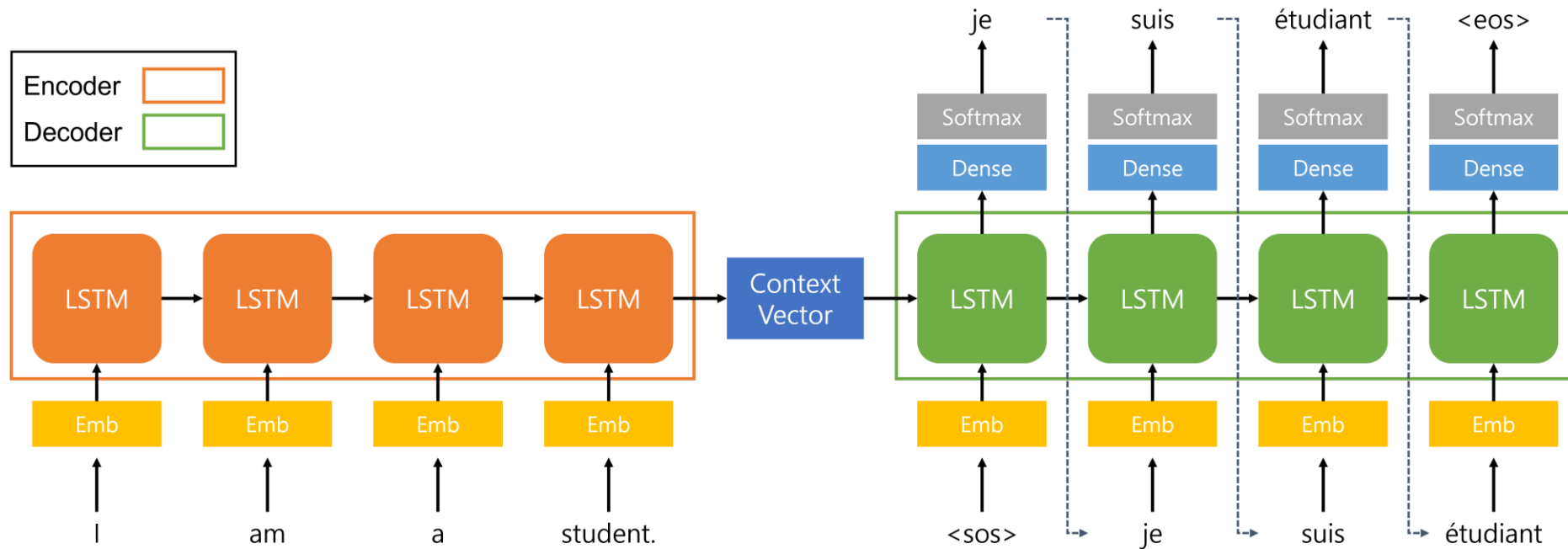


Seq2Seq with Attention

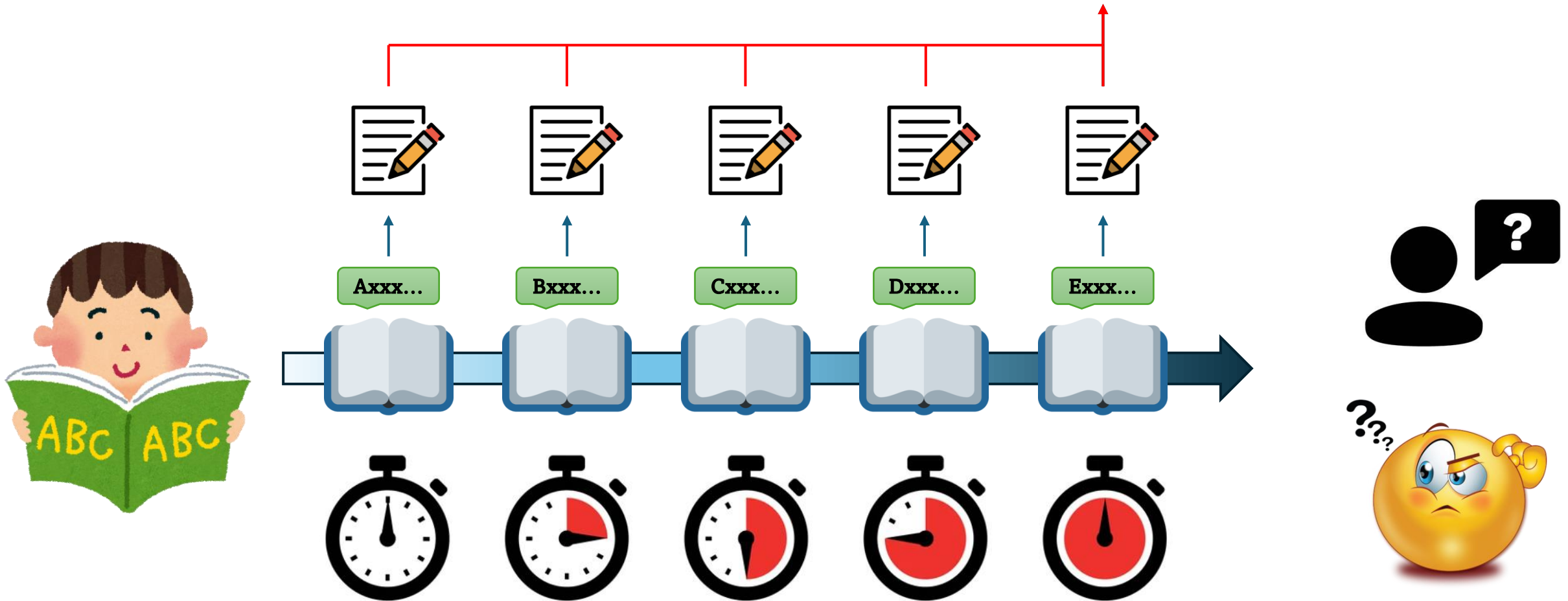


Architecture of seq2seq

Problem : Bottle Neck?

Problem Of seq2seq

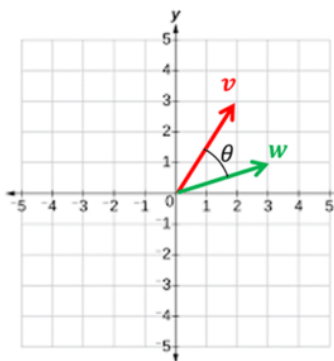
seq2seq **with attention**



Dot Product & Cosine Similarity

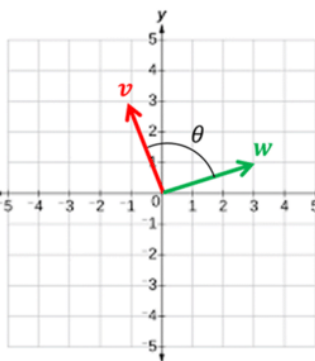
$$\text{sim}(\mathbf{v}, \mathbf{w}) = \cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}$$

$$\text{sim}(\mathbf{v}, \mathbf{w}) = \cos(\theta) \approx 0.79$$



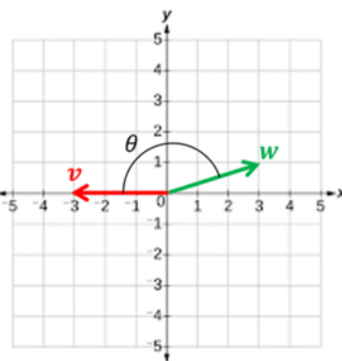
Vectors \mathbf{v} and \mathbf{w} are pointing in similar directions. The angle θ between \mathbf{v} and \mathbf{w} is near 0° . The cosine similarity between vectors \mathbf{v} and \mathbf{w} is near 1.

$$\text{sim}(\mathbf{v}, \mathbf{w}) = \cos(\theta) = 0$$

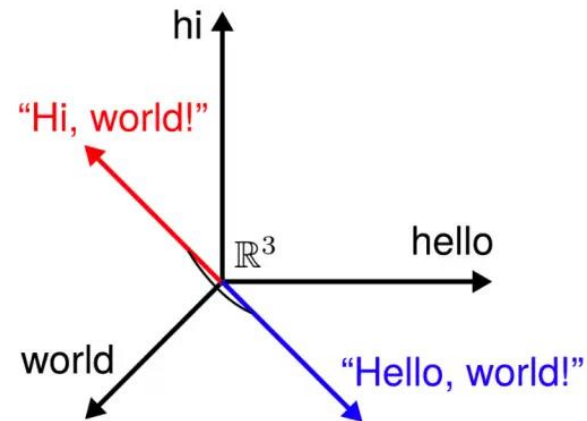


Vectors \mathbf{v} and \mathbf{w} are pointing in perpendicular directions. The angle θ between \mathbf{v} and \mathbf{w} is 90° . The cosine similarity between vectors \mathbf{v} and \mathbf{w} is 0.

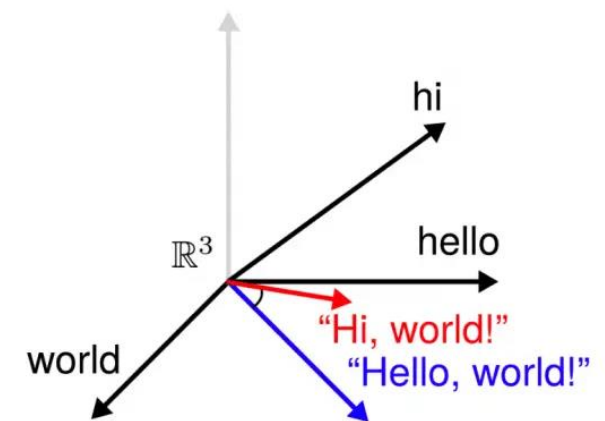
$$\text{sim}(\mathbf{v}, \mathbf{w}) = \cos(\theta) \approx -0.95$$



Vectors \mathbf{v} and \mathbf{w} are pointing in near opposite directions. The angle θ between \mathbf{v} and \mathbf{w} is near 180° . The cosine similarity between vectors \mathbf{v} and \mathbf{w} is near -1.



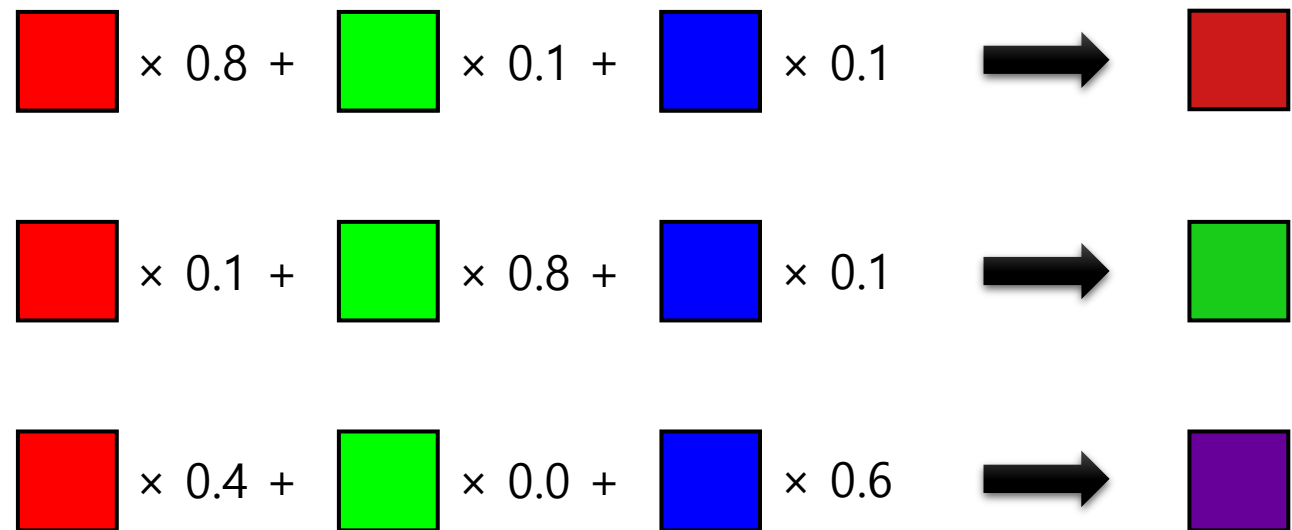
Cosine Similarity



Soft Cosine Measure

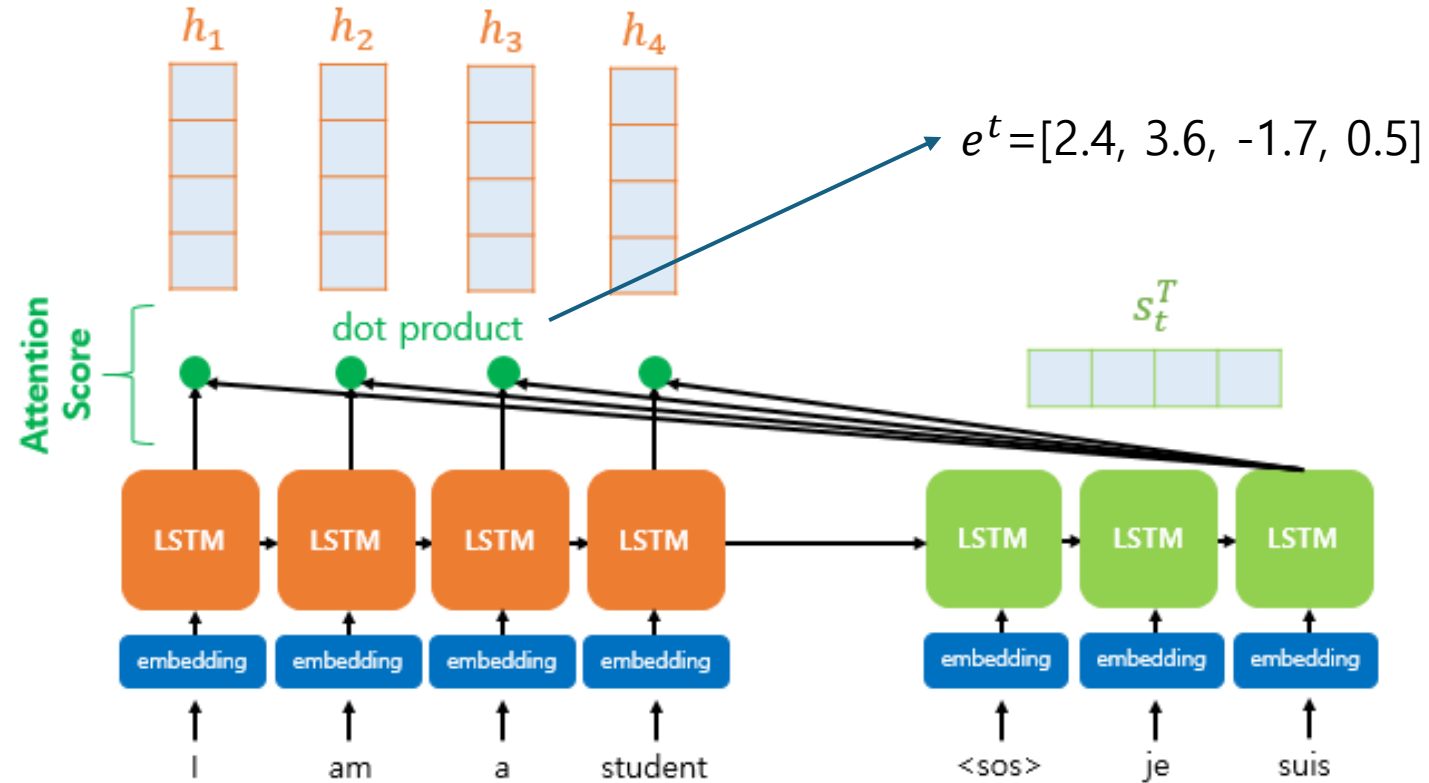
※ 두 벡터 간의 내적이 클 수록 θ 가 작고 유사도가 크다

Weighted Sum



※ weight는 항상 양수이며, 모든 weight의 합은 1이다

Step.1 – Evaluate Attention Score



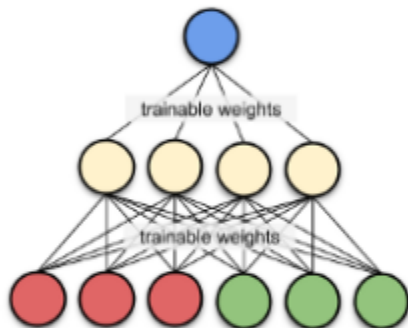
$$score(s_t, h_i) = s_t^T h_i$$

$$e^t = [s_t^T h_1, s_t^T h_2, \dots, s_t^T h_N]$$

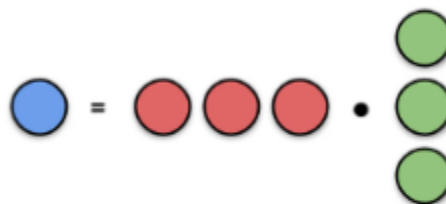
Step.1 – Evaluate Attention Score



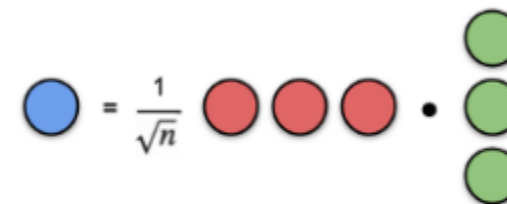
Additive / Concat



Dot product



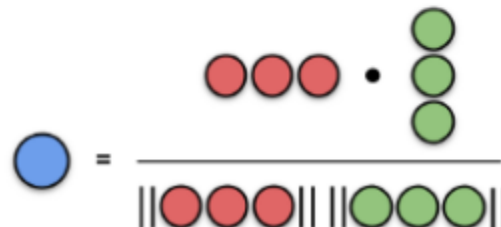
Scaled dot product



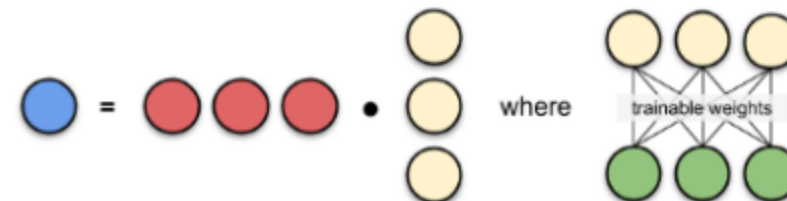
Location-based



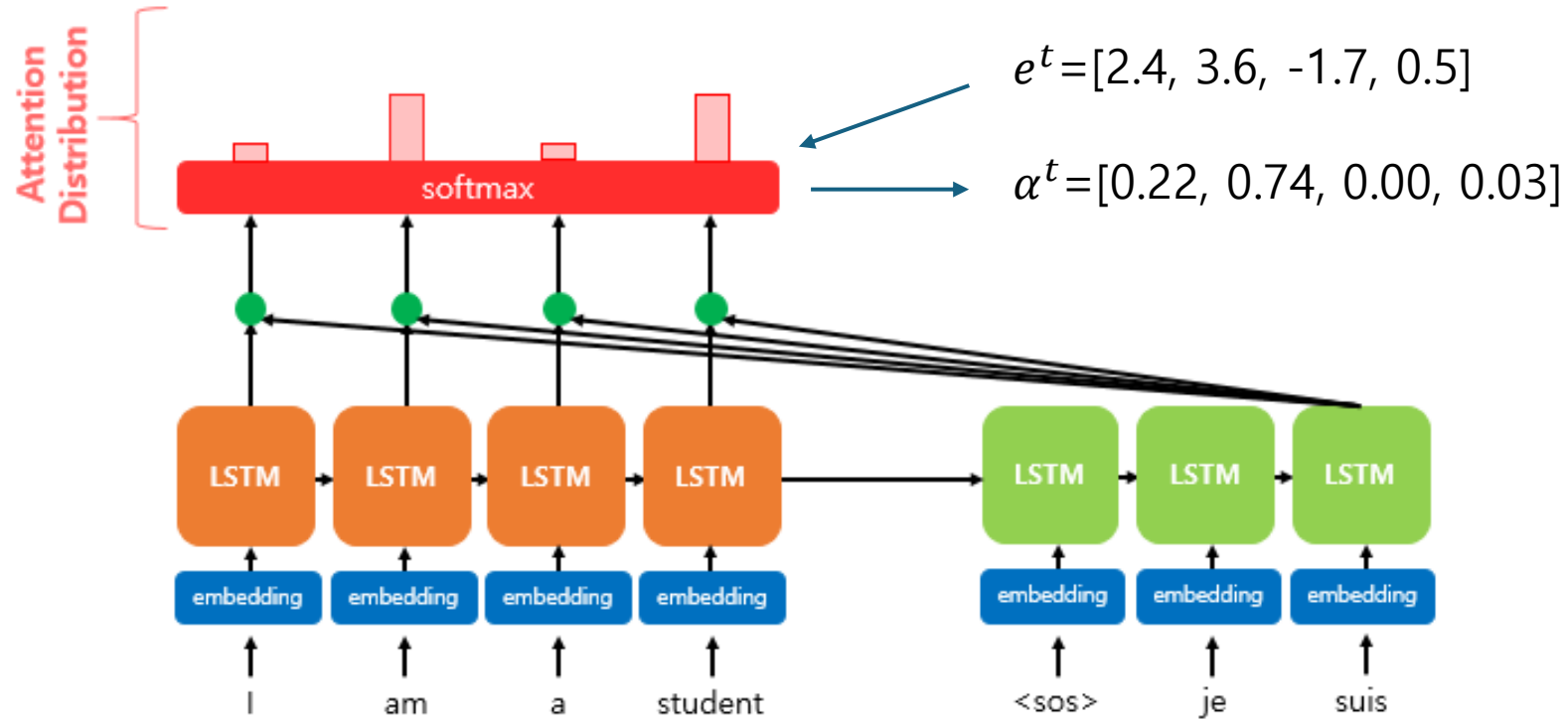
Cosine similarity



General



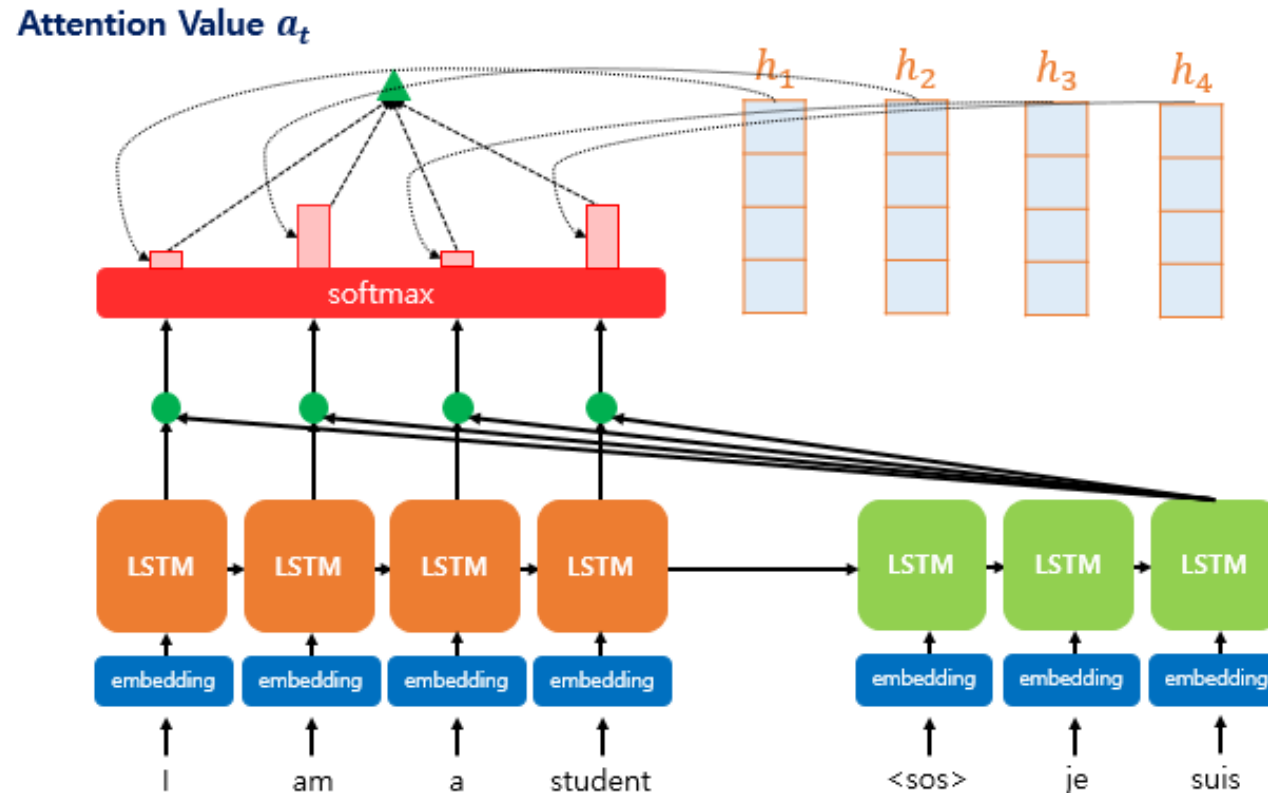
Step.2 – Evaluate Attention Distribution



$$\alpha^t = \text{softmax}(e^t)$$

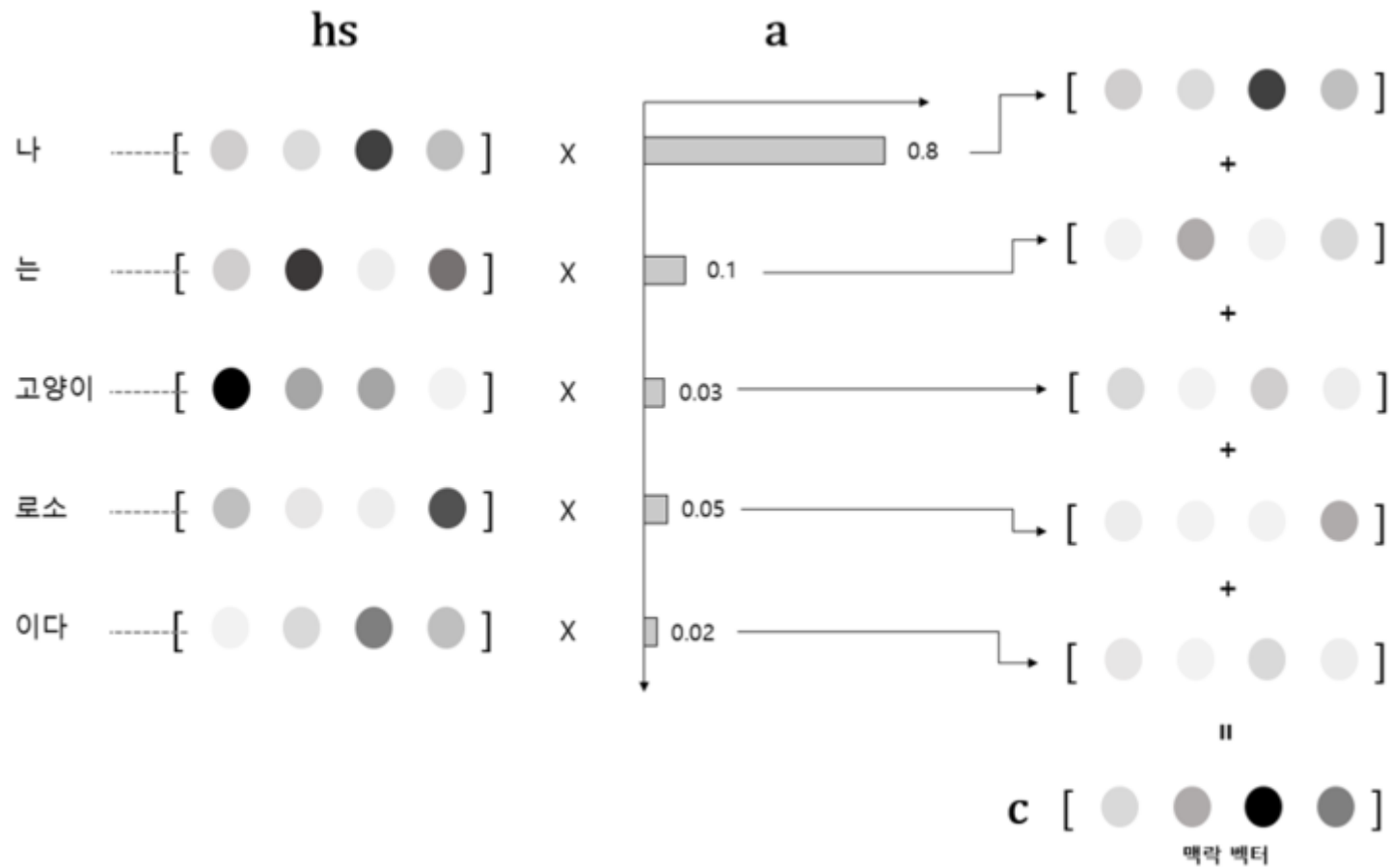
$$\text{softmax} : y_k = \frac{\exp(a_k)}{\sum_{i=1}^n \exp(a_i)}$$

Step.3 – Evaluate Attention Value

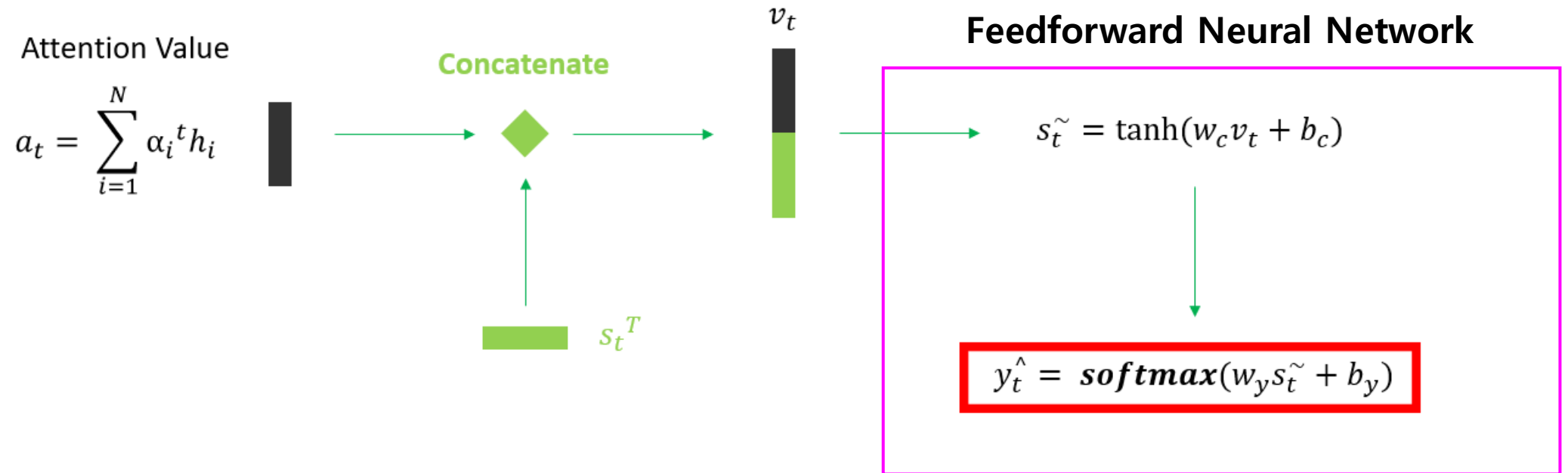


$$a_t = \sum_{i=1}^N \alpha_i^t h_i \text{ (Attention Value or Context Vector)}$$

Step.3 – Evaluate Attention Value



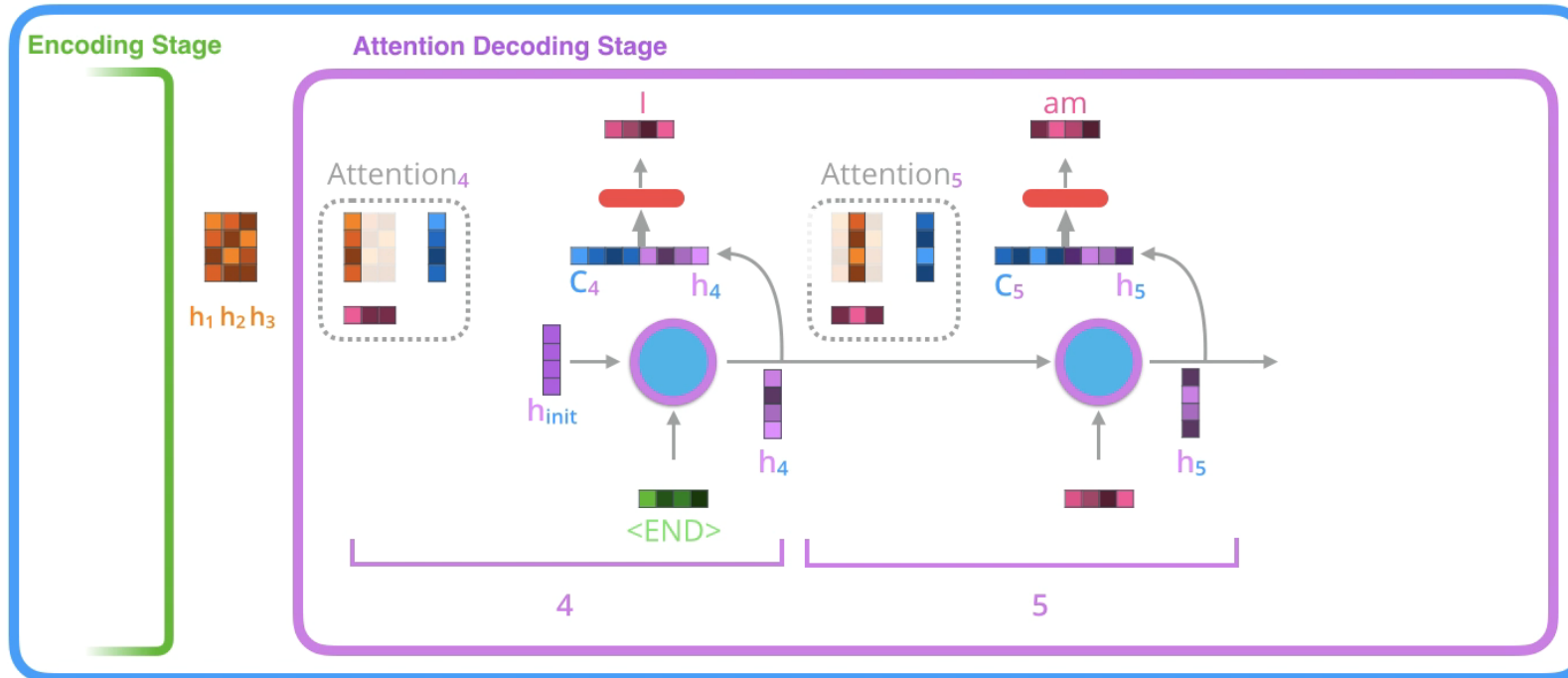
Step.4 – Evaluate New hidden State and Output



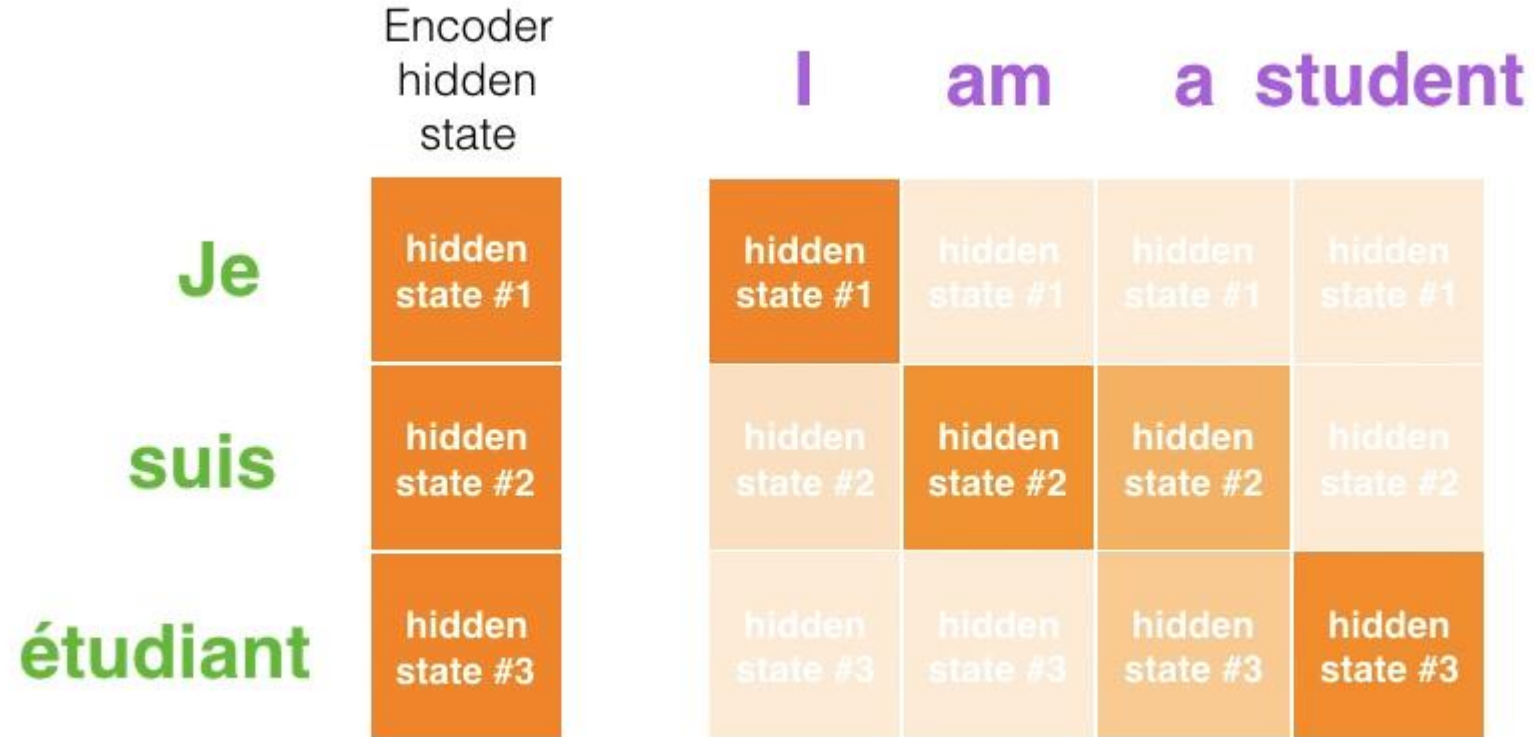
v_t : Concatenation of **Attention Value**(a_t) and **Hidden State of Decoder**(s_t)

Summary

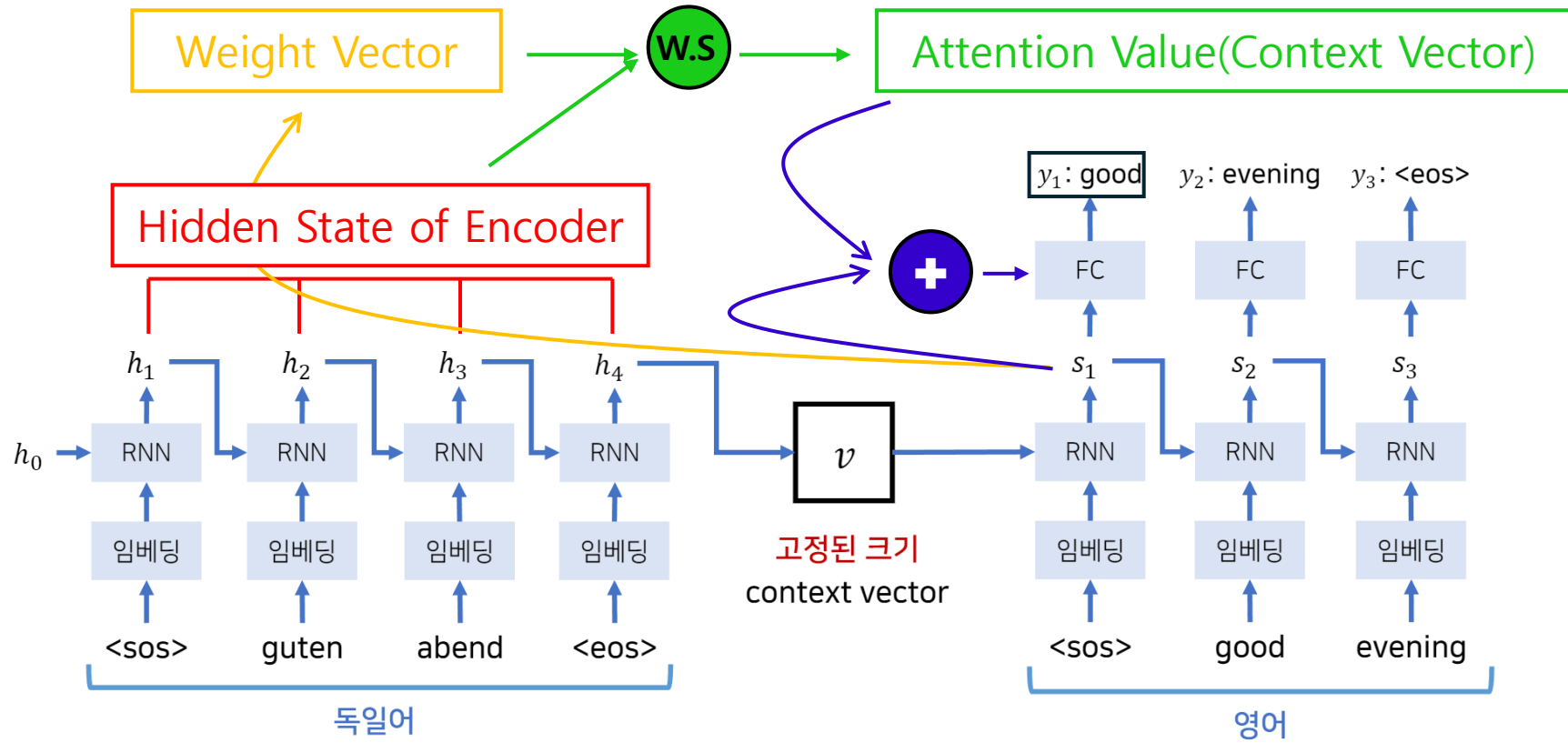
Neural Machine Translation SEQUENCE TO SEQUENCE MODEL WITH ATTENTION



Summary



seq2seq vs seq2seq with attention



Result

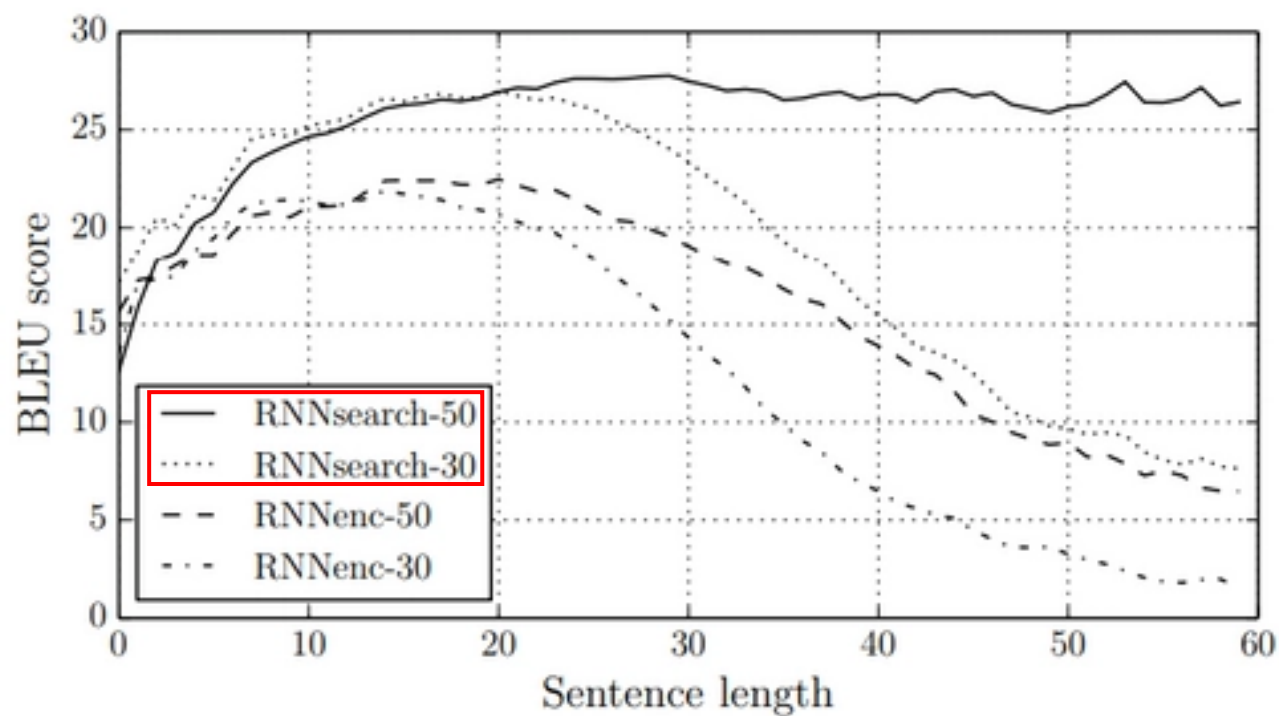
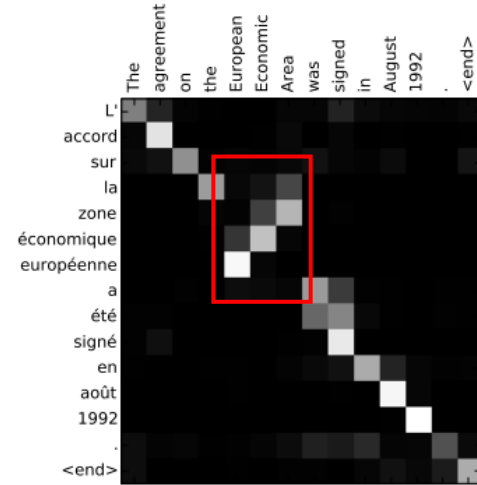
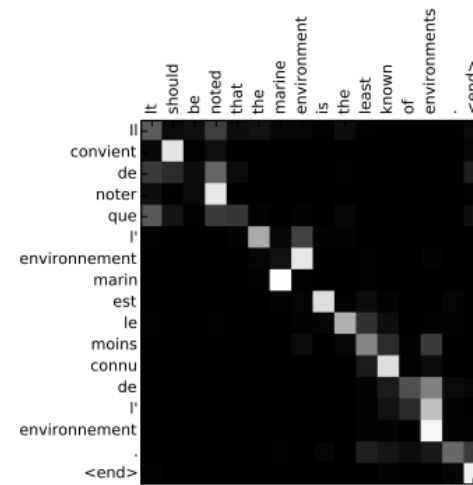


Figure 2: The BLEU scores of the generated translations on the test set with respect to the lengths of the sentences. The results are on the full test set which includes sentences having unknown words to the models.

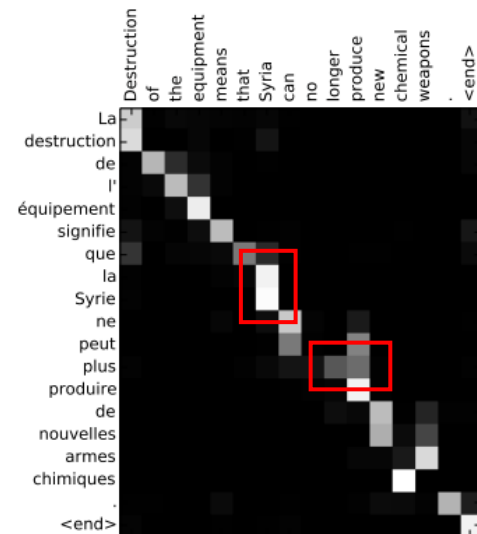
Result - Qualitative Analysis



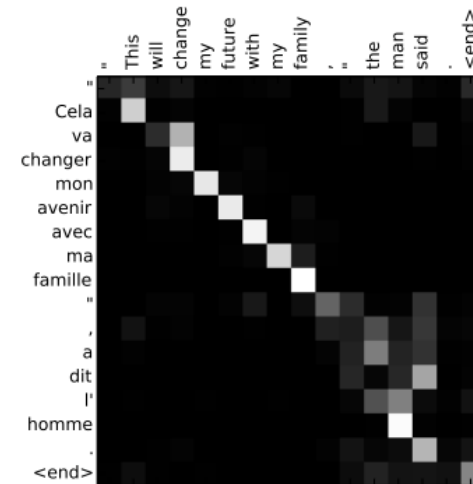
(a)



(b)



(c)



(d)