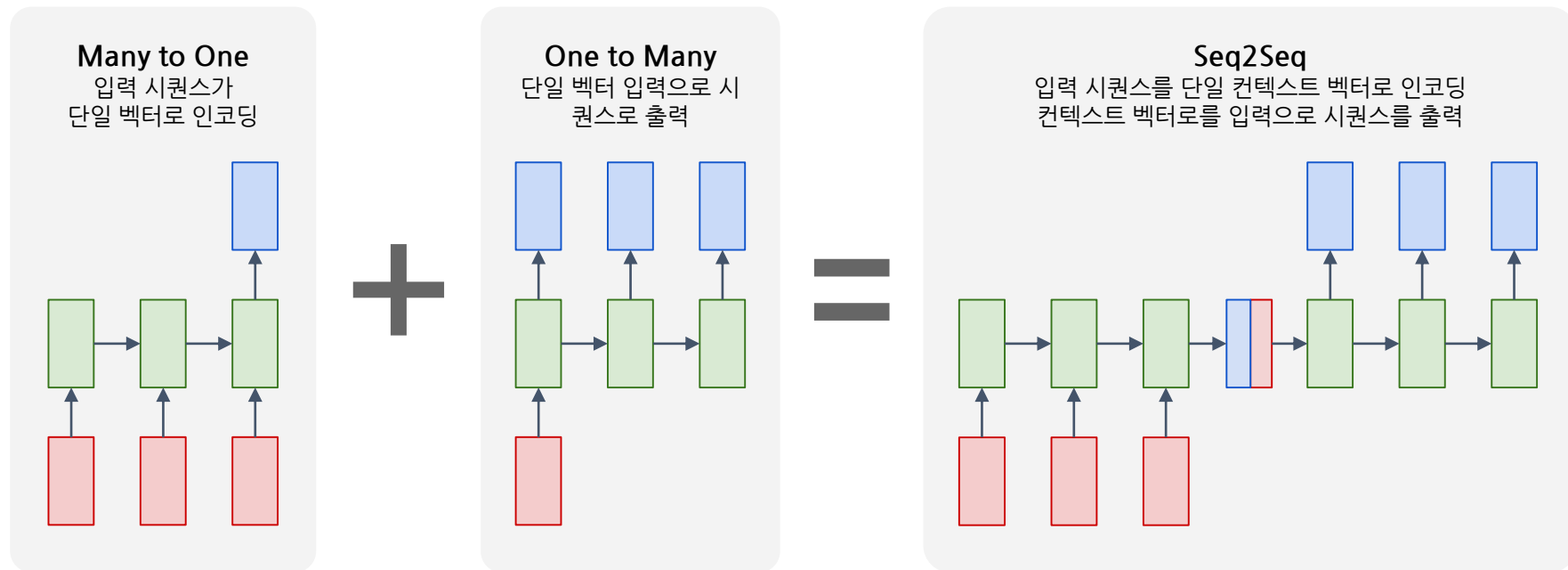
Seq2Seq (1)

Encoder-Decoder



<https://arxiv.org/pdf/1406.1078.pdf>

Seq2Seq (2)



Seq2Seq (3)

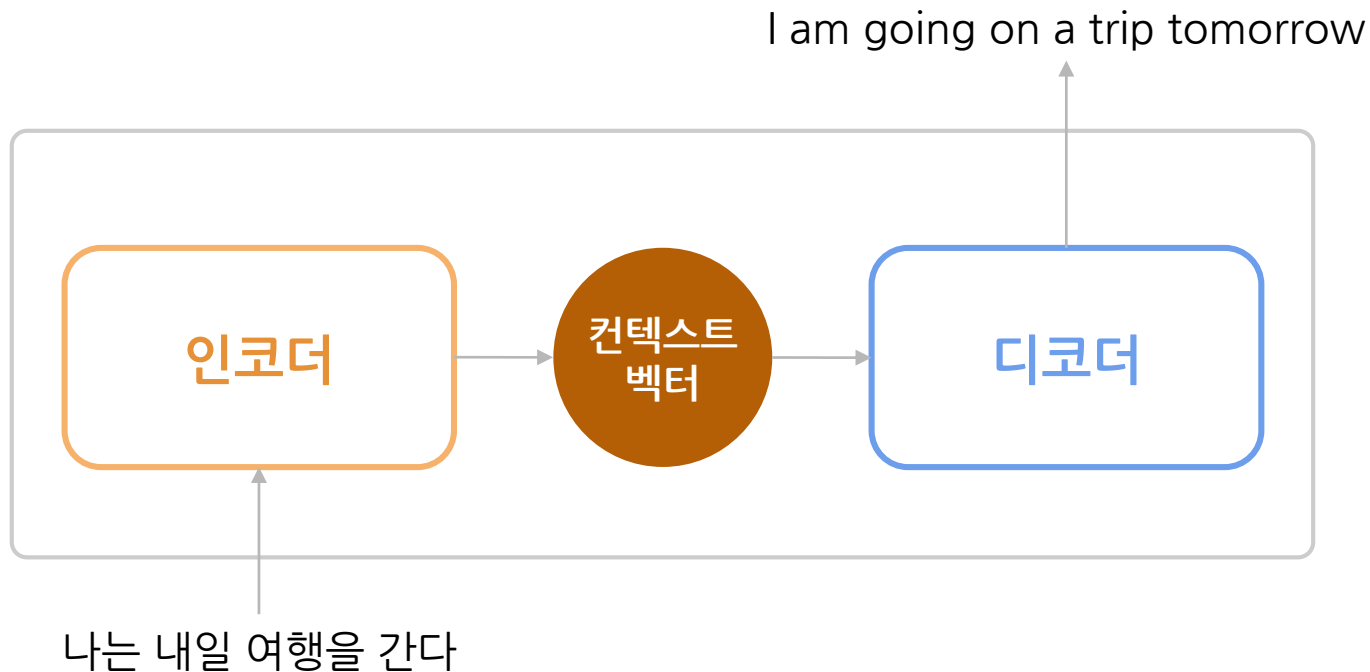
1 2 3 4
나는 / 내일 / 여행을 / 간다.

1 2 3 4 5 6 7
I / am / going / on / a / trip / tomorrow.

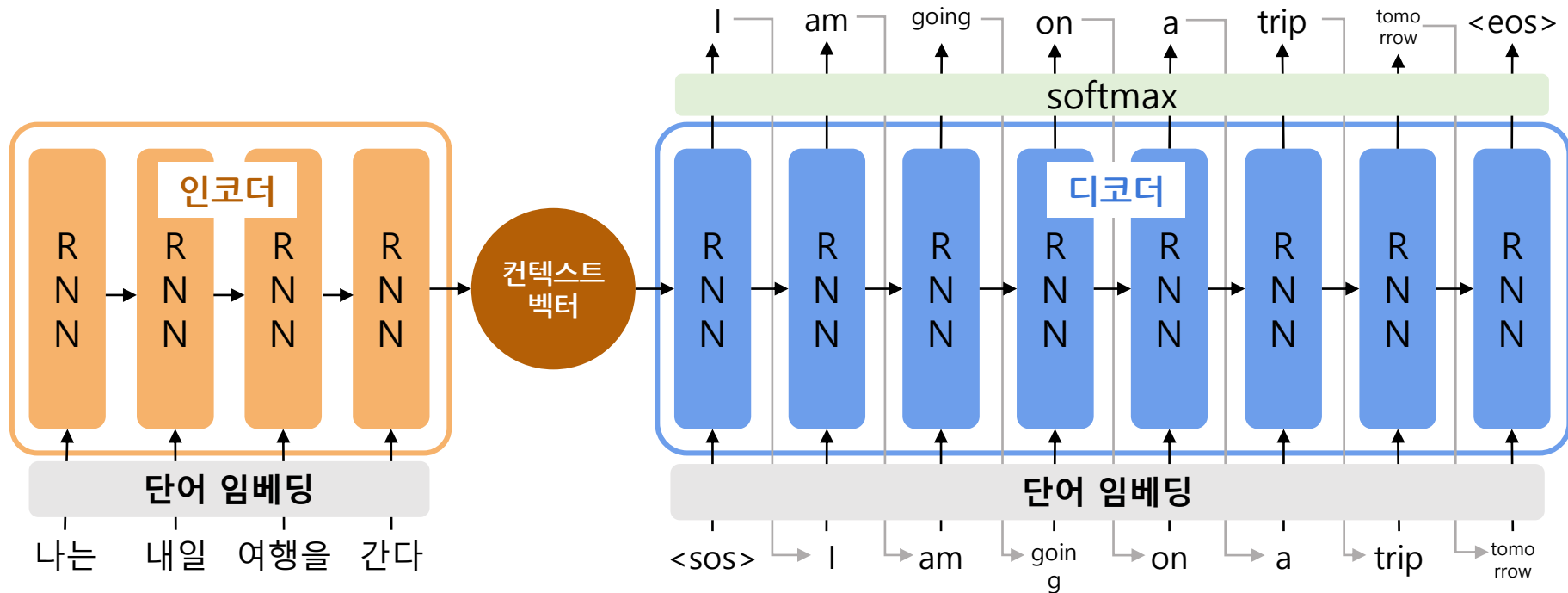
Seq2Seq (3)



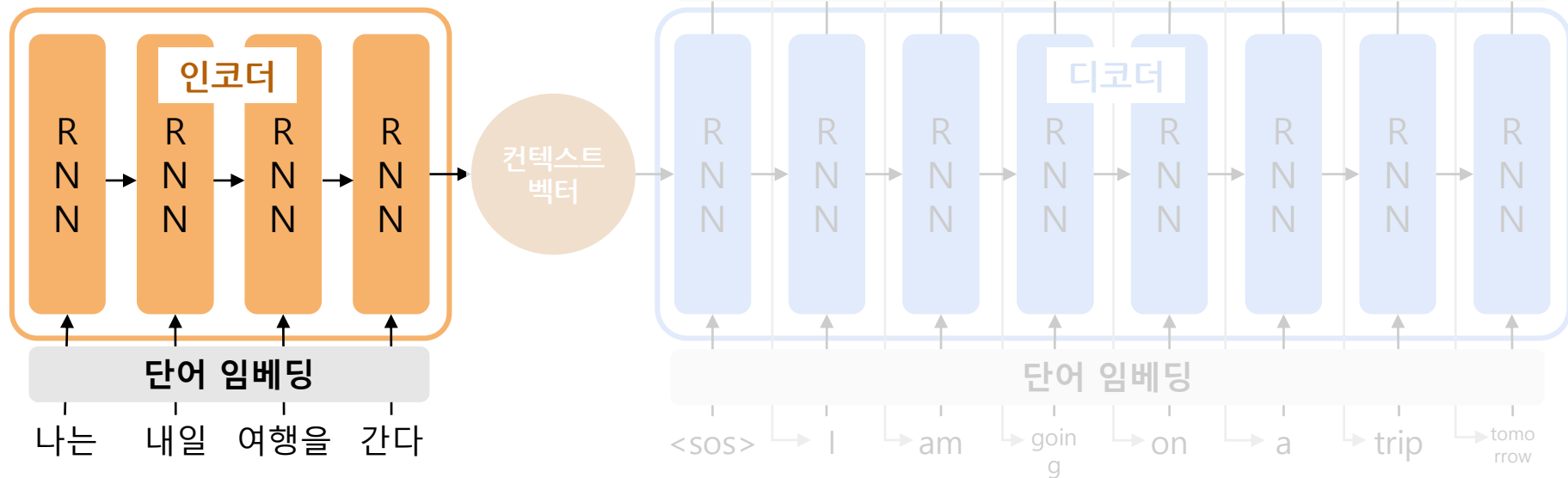
Seq2Seq (4)



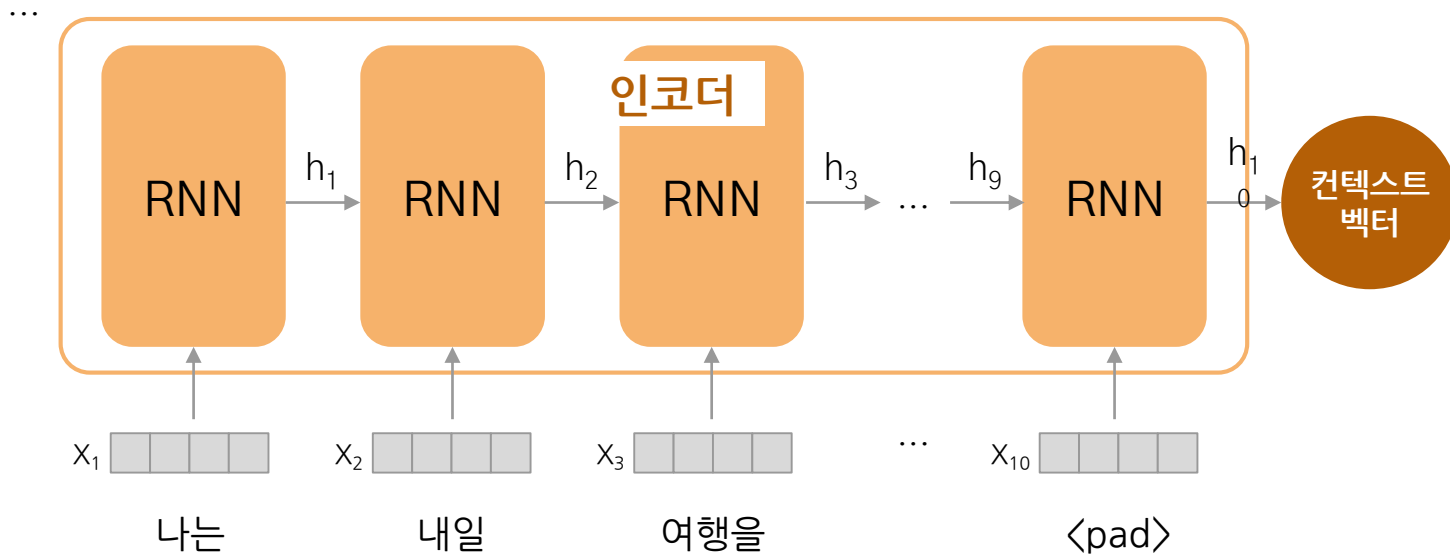
Seq2Seq (5)



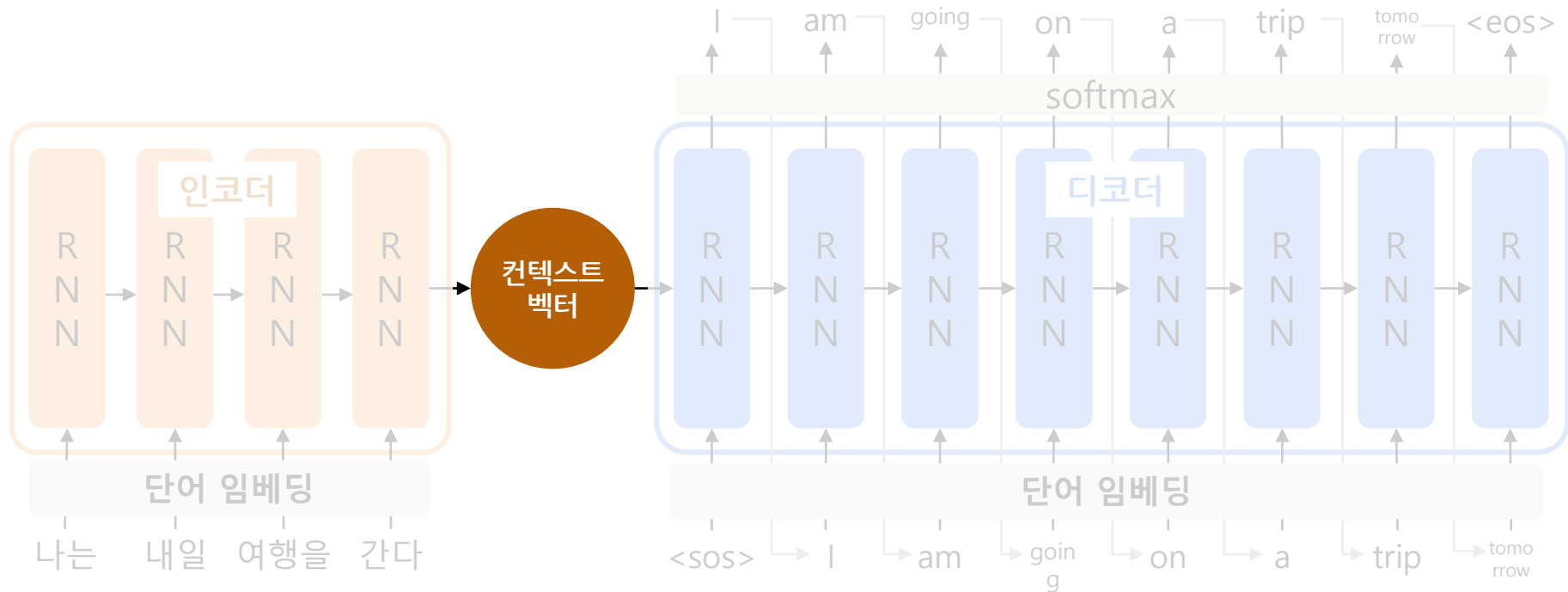
인코더



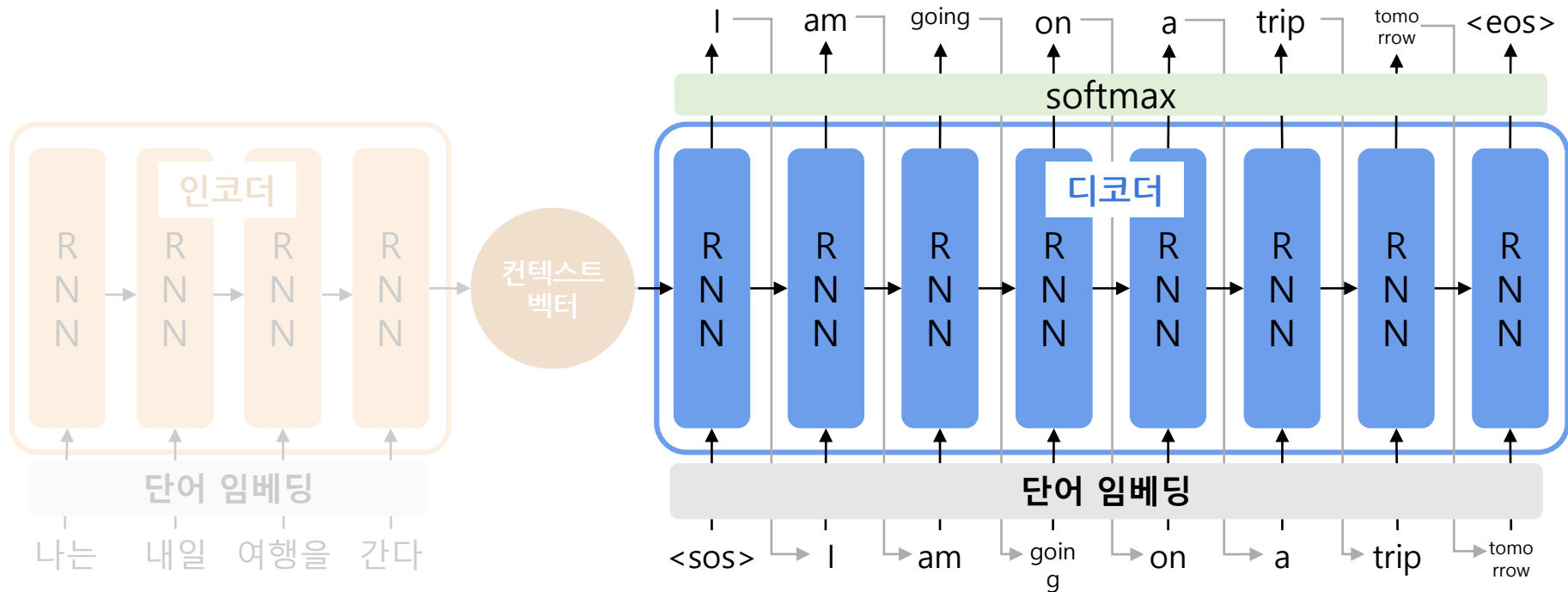
인코더



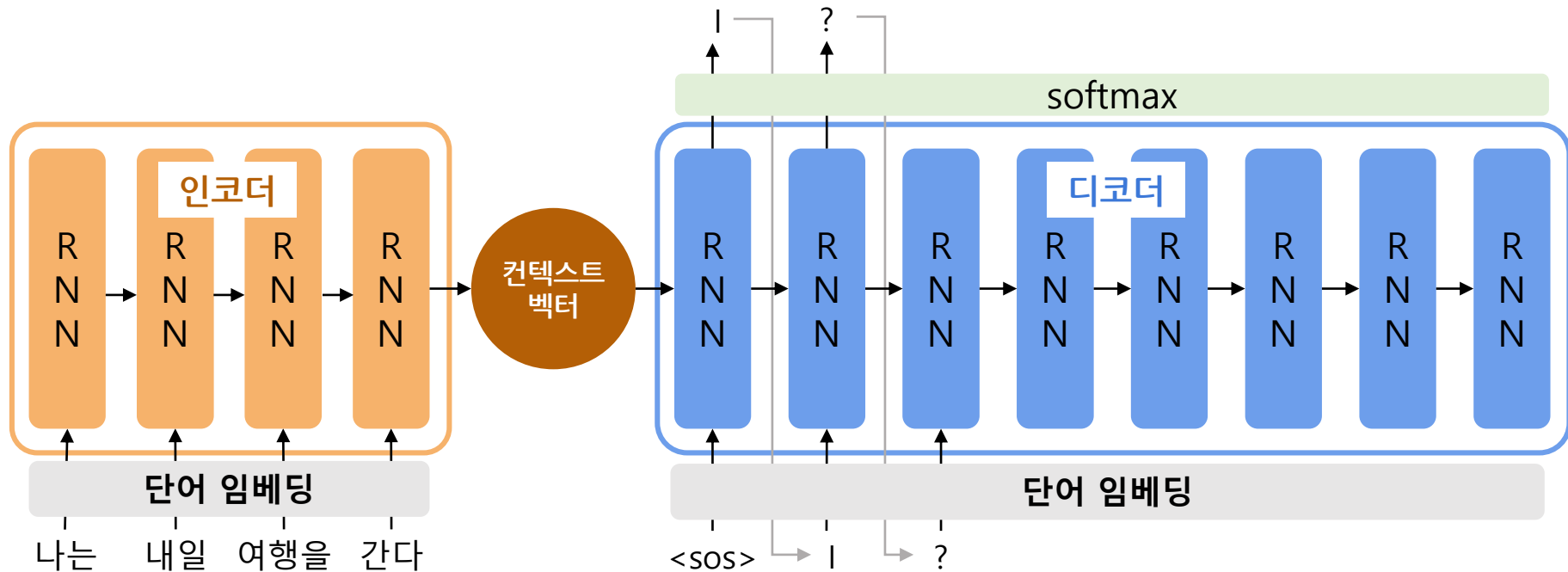
컨텍스트 벡터



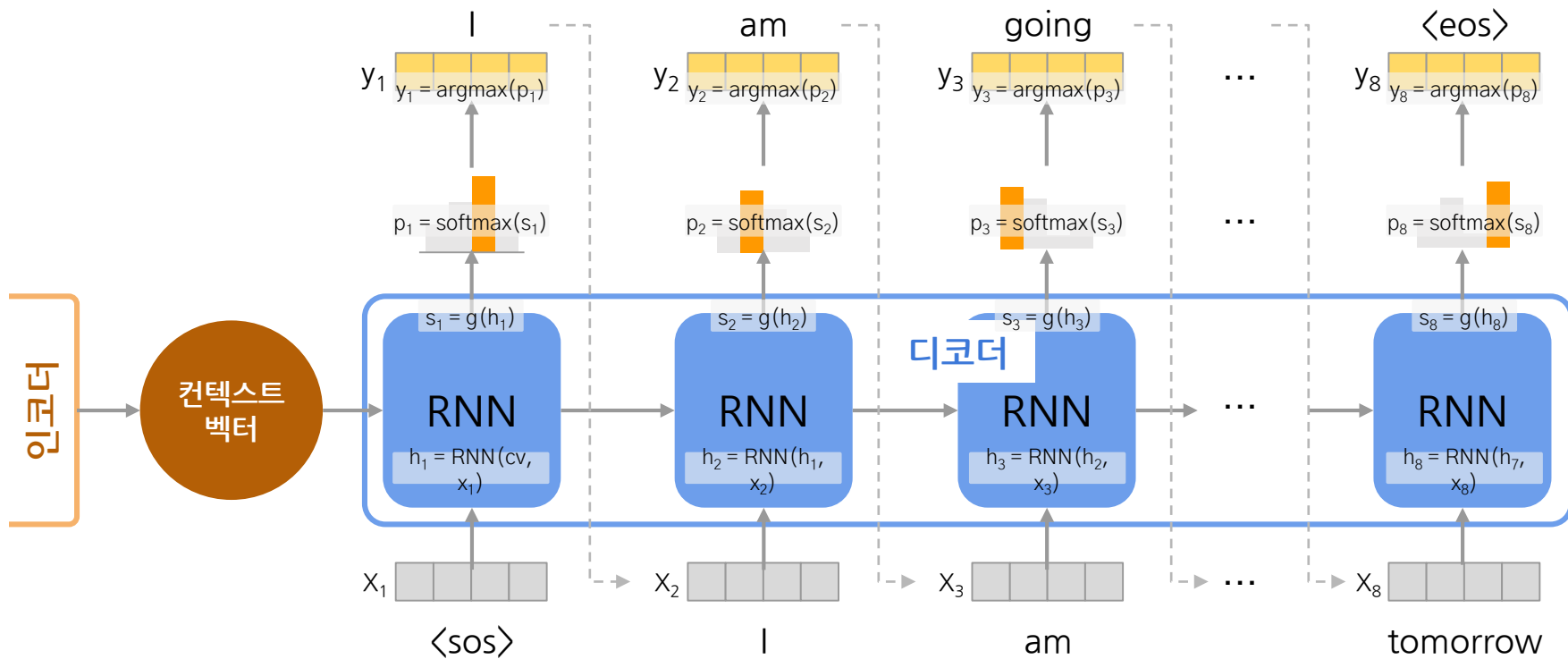
디코더 - greedy decoding



디코더



디코더



디코더

$$y_t = \operatorname{argmax}(p_t)$$

최대확률을 가지는 단어 선택

$$p_t = \operatorname{softmax}(s_t)$$

확률 분포 계산

$$s_t = g(h_t)$$

RNN 출력 계산

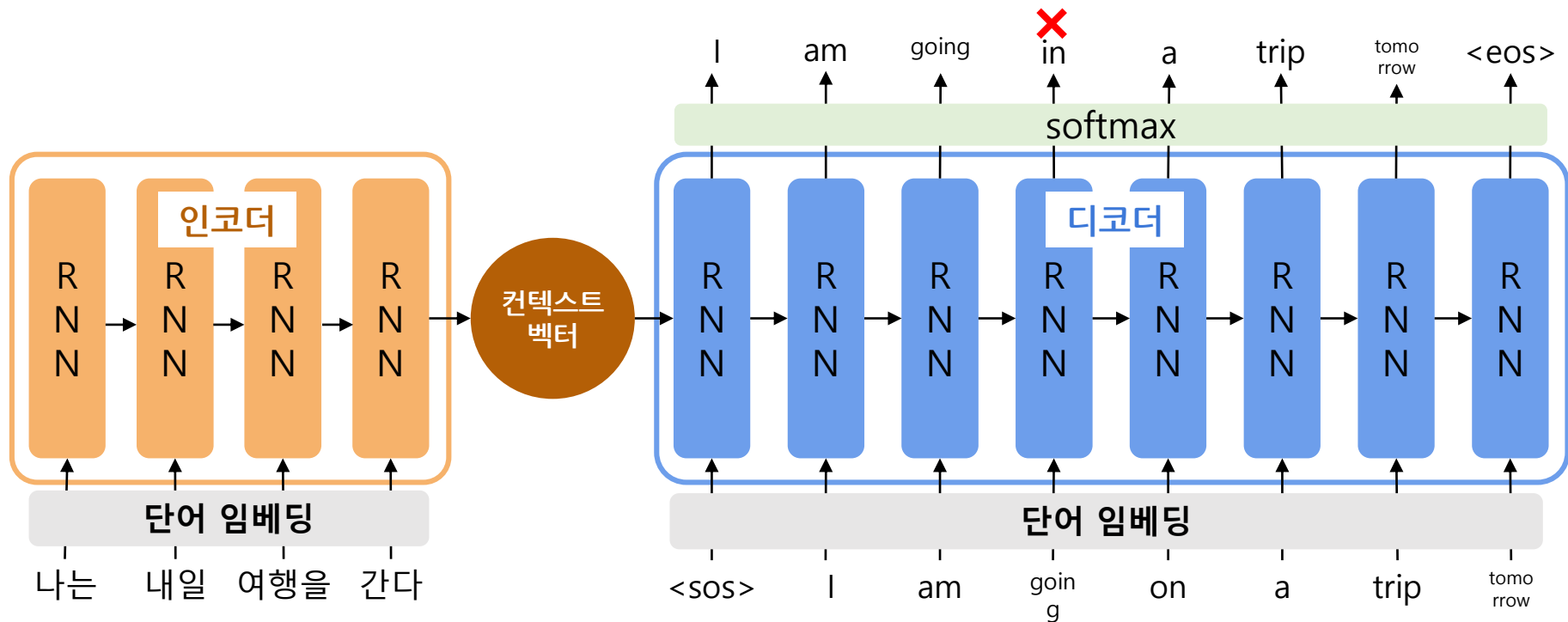
$$h_t = \operatorname{RNN}(h_{t-1}, x_t)$$

Hidden State 계산

h_t : t시점 Hidden State

x_t : t시점 입력

학습 - Teacher Forcing



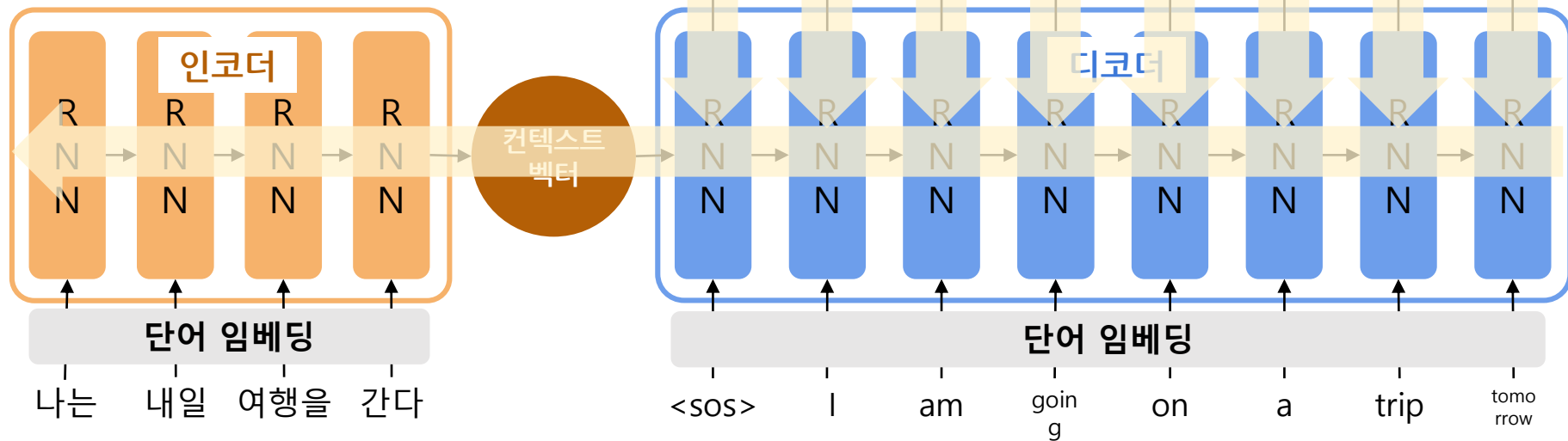
학습

$$-\sum_{j=1}^K y_{t,j} \log \hat{y}_{t,j}$$

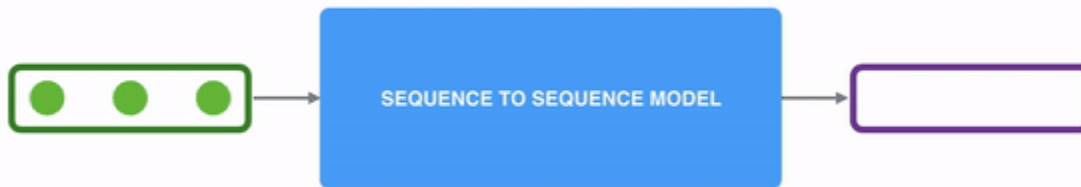
cross-entropy

$$J_t = -\sum_{j=1}^K y_{t,j} \log \hat{y}_{t,j}$$

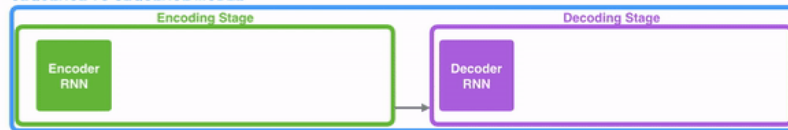
$$J = \frac{1}{T} \sum_{t=1}^T J_t$$



Seq2Seq (3)

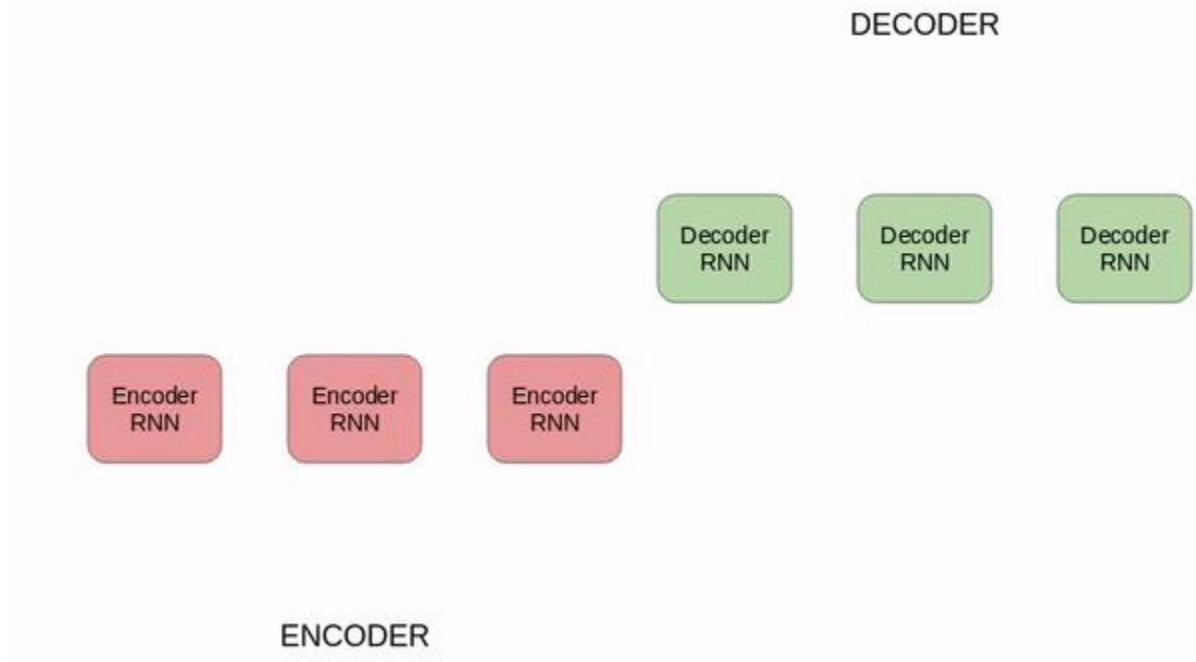


Neural Machine Translation SEQUENCE TO SEQUENCE MODEL



Je suis étudiant

Seq2Seq (4)

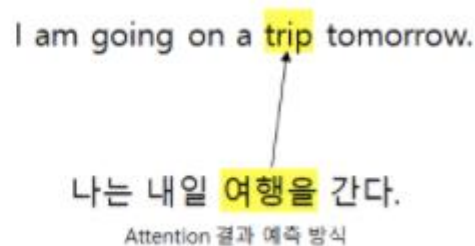


Seq2Seq with Attention

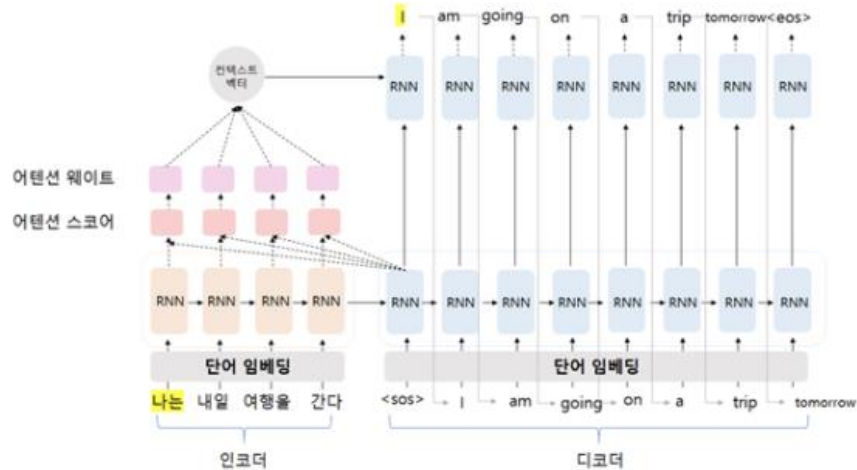
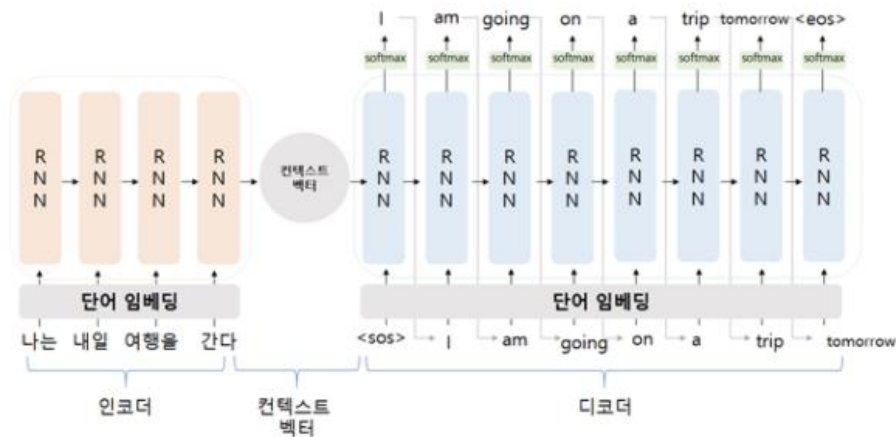
(Effective Approaches to Attention-based Neural Machine Translation)

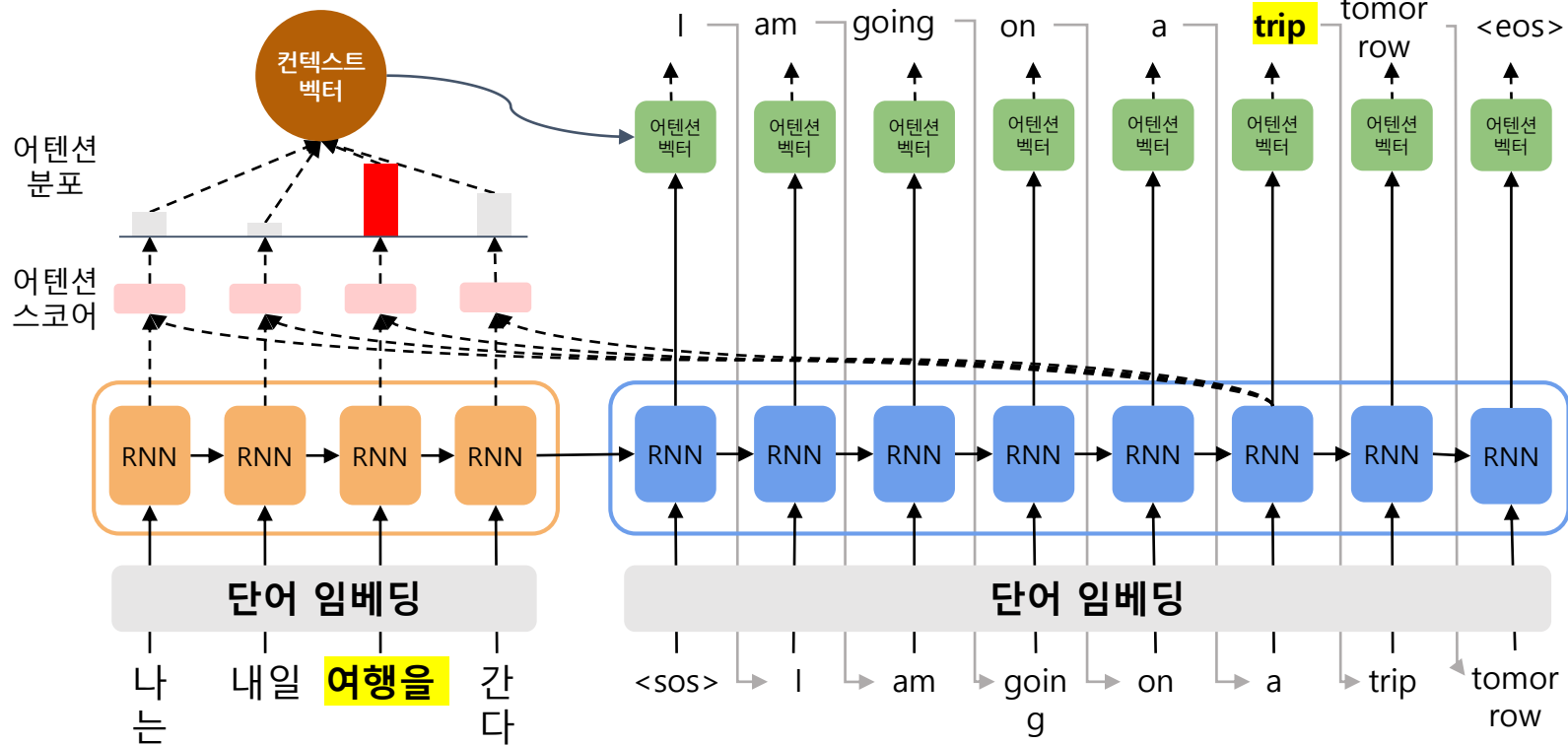
<https://arxiv.org/pdf/1508.04025.pdf>

Seq2Seq with Attention (1)

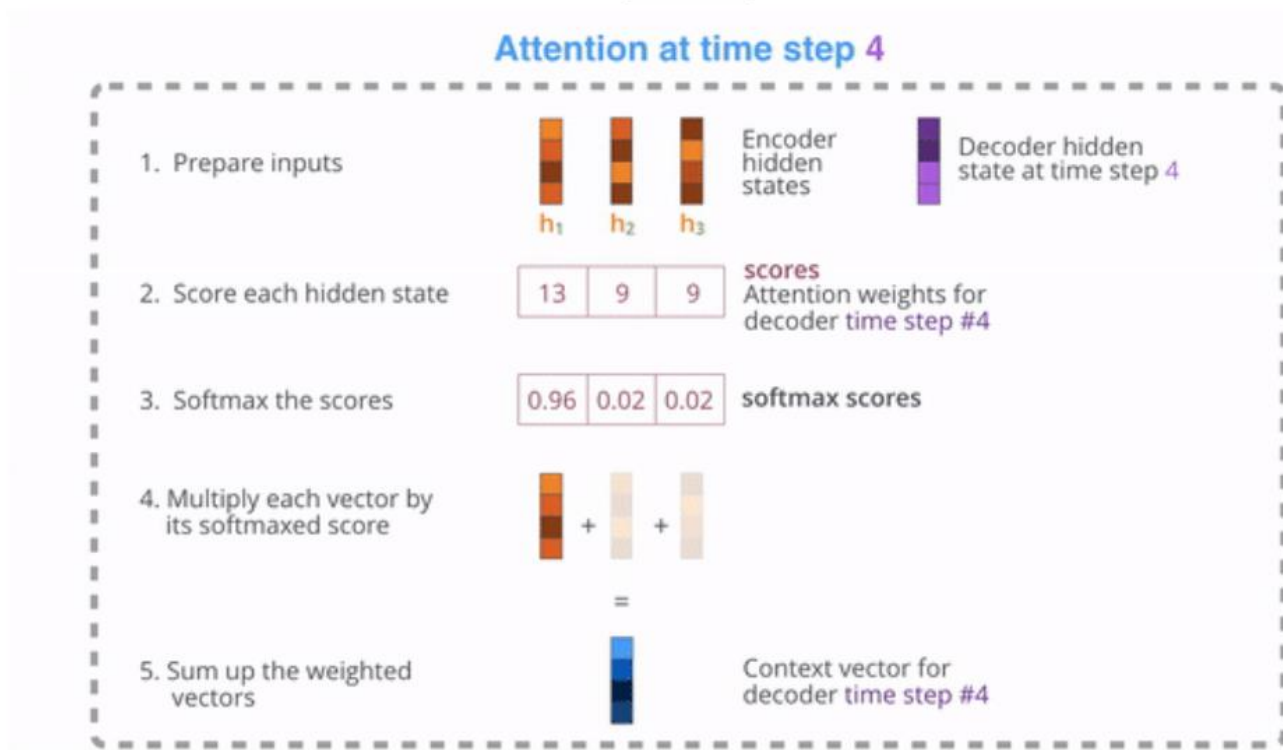


Seq2Seq with Attention (1)

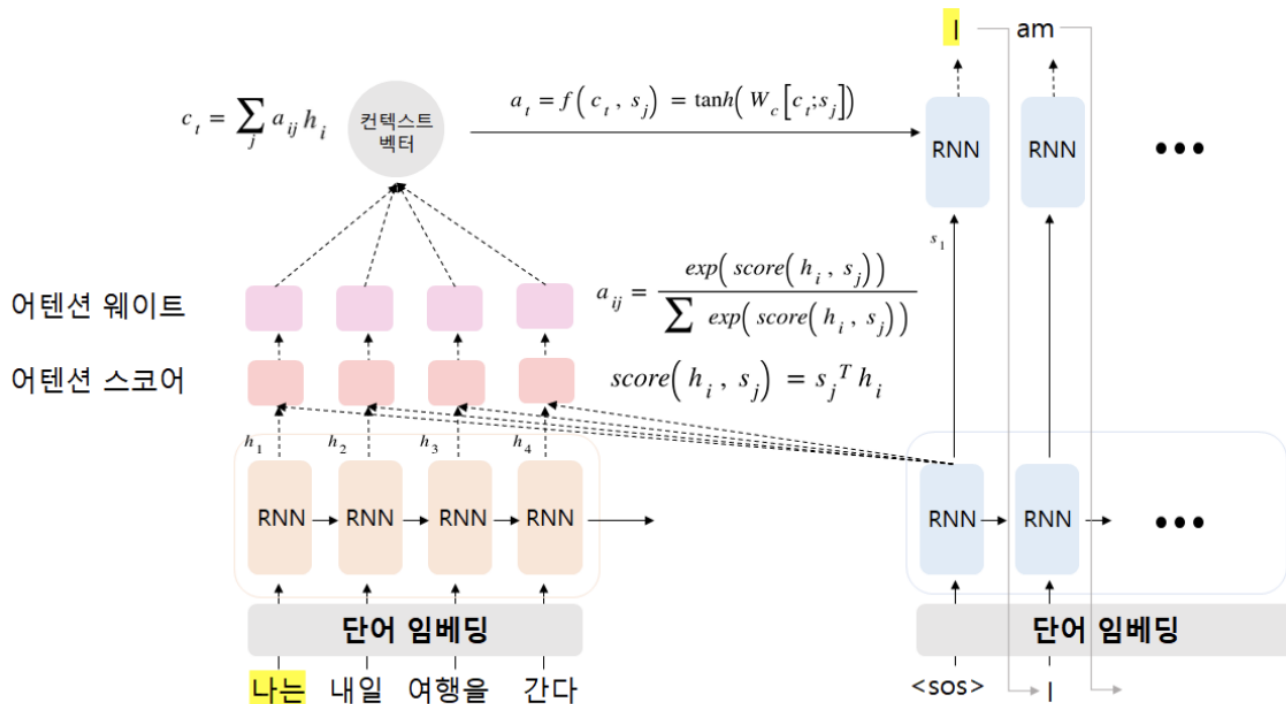




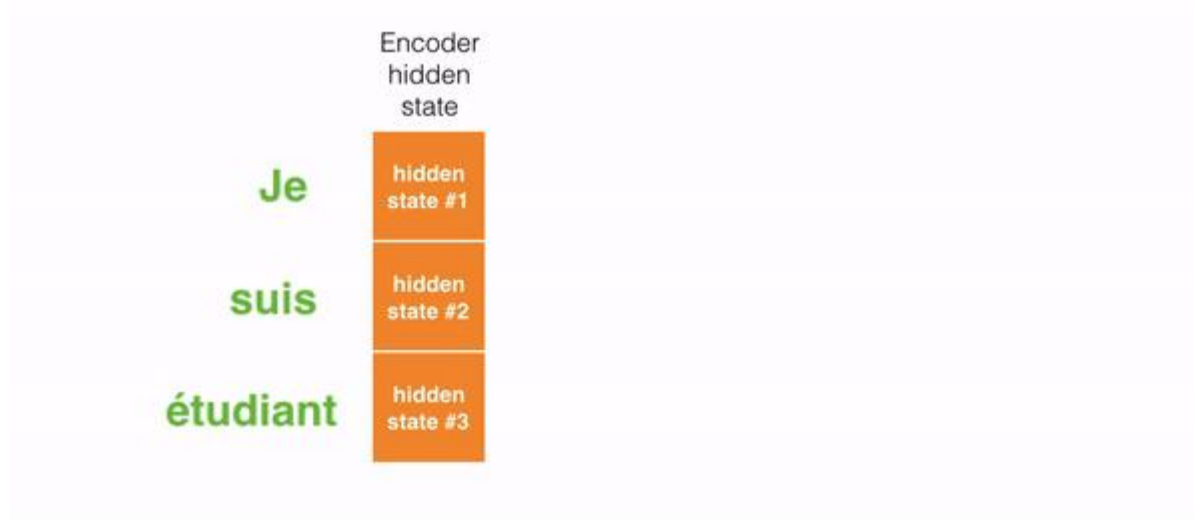
Seq2Seq with Attention (2)



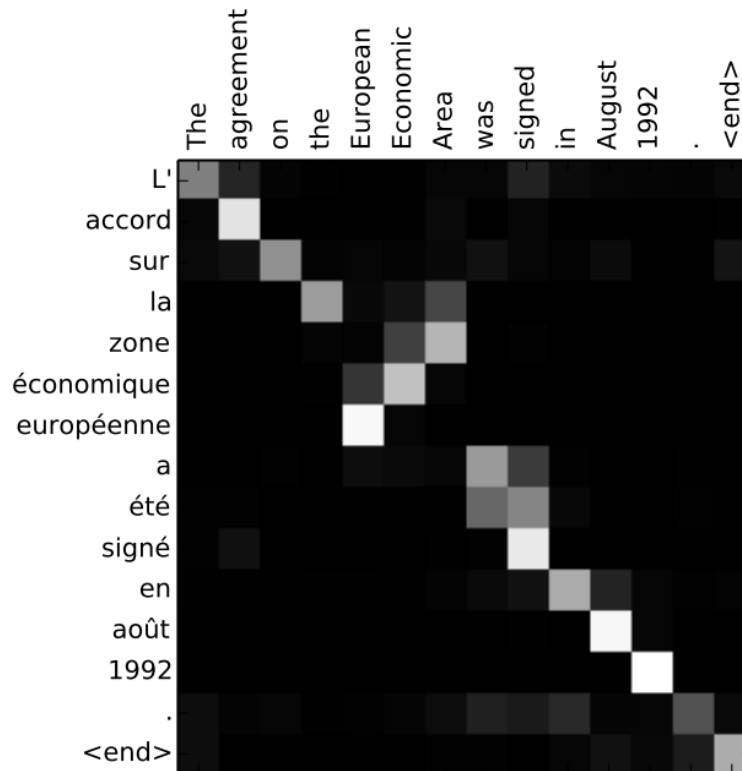
Seq2Seq with Attention (2)



Seq2Seq with Attention (3)



Seq2Seq with Attention (4)



BLEU

$$BLEU = \min\left(1, \frac{\text{predict length}}{\text{reference length}}\right) \left(\prod_{i=1}^4 \text{precision}_i\right)^{\frac{1}{4}}$$