

Seq2Seq with Attention

임세진

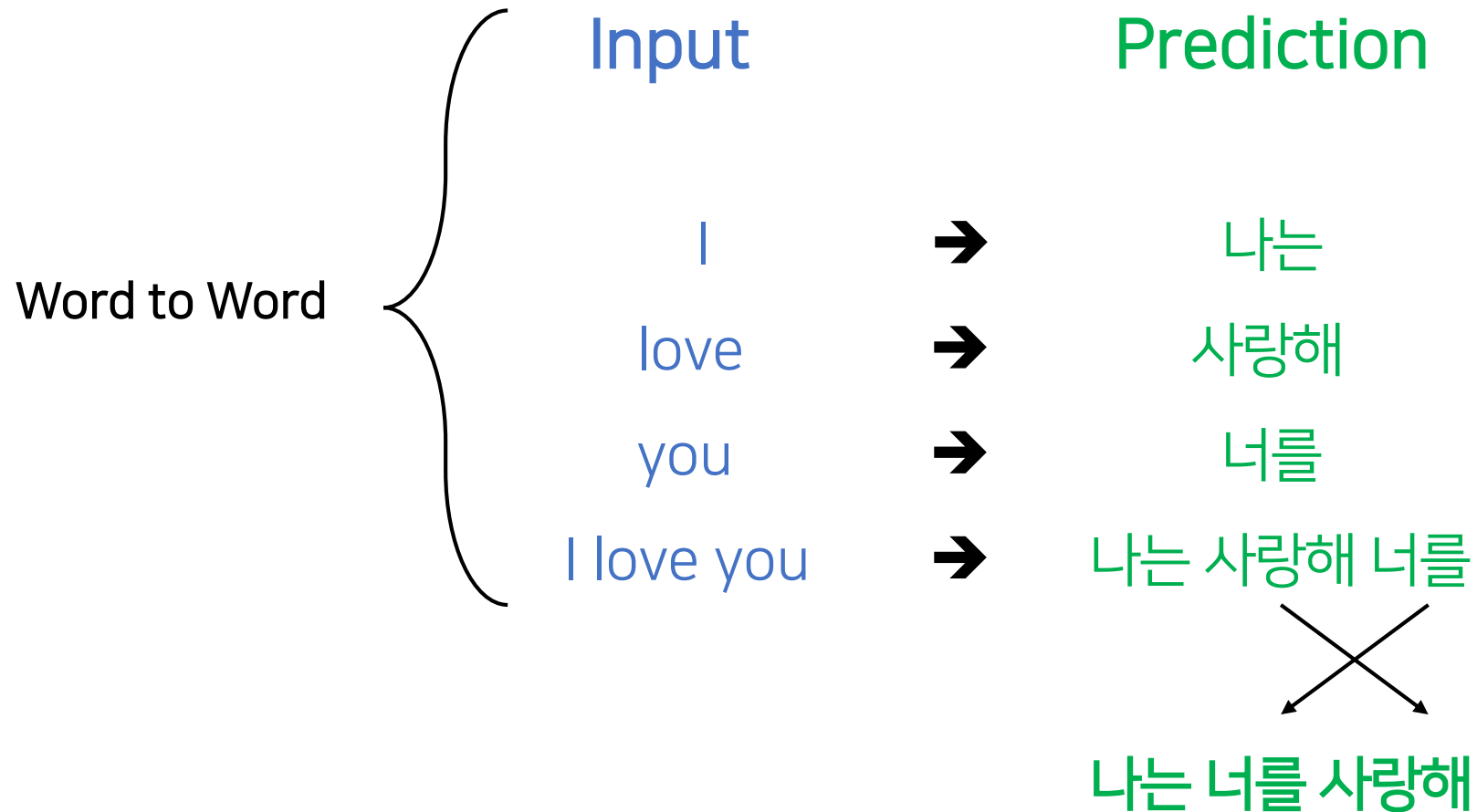
<https://youtu.be/iT6rT3kLOm4>

01. Seq2Seq

02. Seq2Seq with Attention

01. Seq2Seq

기계번역의 문제점 1 : 언어별 단어 순서가 다른 경우



01. Seq2Seq

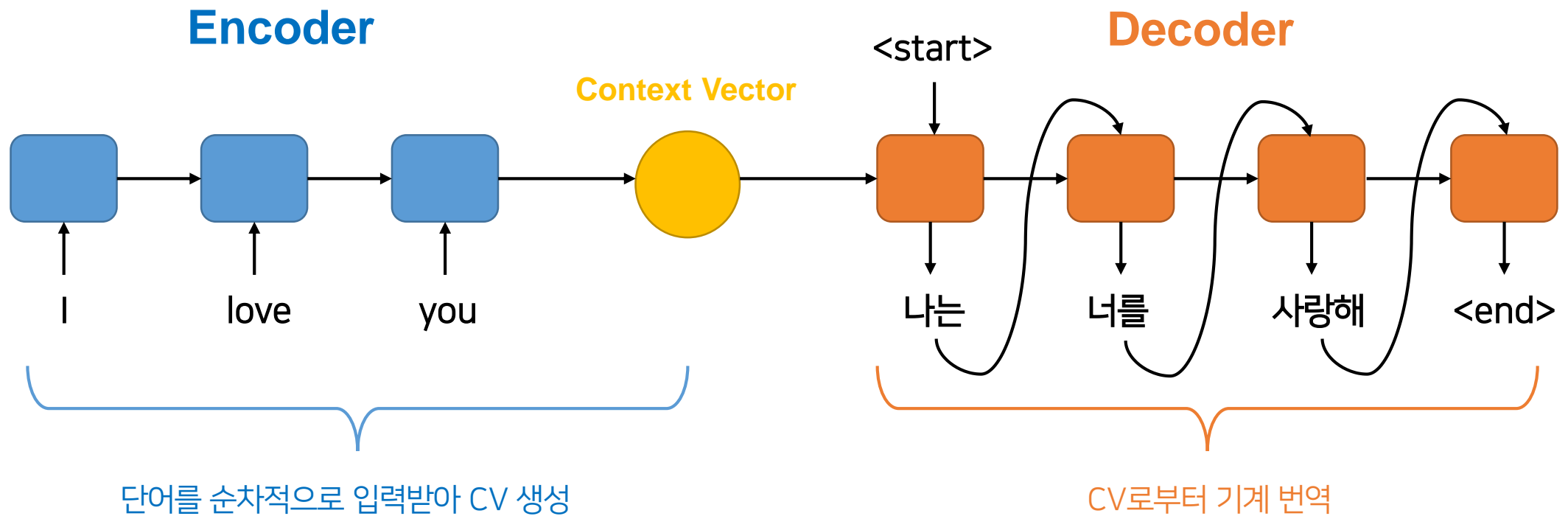
기계번역의 문제점 2 : 단어별 해석 △

How/are/you ? = 잘/지내 ?
3 words 2 words

01. Seq2Seq

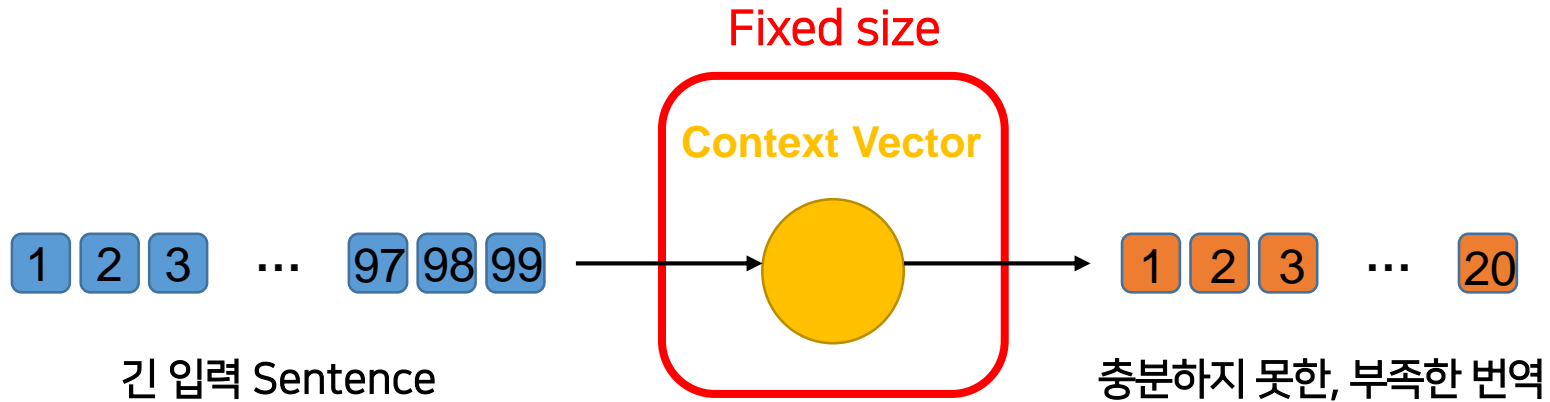
앞의 두가지 문제점 해결 : RNN 기반 Encoder Decoder (Seq2Seq)

Sequence-to-Sequence



01. Seq2Seq

문제점 : 고정된 크기의 Context Vector

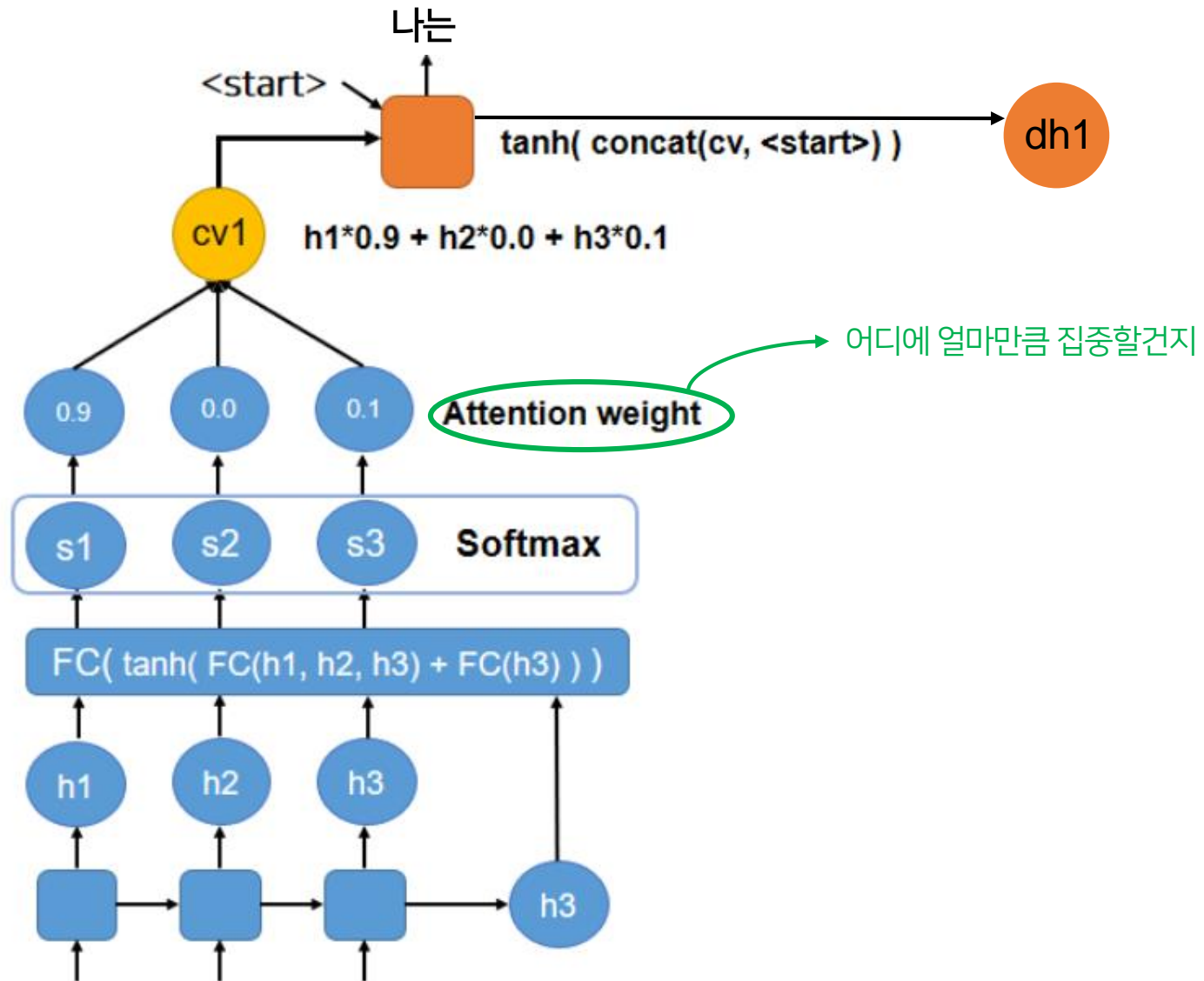


하나의 문맥 벡터가 소스 문장의 모든 정보를 가지고 있어야함 → 성능저하

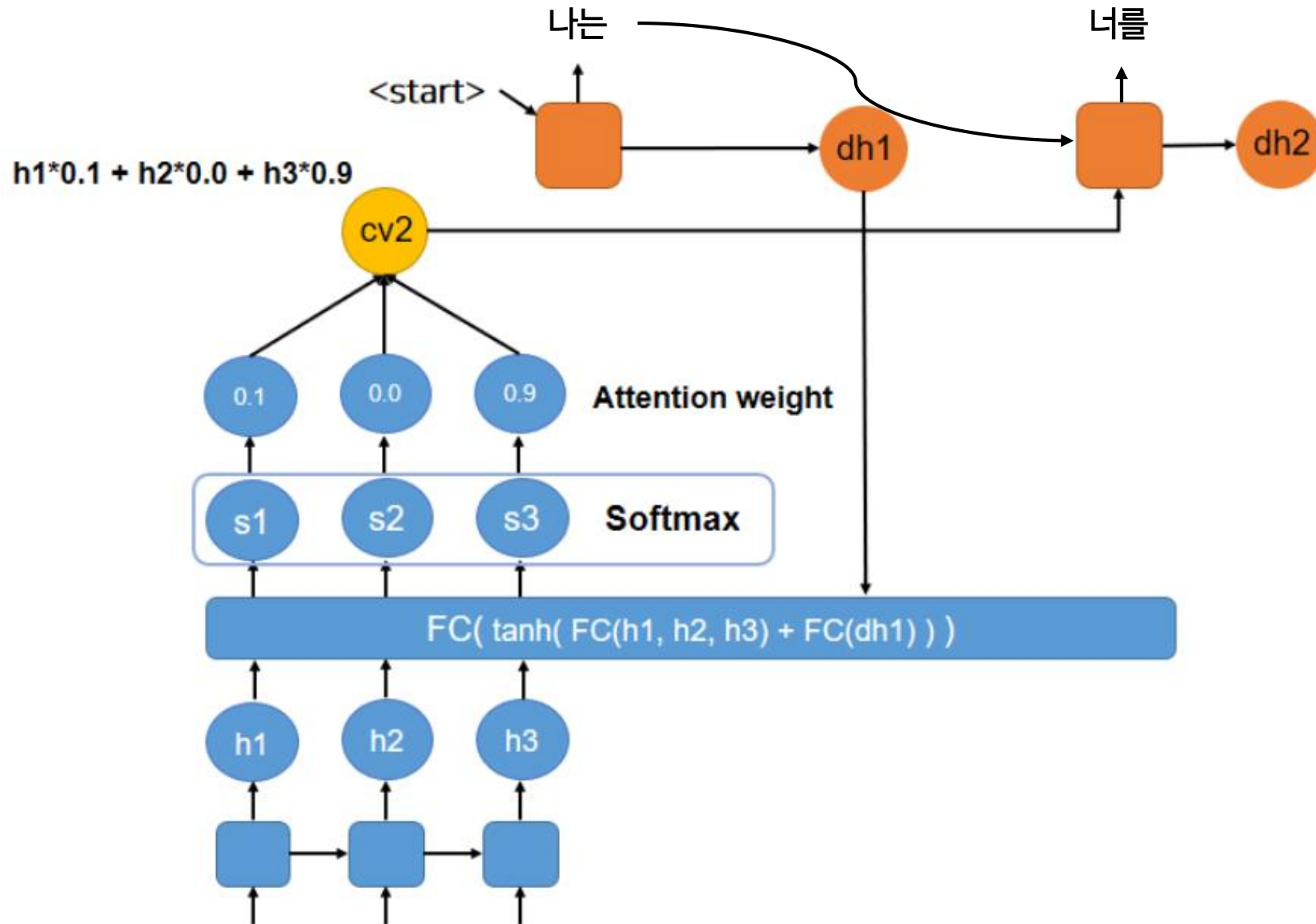


입력 문장 전체에서 정보 추출 (기술 발전으로 GPU는 많은 메모리와 빠른 병렬 처리 지원 가능), 고정된 사이즈 문제 해결

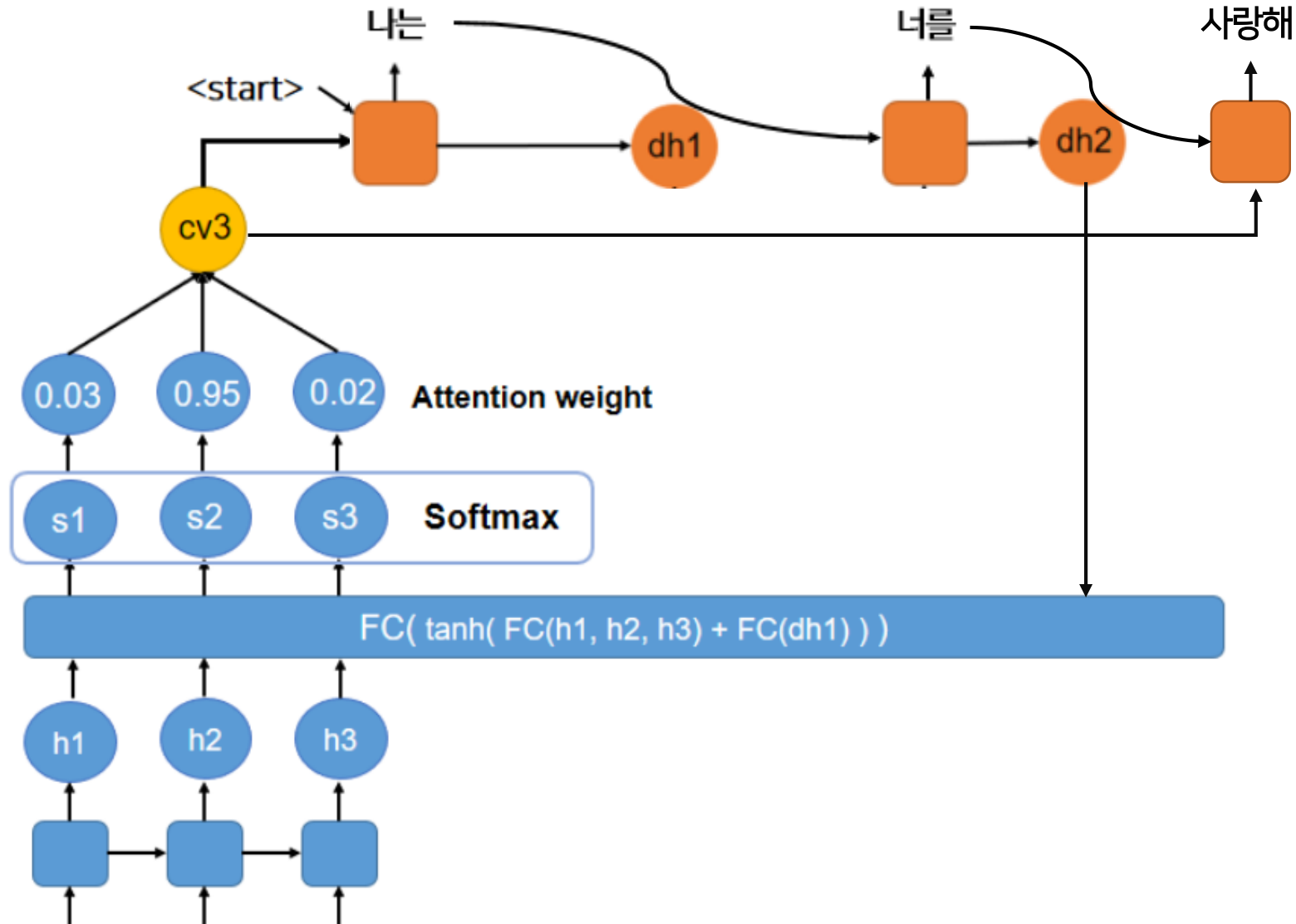
02. Seq2seq + Attention



02. Seq2seq + Attention



02. Seq2seq + Attention



02. Seq2seq + Attention

- 디코더는 매번 인코더의 모든 출력 중 어떤 정보가 중요한지 계산

- i = 현재 디코더가 처리 중인 인덱스
- j = 각각의 인코더 출력 인덱스

✓ 에너지 (Energy)

입력 문장 중 어떤 것과 가장 연관이 있는지 수치화

✓ 가중치 (Weight)

상대적 확률값

$$e_{ij} = a(s_{i-1}, h_j)$$

각 단어별 인코더 출력 값

디코더가 사용했던 hidden state

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

softmax

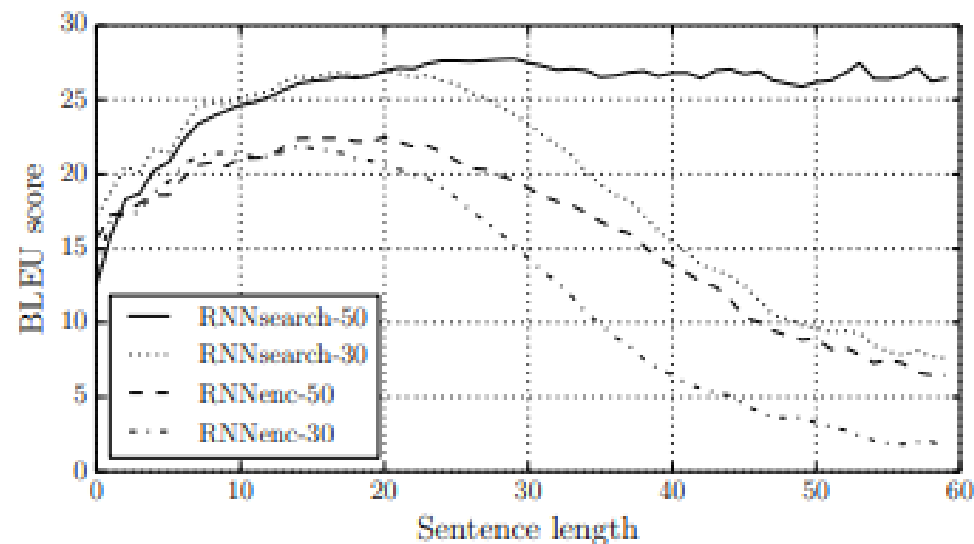
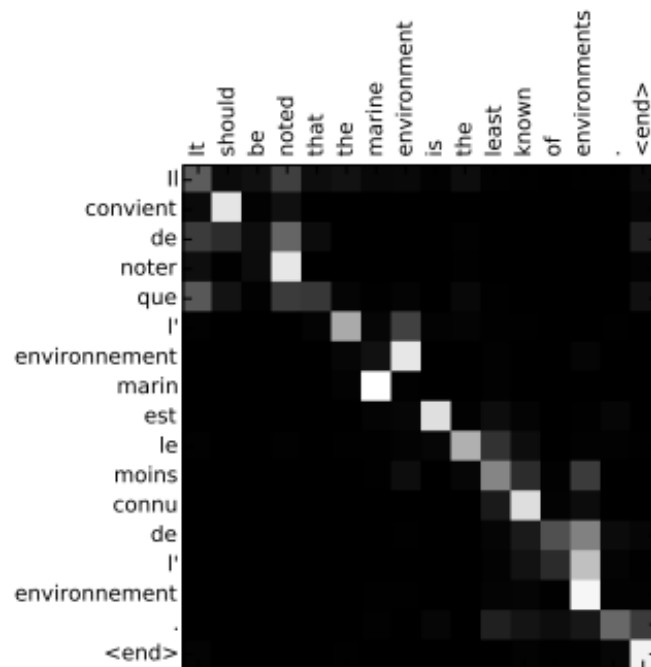
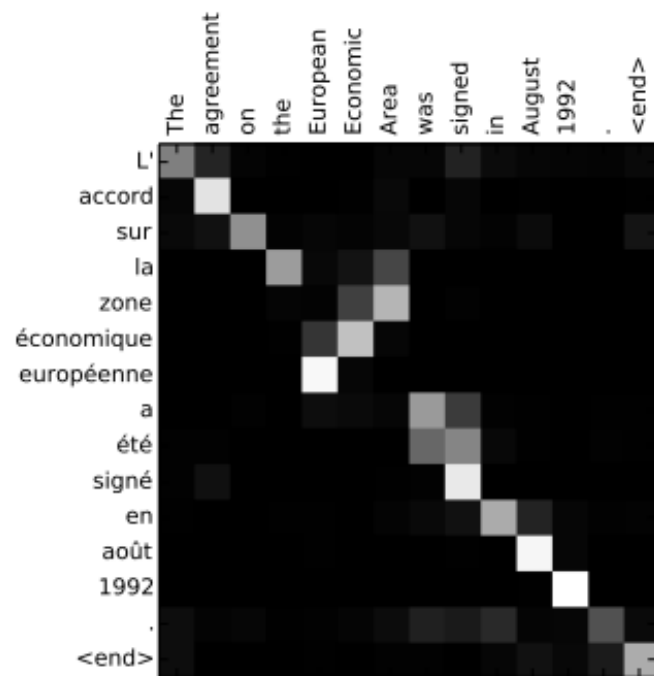


Weighted sum

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

디코더의 입력 (context vector)

02. Seq2seq + Attention



Attention 가중치를 사용하여 각 출력이 어떤 입력 정보를 참고했는지 시각화 할 수 있음

Q & A