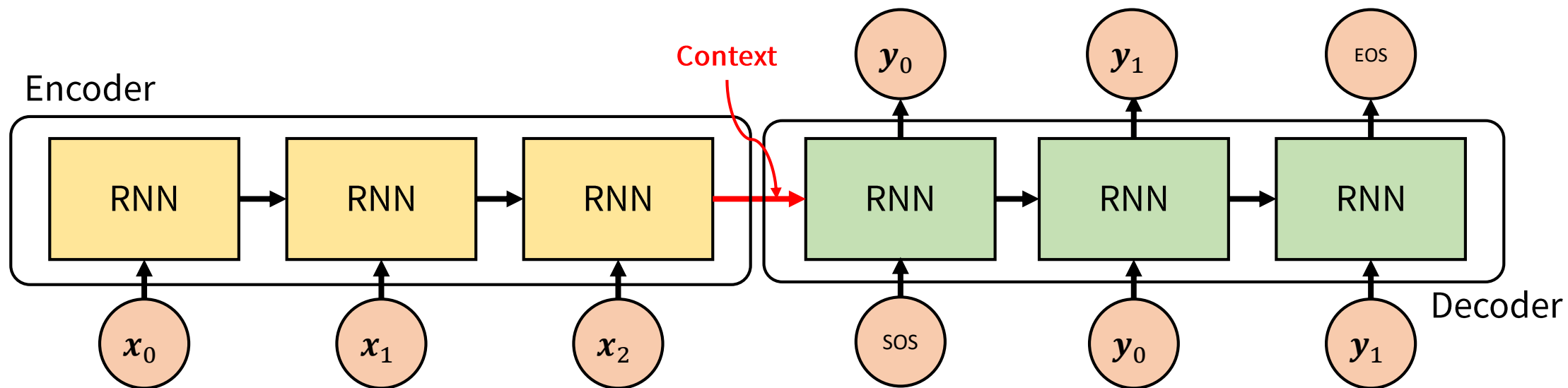


Chapter 07. 맥락을 파악하는 Attention 기법

STEP1. Attention 기법의 이해

Sequence-to-sequence model



Seq2seq 모델은 번역 문제를 학습하기 위해 널리 사용되고 있는 RNN 구조이다.

영어 문장의 데이터화

단어 분리, 문장부호 제거

What a beautiful place!



['what', 'a', 'beautiful', 'place']

Tokenizer



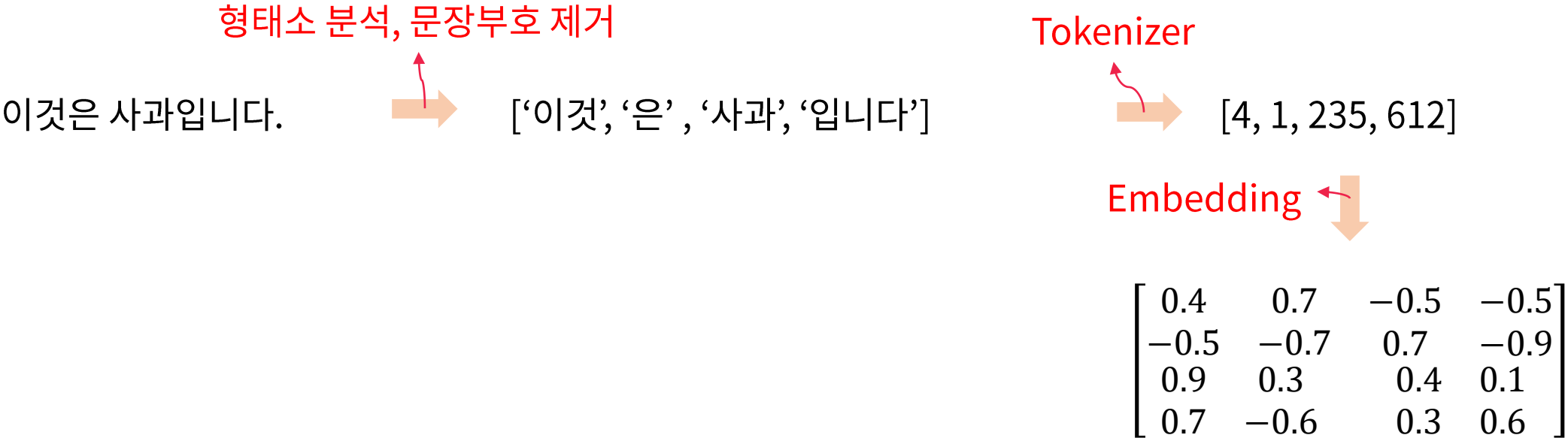
[4, 1, 235, 612]

Embedding


$$\begin{bmatrix} 0.4 & 0.7 & -0.5 & -0.5 \\ -0.5 & -0.7 & 0.7 & -0.9 \\ 0.9 & 0.3 & 0.4 & 0.1 \\ 0.7 & -0.6 & 0.3 & 0.6 \end{bmatrix}$$

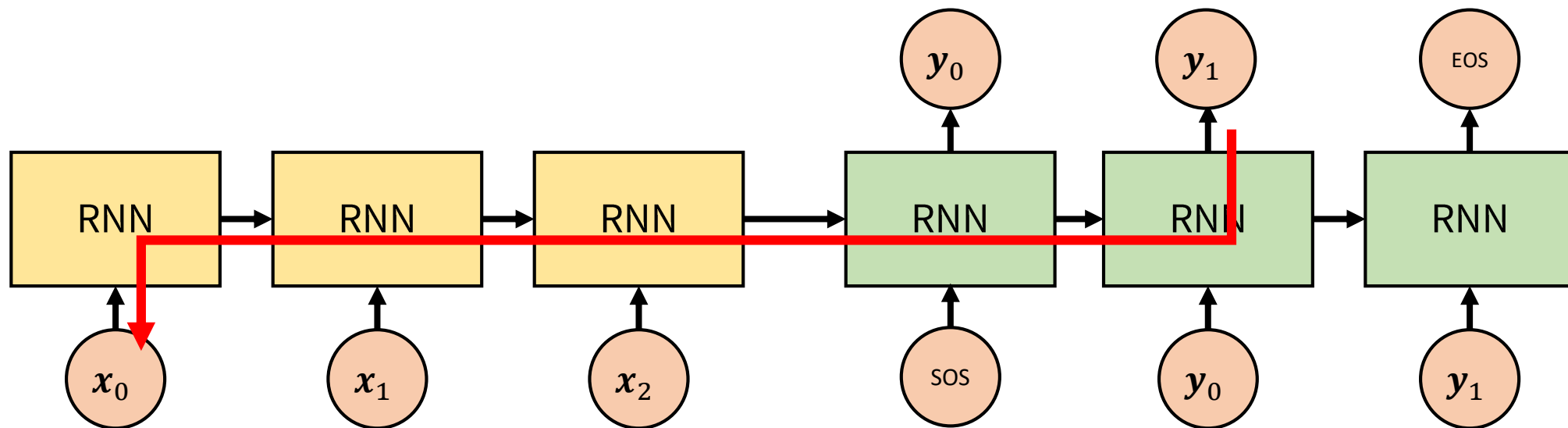
영어 문장을 데이터화 하기 위해서는, Tokenizer와 Embedding이 필요하다.

한글 문장의 데이터화



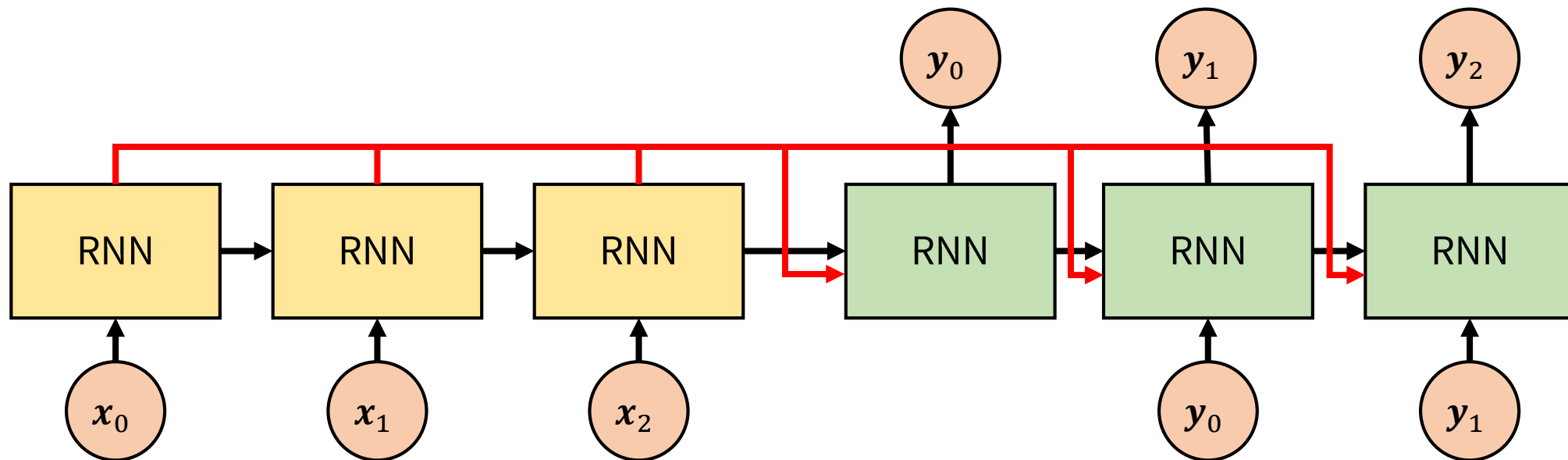
한글 문장의 경우, 띄어쓰기만으로 단어를 구분하지 않고 형태소 분석이 필요하다.

Gradient Vanishing in RNN



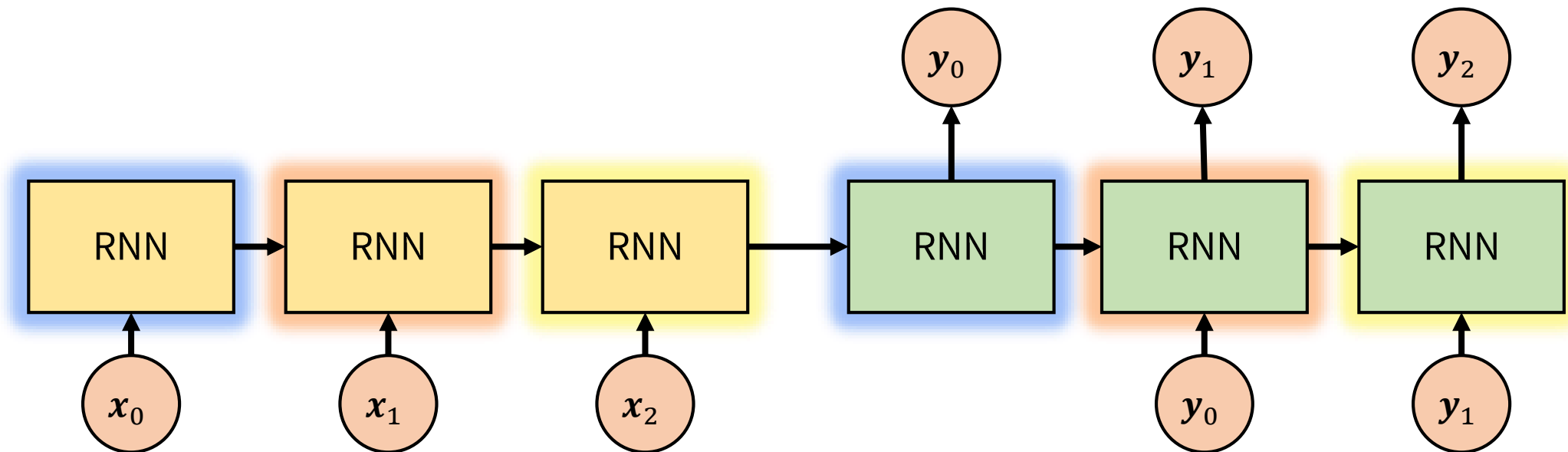
입출력 연관 관계가 너무 멀리 떨어져 있으면 기울기 소실이 일어나 잘 학습되지 않는다.

No more gradient vanishing!



이렇게 모든 Encoder hidden state를 모아서 Decoder로 각각 전달하면 기울기 소실을 해결할 수 있을 것이다. 이 구조를 어떻게 하면 효율적으로 구성할 수 있을까?

Attention!



Decoder 단에서 어떤 Encoder 정보에 ‘집중’해야 하는지 알 수 있다면, 도움이 될 것이다.
이것이 Attention mechanism의 기본 아이디어!