# Sequence to Sequence

2022.02.16

지능융합학과 정민지

# Contents

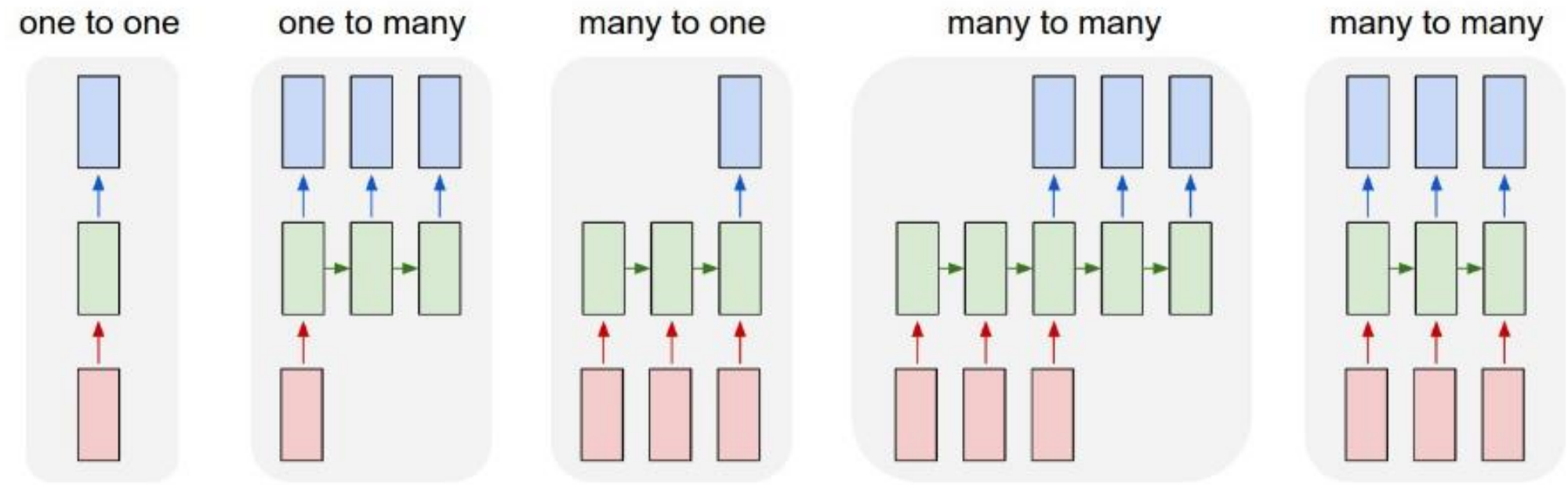
1. Sequence to Sequence

2. Improving Sequence to Sequence

# 1. Sequence to Sequence

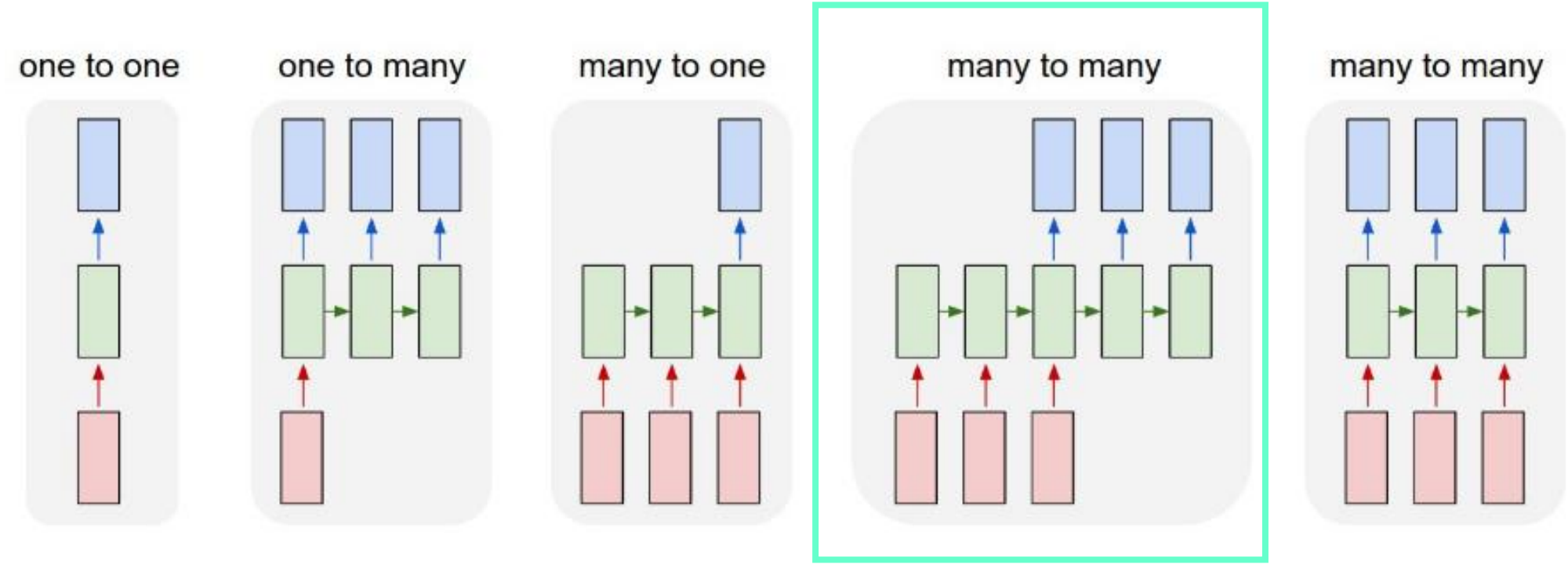
# Recall: RNN





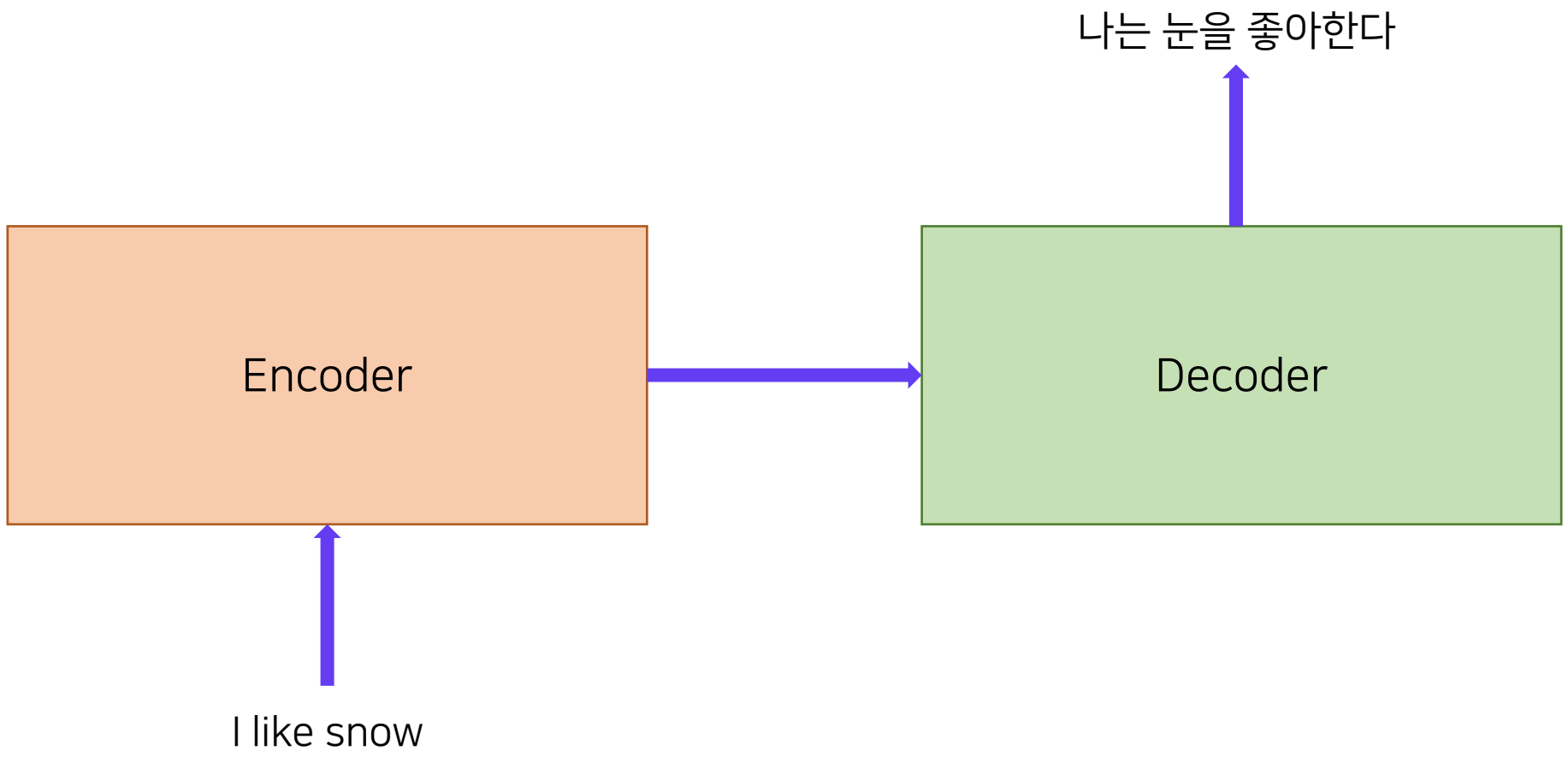
# Recall: RNN





# Seq2Seq model

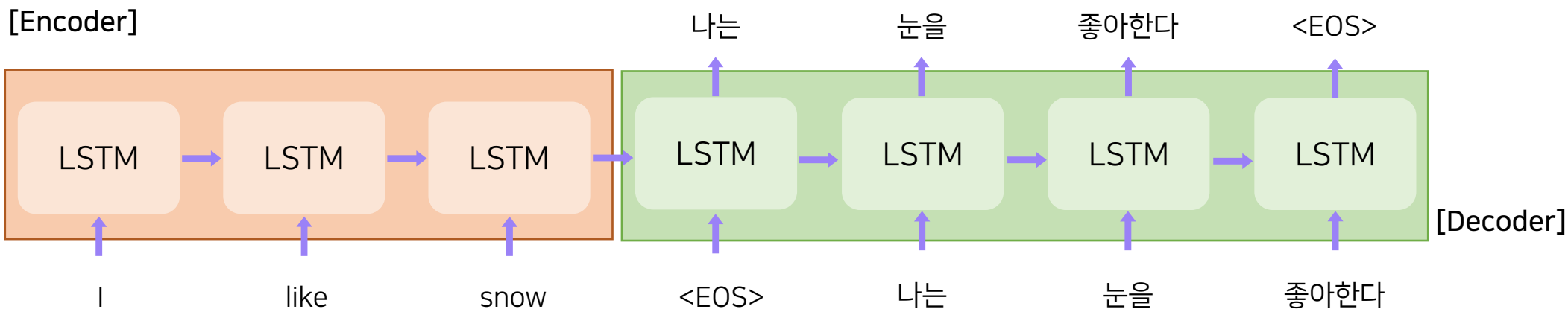




# Seq2Seq model

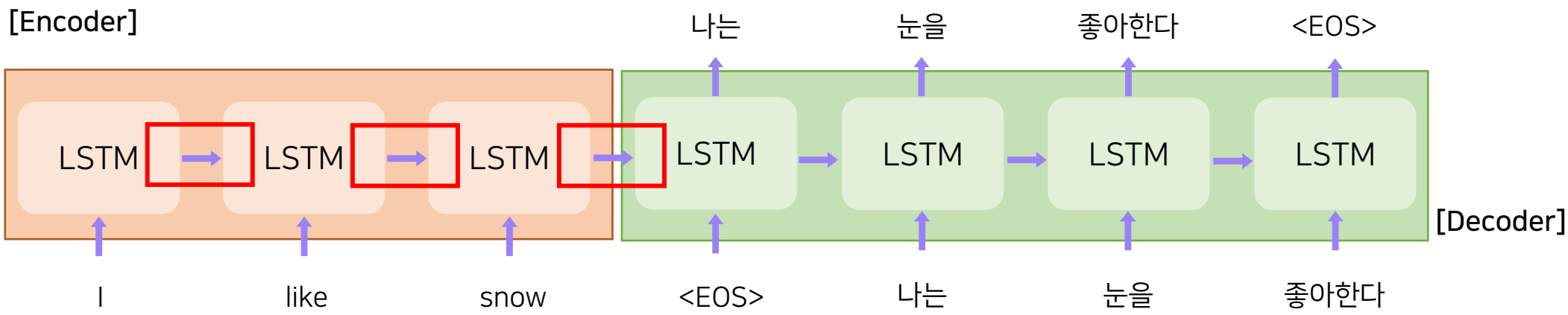


- 기계 번역
- 챗봇 다음 대화 생성 등…



# Seq2Seq model

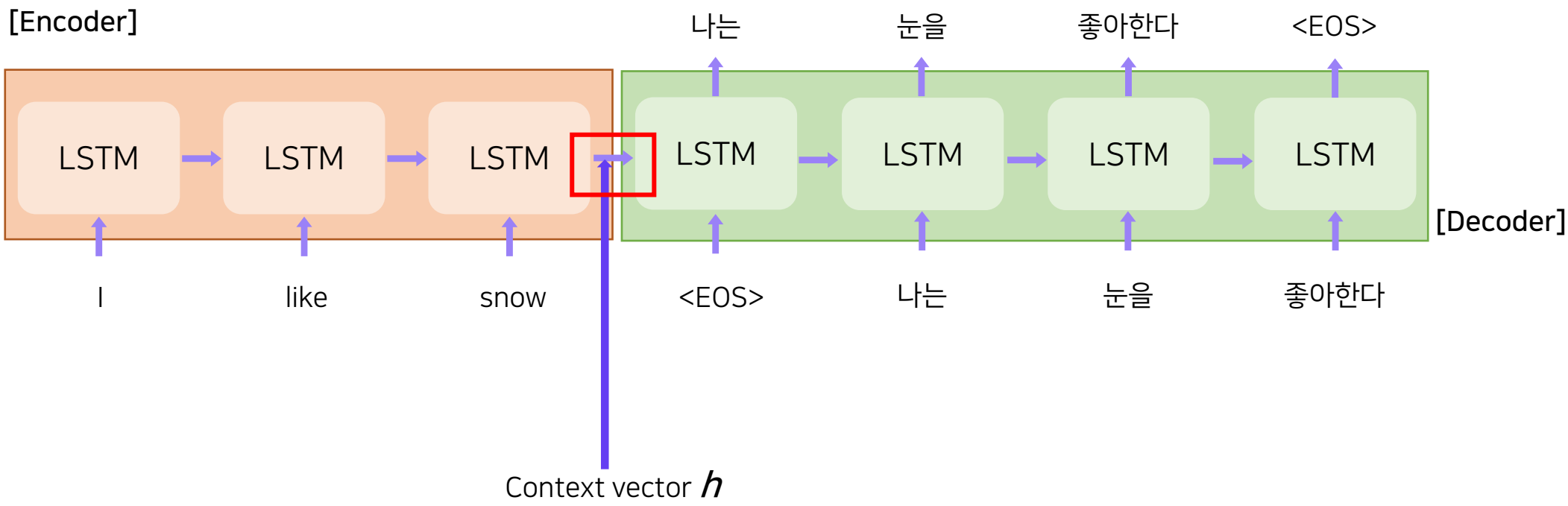





- Encoder에서 출력되는 output은 다음 state에 반영됨

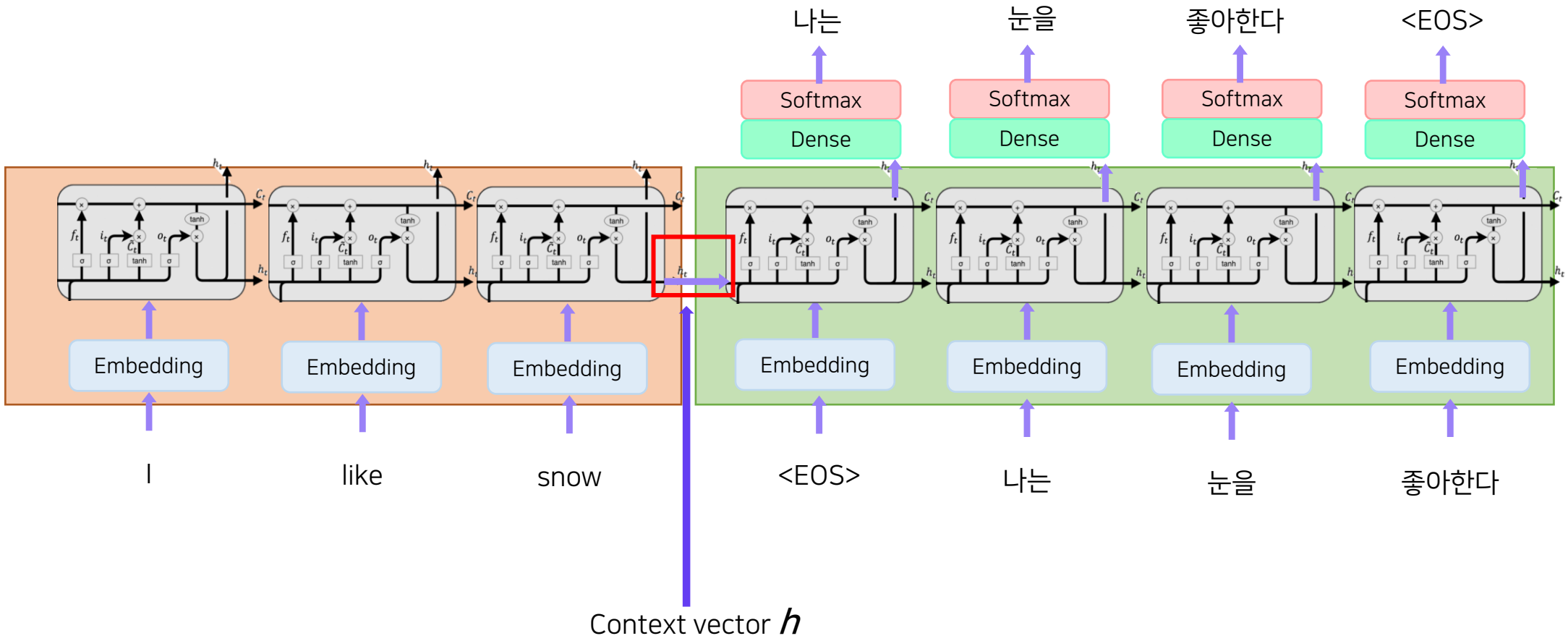
# Seq2Seq model





# Seq2Seq model

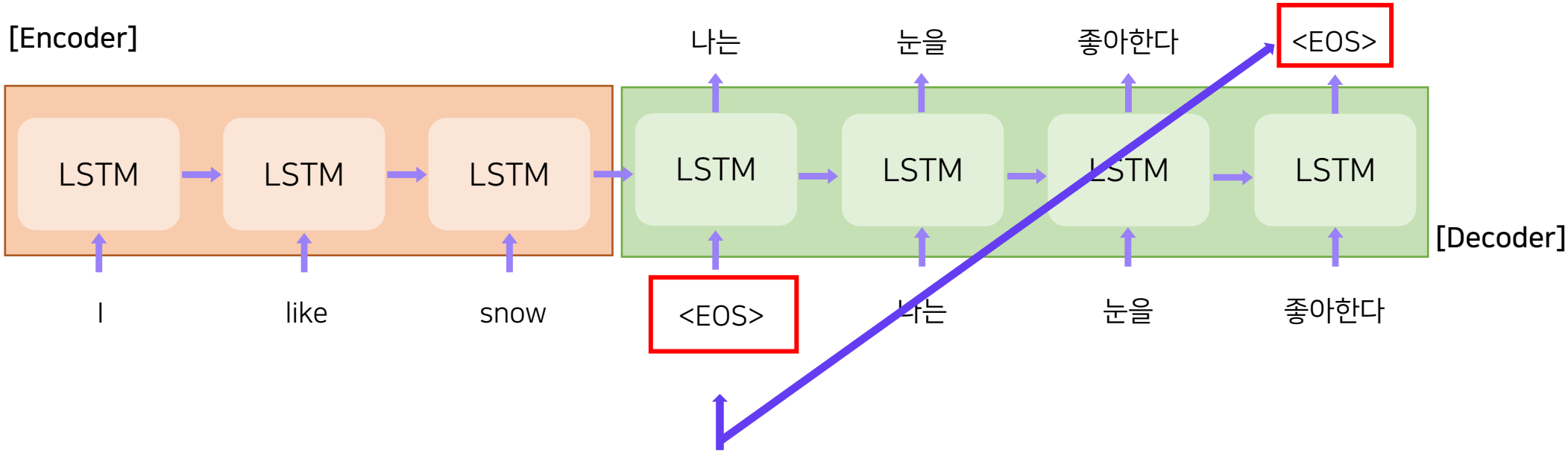





- Encoder의 마지막 timestep의 hidden state vector로 Input 문장에 대한 모든 정보를 담고 있다.
- 이 벡터가 Decoder의 첫번째 LSTM으로 들어가게 된다.

# Seq2Seq model



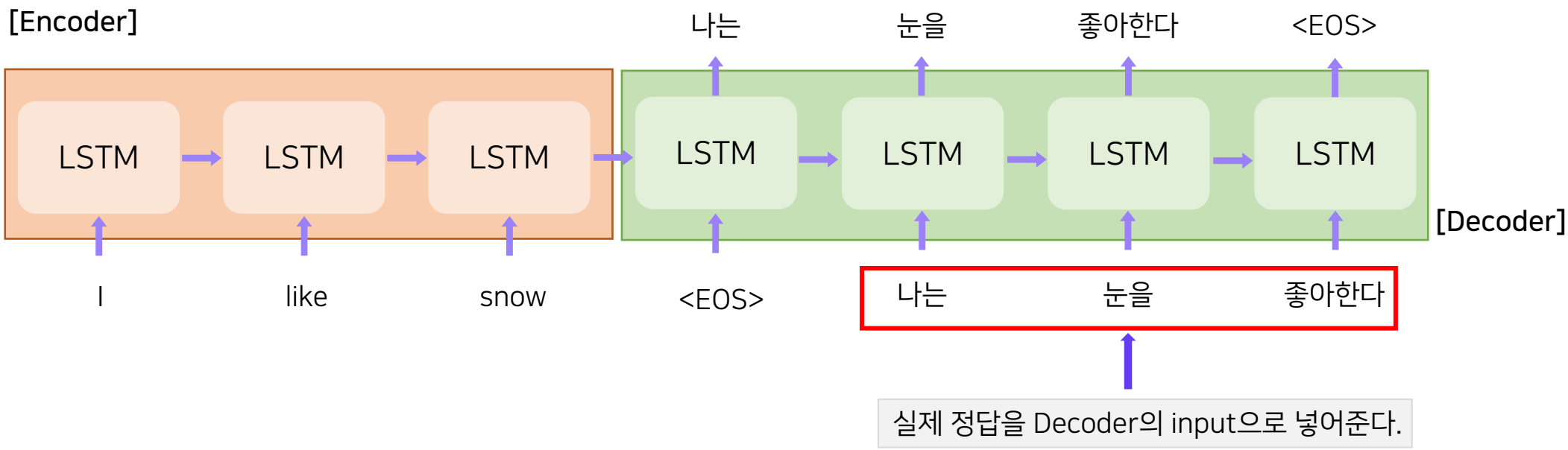



- 논문에서는 <EOS>라는 구분 기호를 이용하였음
- 입력 문장 맨 뒤에 들어옴
- Decoder에서 문장 생성 시작 / 종료를 알리는 구분자

# Train and Test



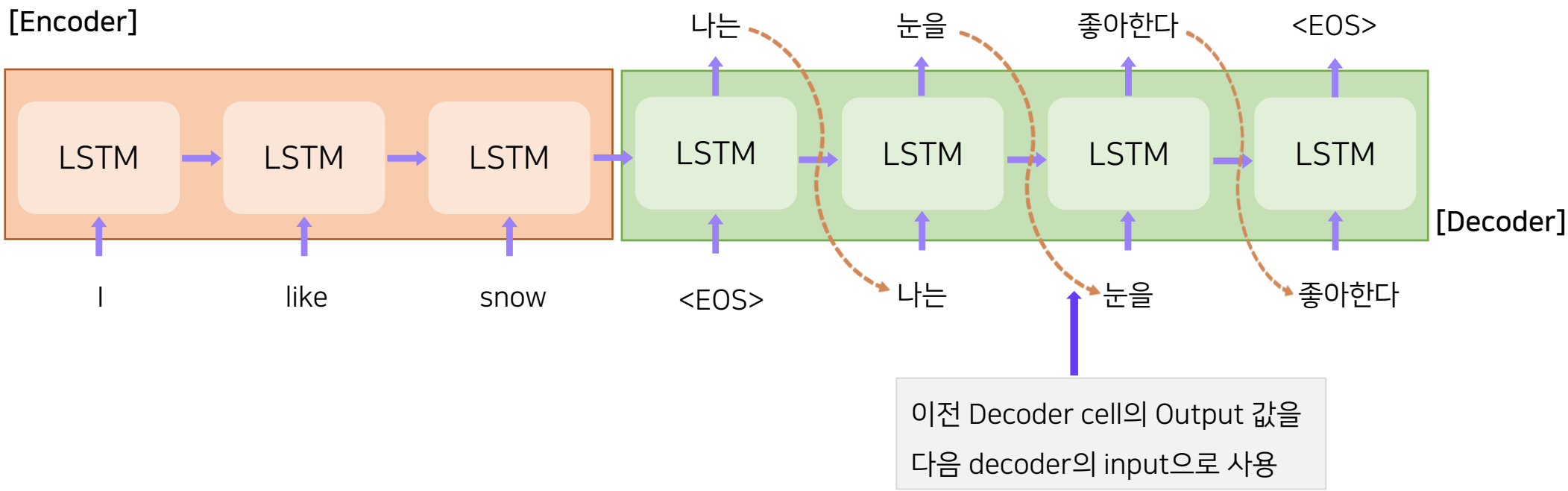
- **Train** – teacher forcing : 예측값 대신 실제 정답을 input으로 넣어주는 방법
  - 입력 문장: I like snow
  - GT: 나는 눈을 좋아한다



# Train and Test



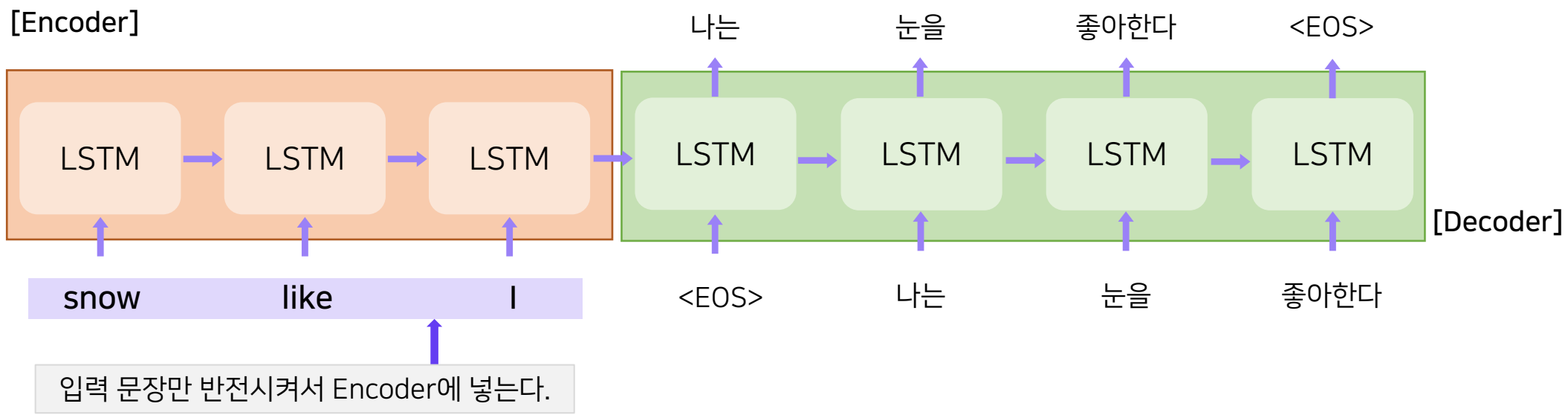
- Test
  - 입력 문장: I like snow



# 2. Improving Sequence to Sequence

# Reverse






Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 (2014).

# Reverse



| Method | test BLEU score (ntst14) |
|---|---|
| Bahdanau et al. [2] | 28.45 |
| Baseline System [29] | 33.30 |
| Single forward LSTM, beam size 12 | 26.17 |
| Single reversed LSTM, beam size 12 | 30.59 |
| Ensemble of 5 reversed LSTMs, beam size 1 | 33.00 |
| Ensemble of 2 reversed LSTMs, beam size 12 | 33.27 |
| Ensemble of 5 reversed LSTMs, beam size 2 | 34.50 |
| Ensemble of 5 reversed LSTMs, beam size 12 | **34.81** |

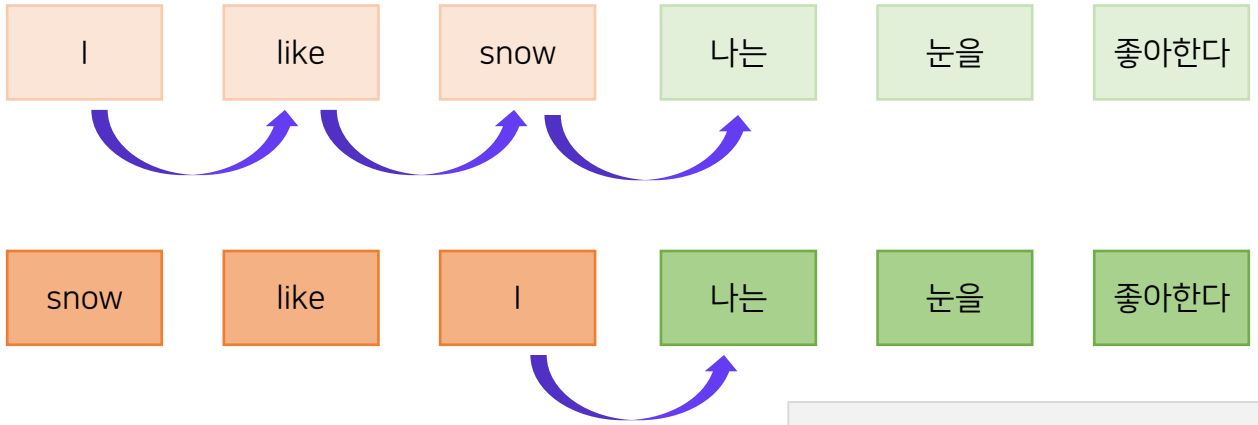
Table 1: The performance of the LSTM on WMT'14 English to French test set (ntst14). Note that an ensemble of 5 LSTMs with a beam of size 2 is cheaper than of a single LSTM with a beam of size 12.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 (2014).

# Reverse



- **왜 개선될까?**
  - 논문에서도 명확한 이유를 제시하진 않음
    - Dataset에 Short-term dependency가 존재할 것으로 예상



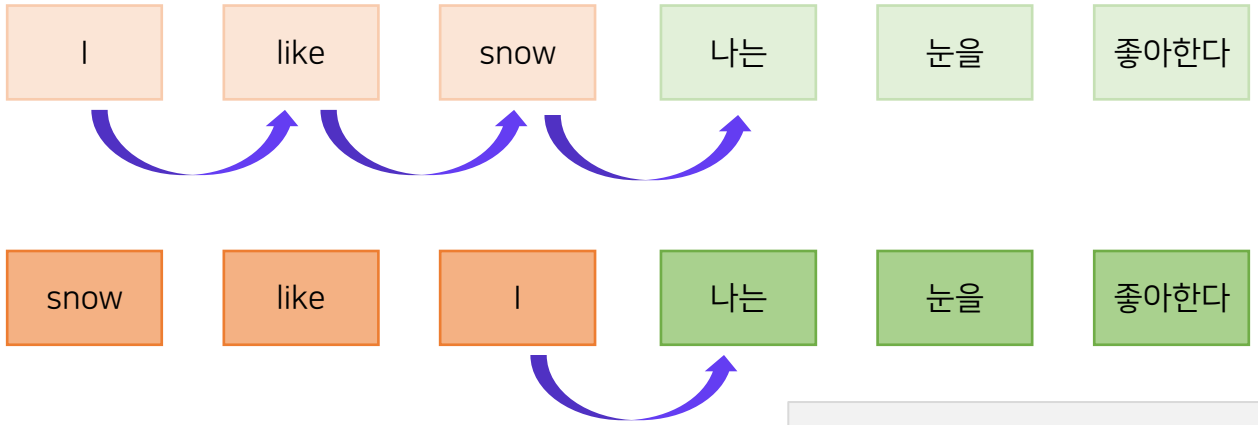

- 입력 문장의 첫 부분의 경우 반전 이후 변환 후의 단어와 더 가까워진다.
- 처음부터 잘 생성 해야지 **다음 단어를 제대로 생성할 가능성이 높아진다**.
- 입력 데이터를 반전시켜도 단어 사이의 **평균 거리는 그대로**이다.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 (2014).

# Reverse



- **왜 개선될까?**
  - 논문에서도 명확한 이유를 제시하진 않음
    - Dataset에 Short-term dependency가 존재할 것으로 예상

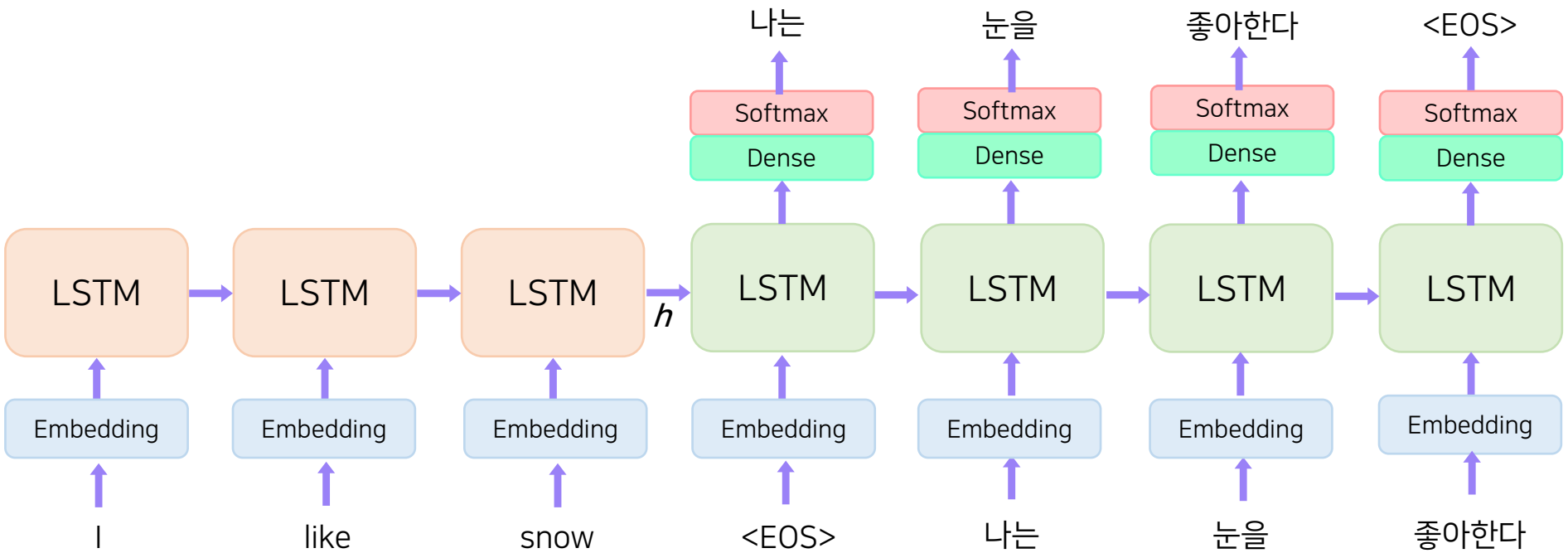



- 입력 문장의 첫 부분의 경우 반전 이후 변환 후의 단어와 더 가까워진다.
- 처음부터 잘 생성 해야지 **다음 단어를 제대로 생성할 가능성이 높아진다**.
- 입력 데이터를 반전시켜도 단어 사이의 **평균 거리는 그대로**이다.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 (2014).

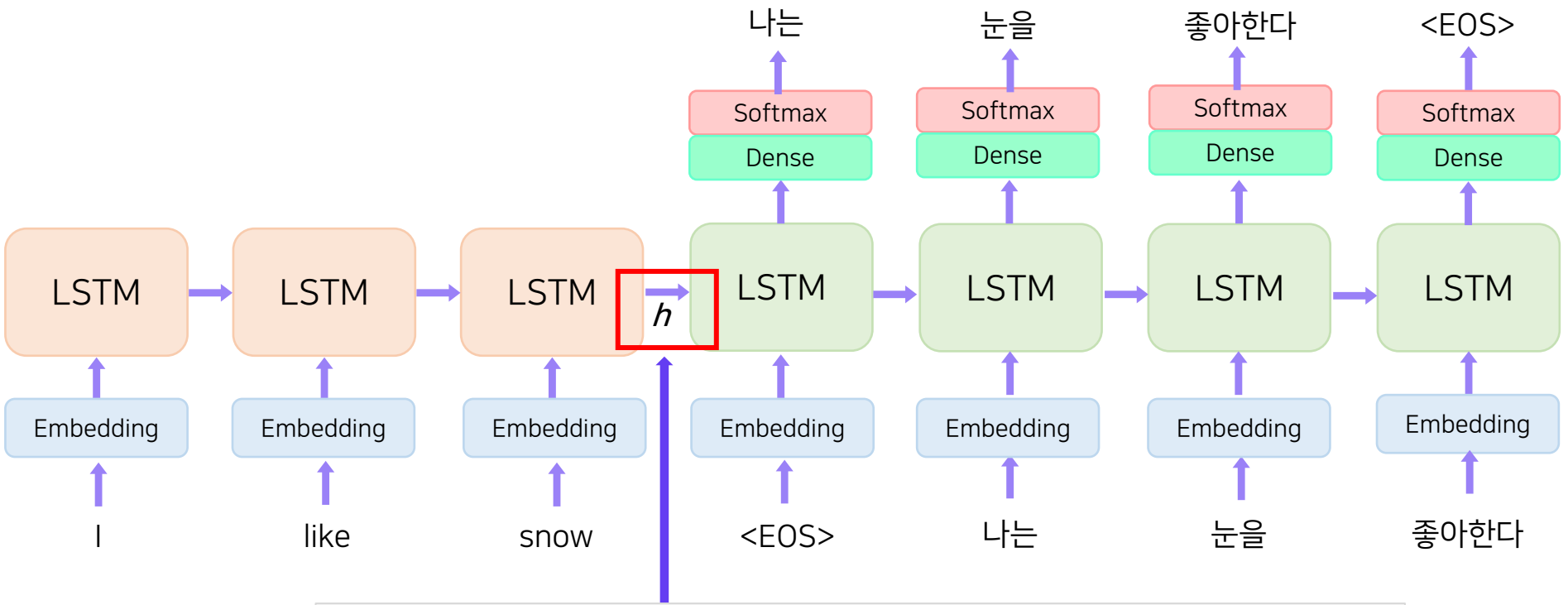
# Peeky




나는
눈을
좋아한다
<EOS>
Softmax
Dense
LSTM
Embedding
I
like
snow
h
<EOS>
나는
눈을
좋아한다

# Peeky

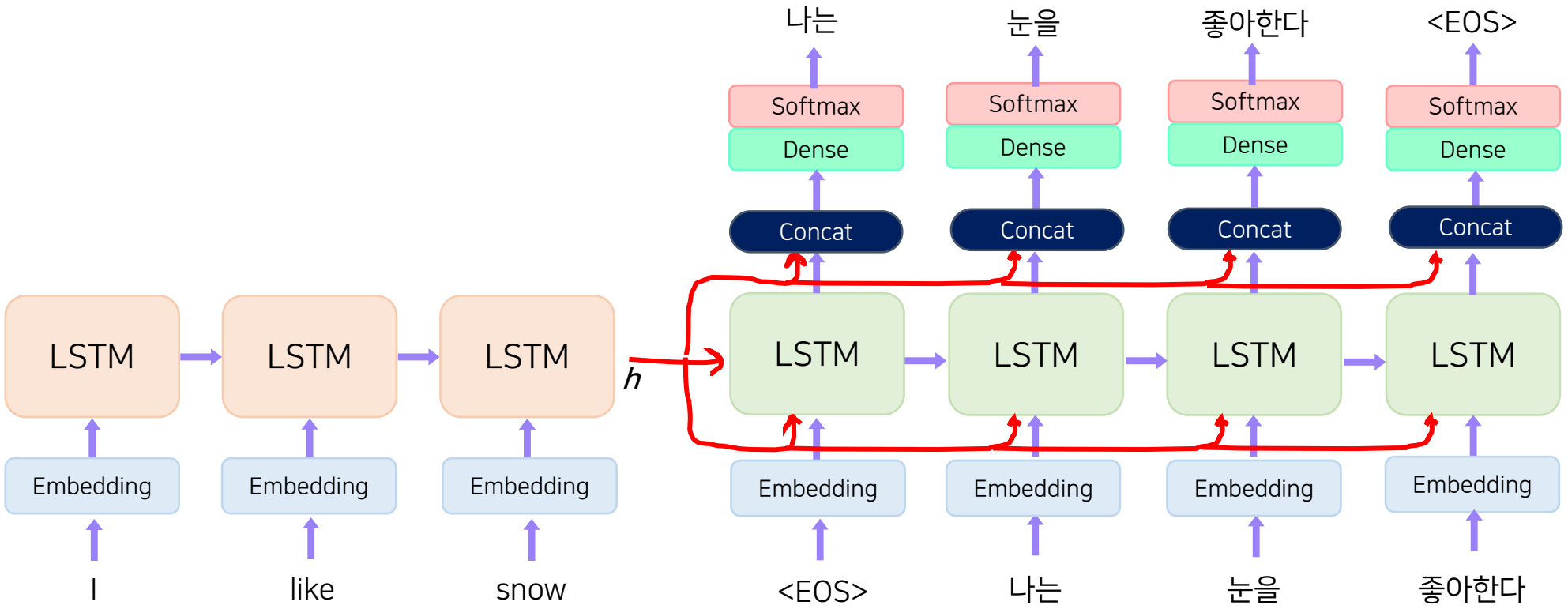





- Input 문장에 대한 정보가 Decoder에 넘어갈 때, Decoder의 첫번째 cell만 ***h*** 를 이용하고 있음
- 이 정보를 더 잘 활용할 수 없을까?

# Peeky






Cho, Kyunghyun, et al. "Learning phrase representations usingRNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 (2014).

# Reference


- Sequence to Sequence Learning with Neural Networks (NIPS 2014)
  https://proceedings.neurips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html
- Learning phrase representations using RNN encoder-decoder for statistical machine translation (EMNLP 2014)
  https://arxiv.org/abs/1406.1078
- https://wikidocs.net/24996