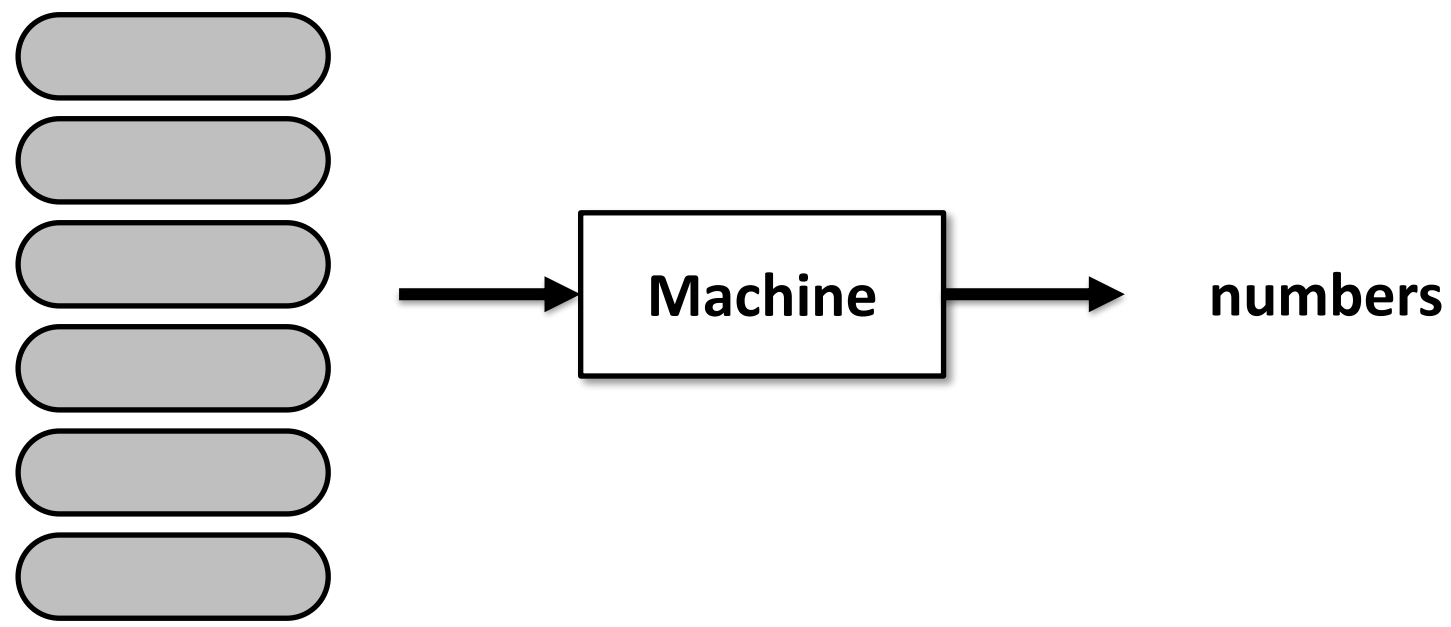


Understanding Attention Mechanism

정상근

Big Picture



Motivation and Metaphor of Attention

BLENDING



Blending

Sequence of Information 을 Blending 해보는 건 어떨까?

<http://artquestionsanswered.com/wp-content/uploads/2014/05/blending-Colored-Pencil-Techniques.jpg>

We want to create best **reactive material** to liquid Q
using multiple materials I_1 to I_5



I_1

I_1

I_2

I_1

I_3

I_1

I_4



I_5

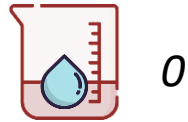
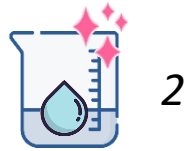
Idea

Blend Q and each item, and mix tasty ones

① Mix



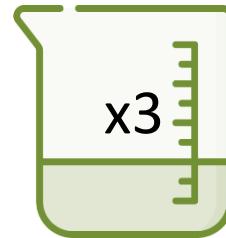
② Check degree of reactivity



③ Mix materials with reactivity proportion



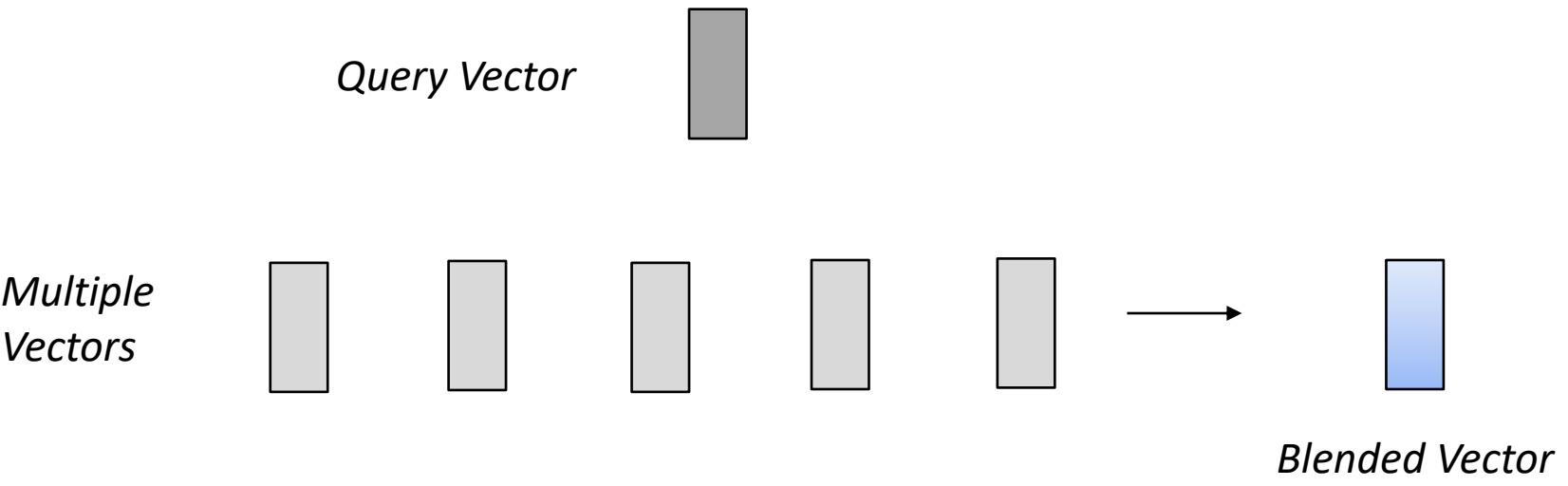
+



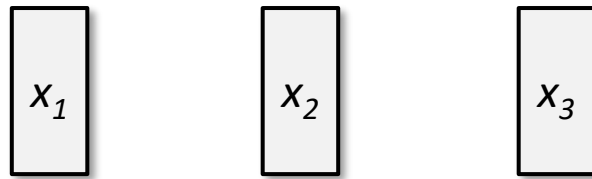
④ Blended Best Reactive material!



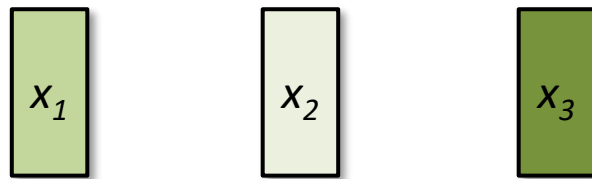
Attention as blending metaphor



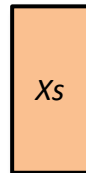
Attention Without Query – In/Out View



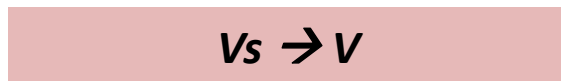
- There are more important items than other



- Consider importance (reactivity)




- Element wise summation
- = Blending

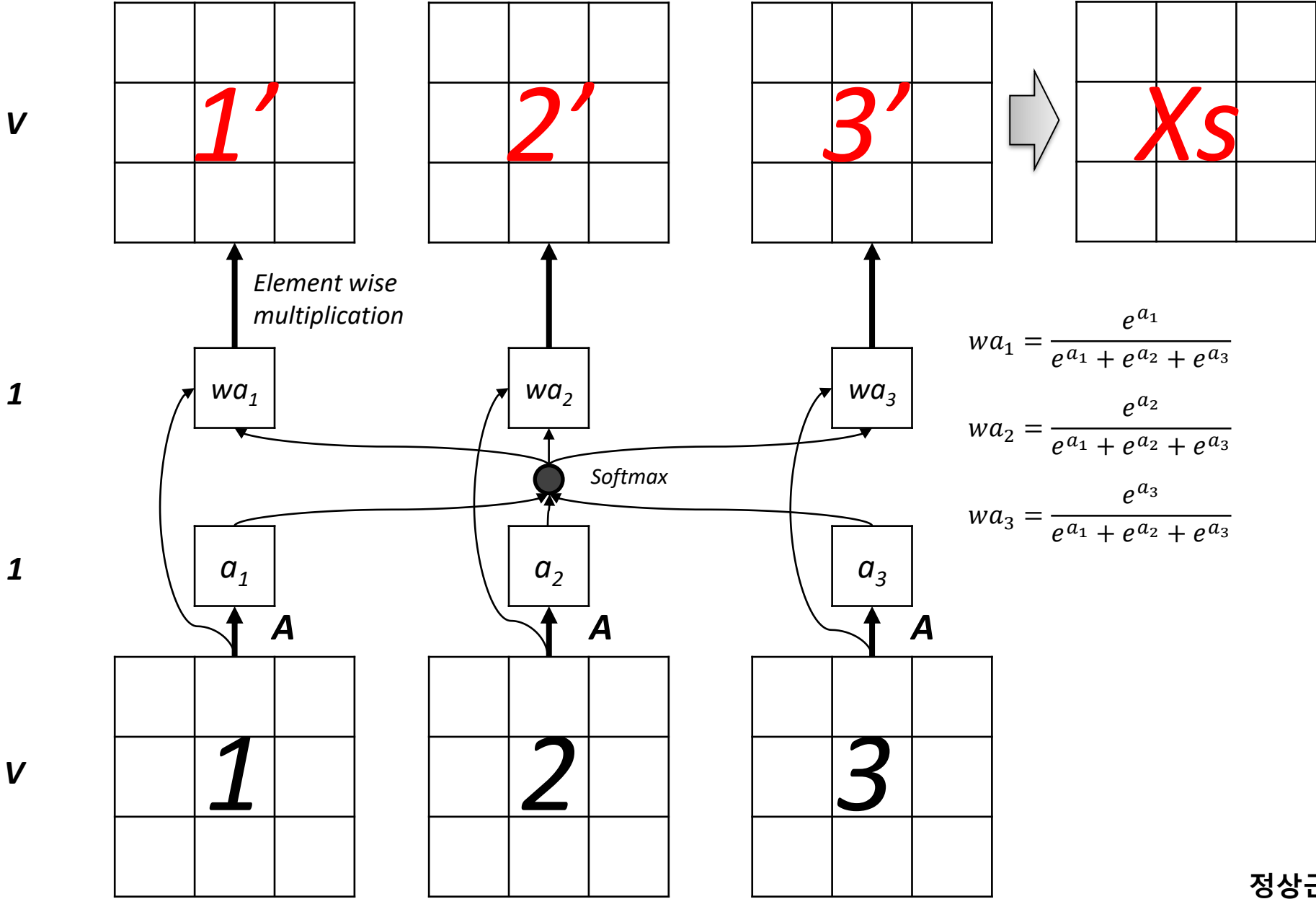


Summary the items using the 'importance' of the items

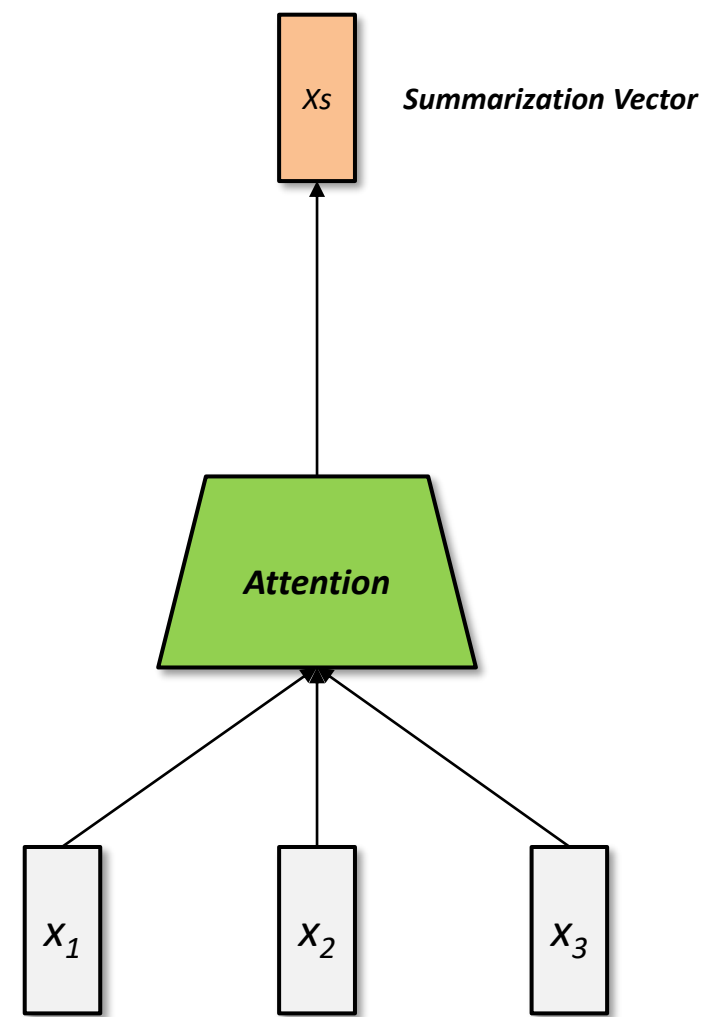
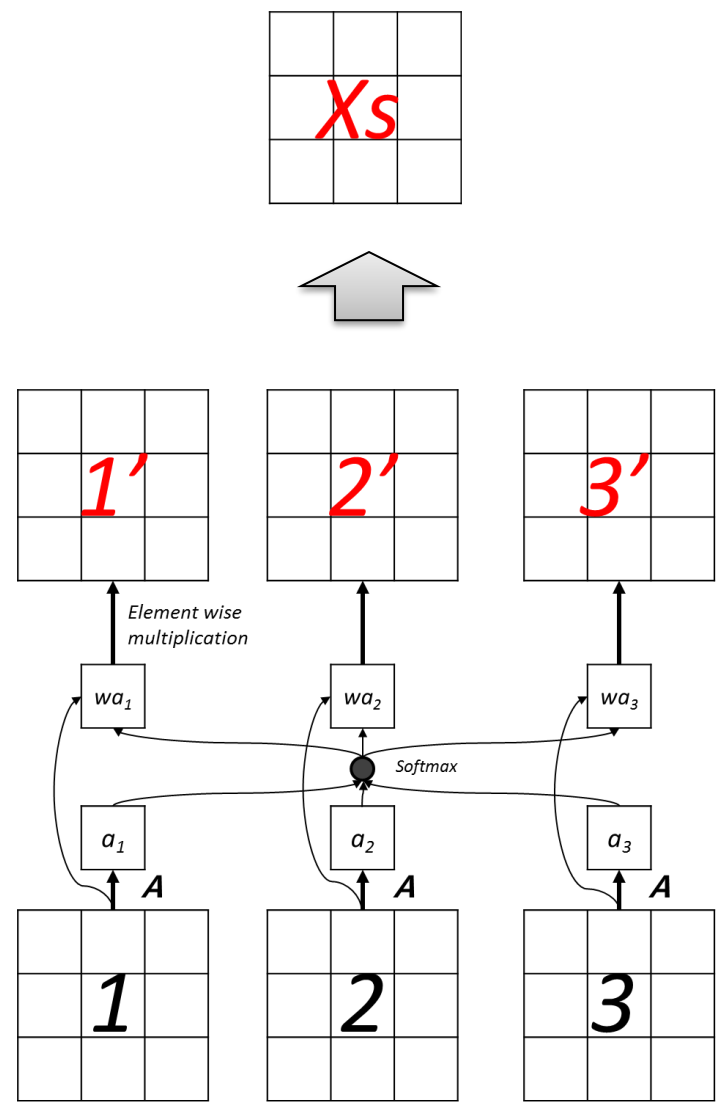
Attention

Attention without Query

Element wise
Summation 

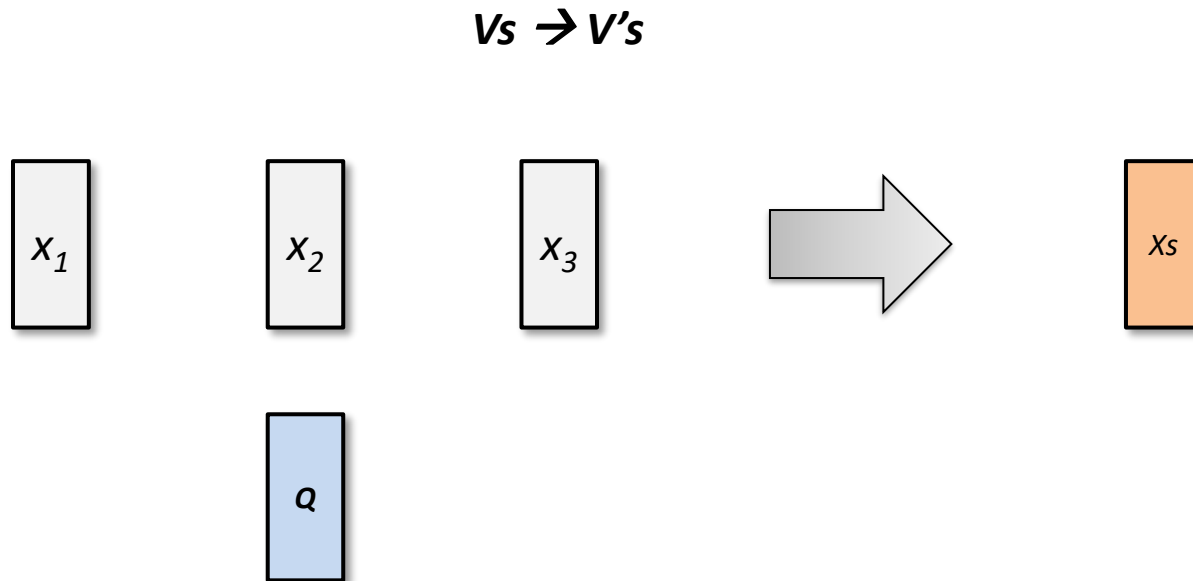


Attention Without Query – Simple View



Attention Query – In/Out View

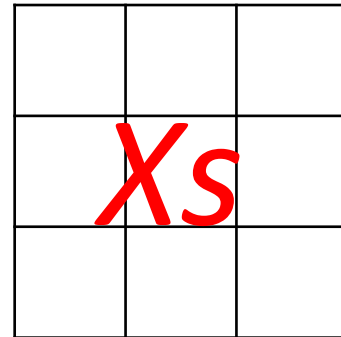
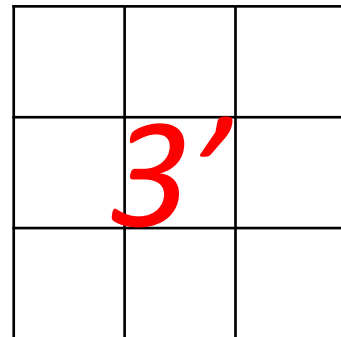
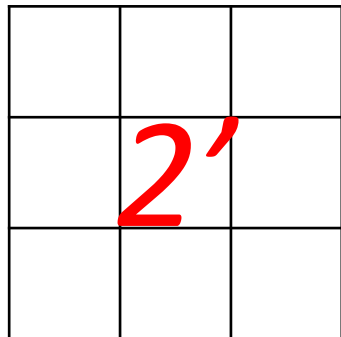
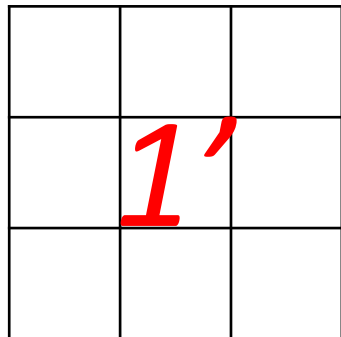
- 1) Calculating attention scores from the sequence data
- 2) Applying the attention scores to the original data



Attention with Query

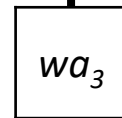
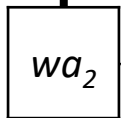
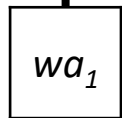
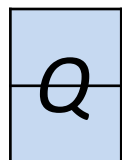


v

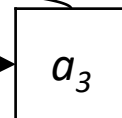
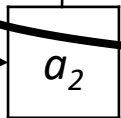
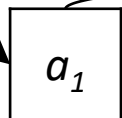


Element wise multiplication

1



1



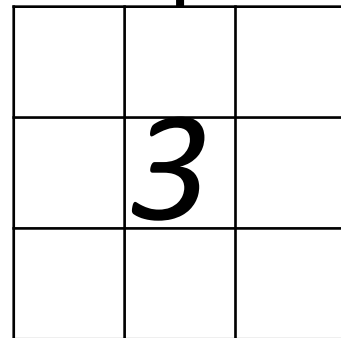
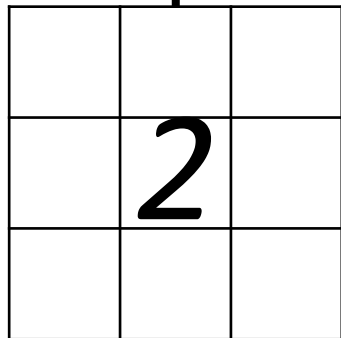
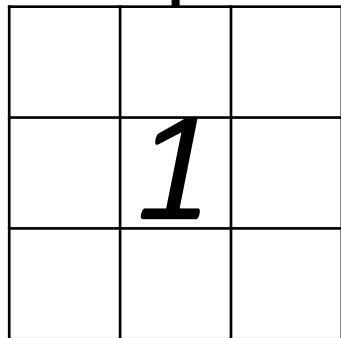
Softmax

$$wa_1 = \frac{e^{a_1}}{e^{a_1} + e^{a_2} + e^{a_3}}$$

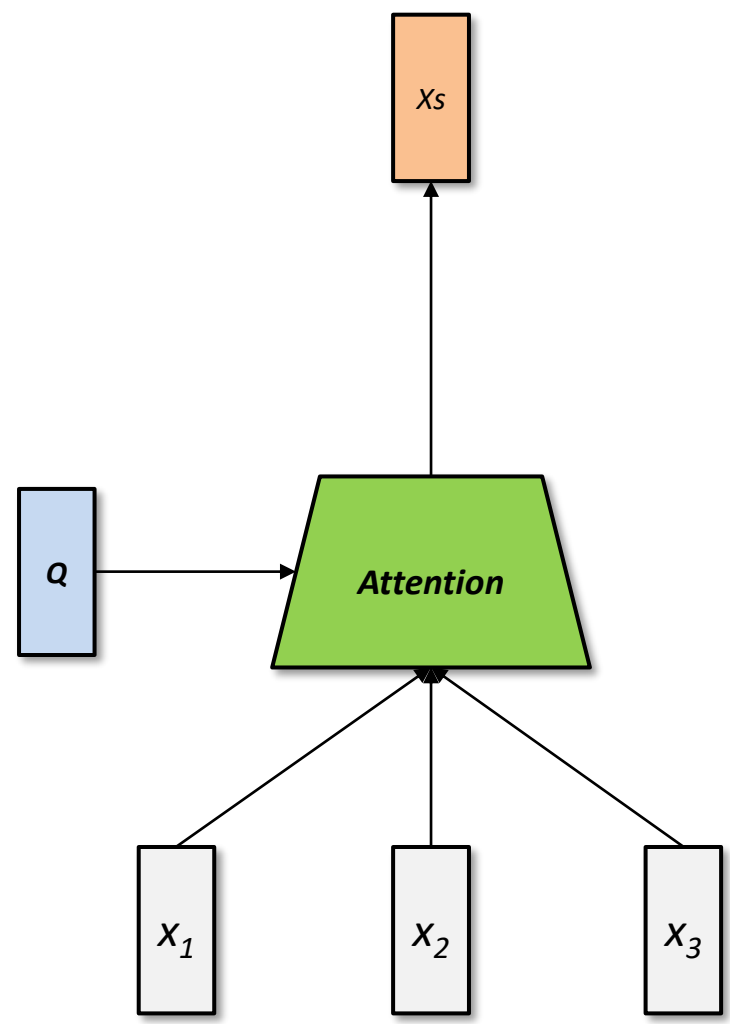
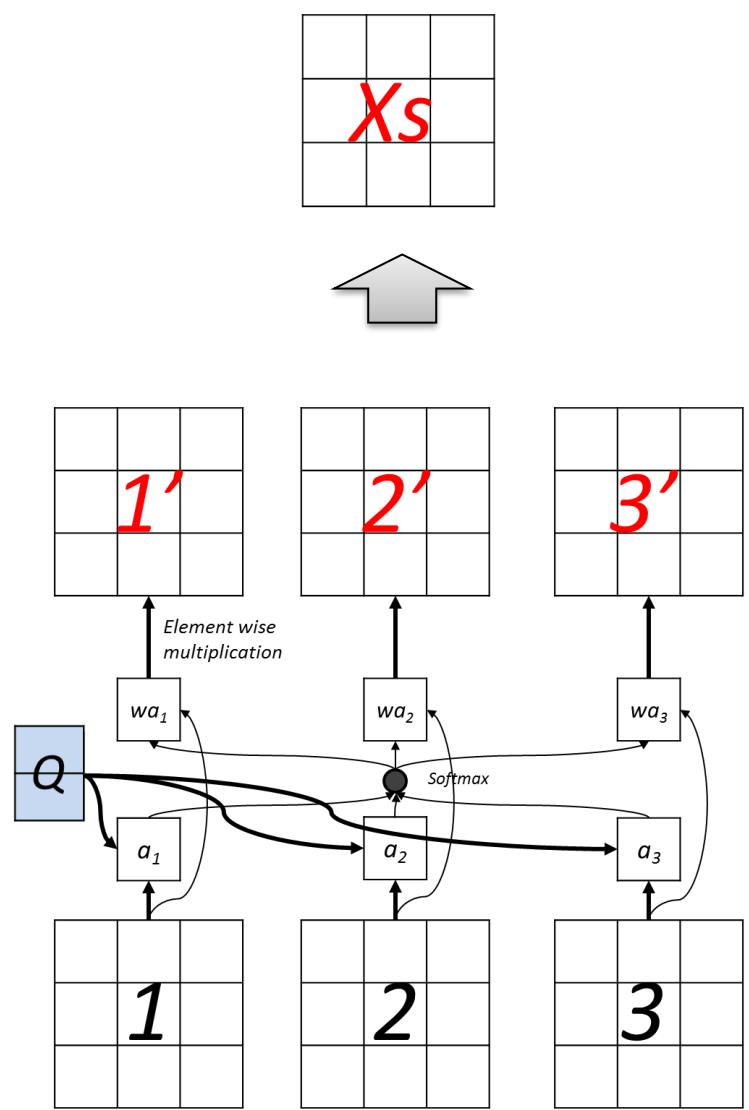
$$wa_2 = \frac{e^{a_2}}{e^{a_1} + e^{a_2} + e^{a_3}}$$

$$wa_3 = \frac{e^{a_3}}{e^{a_1} + e^{a_2} + e^{a_3}}$$

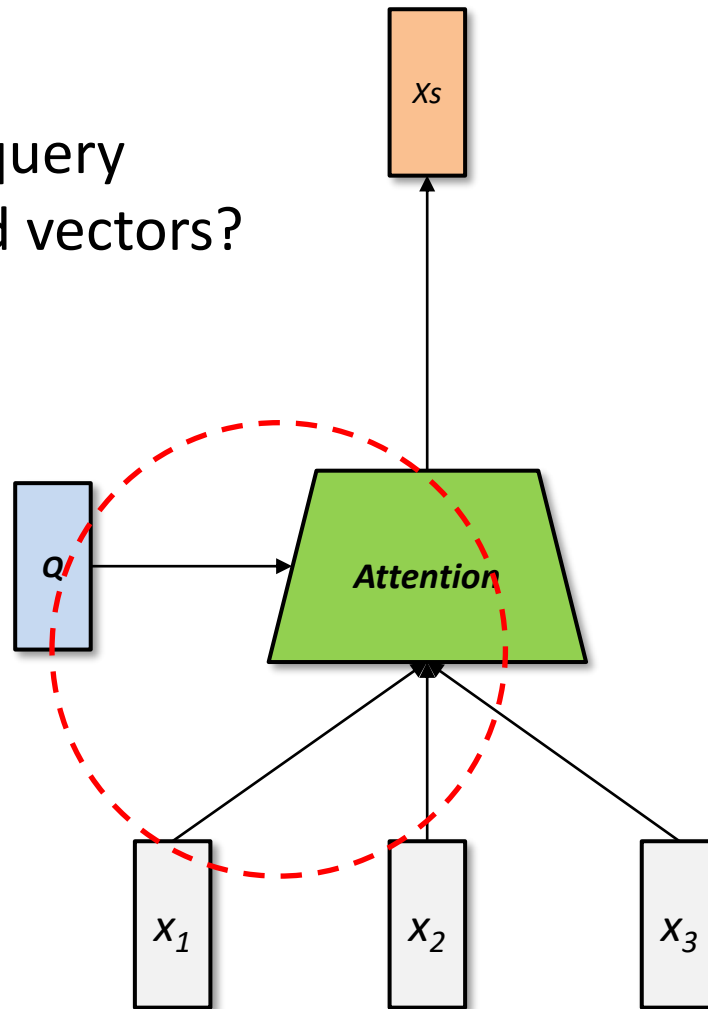
v



Attention With Query – Simple View



[Question]
How to apply query
vector to blend vectors?



Query Apply Methods

Additive Approach

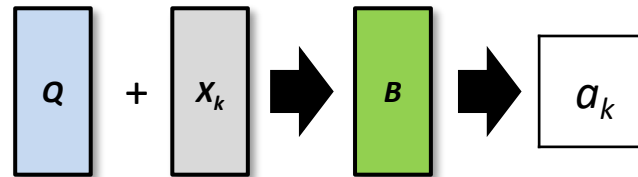
Reactivity Scoring
by Addition

Multiplicative Approach

Reactivity Scoring by
Multiplication

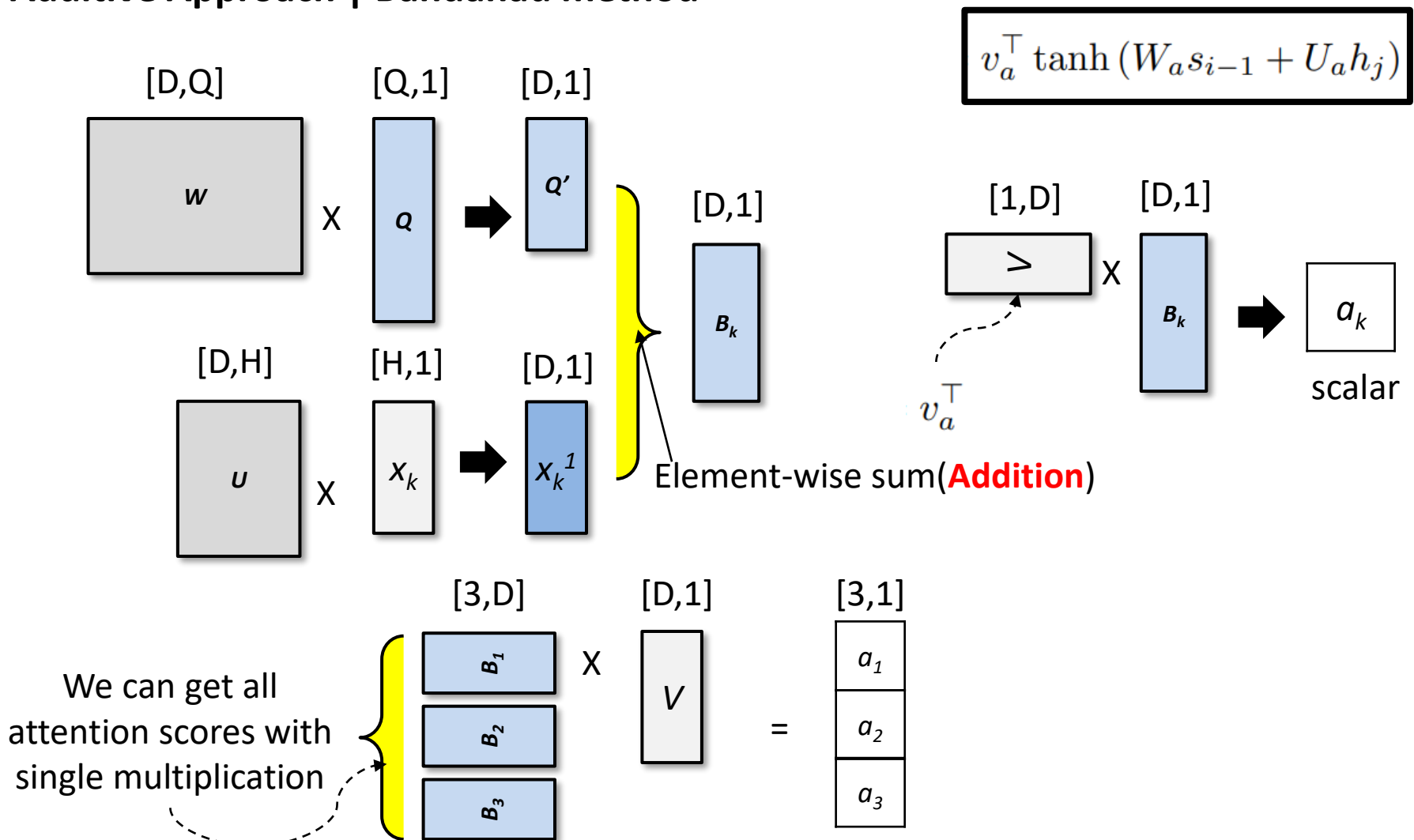
Additive Approach | Bahdanau method

Bahdanau et al. (2014)



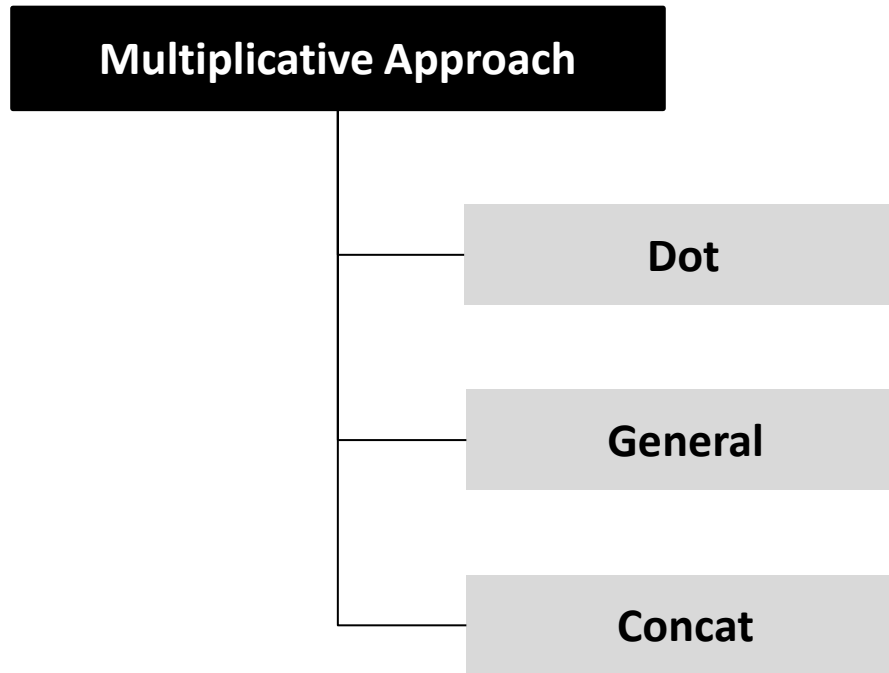
- Bahdanau attention
- First proposed mechanism

Additive Approach | Bahdanau method

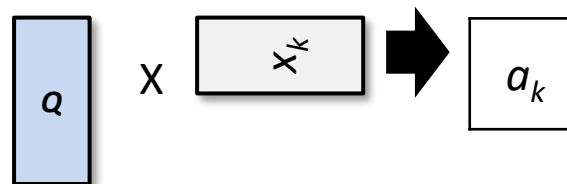


- It is working even when **dim(Q) != dim(H)**
- Blend by addition. (it's **additive** attention)
- Use additional parameter V to make added vector to be scalar
- You can find more optimized operation – bmm (batch mat-mul)

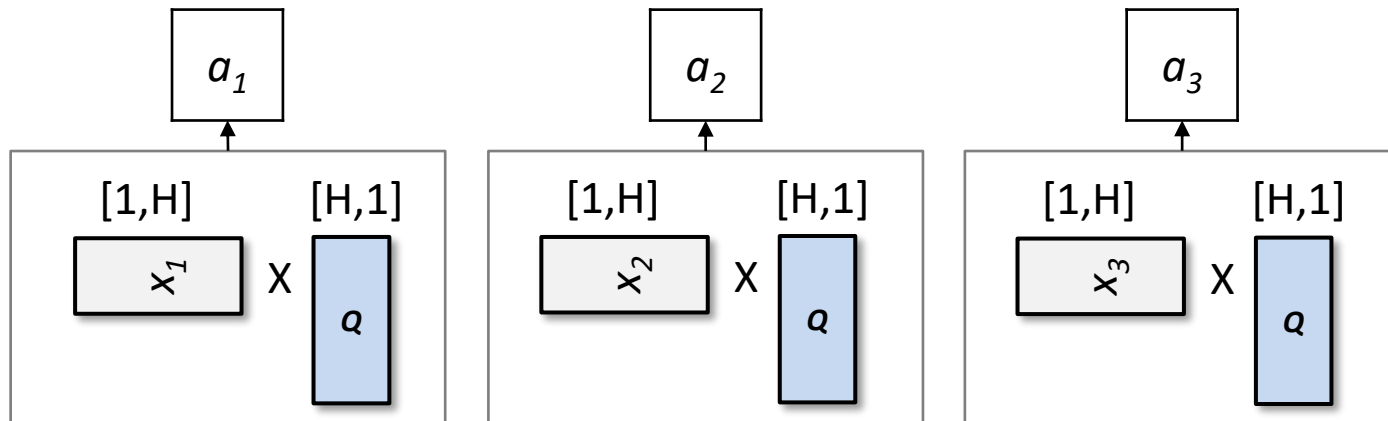
Multiplicative Approach



Luong et al. (2015)



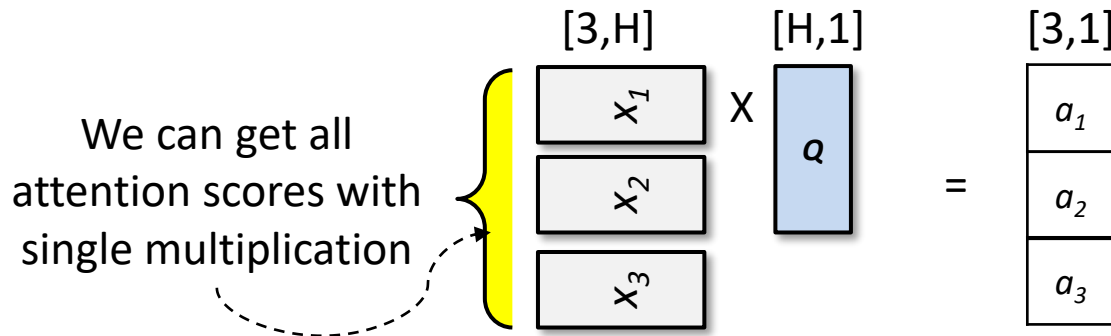
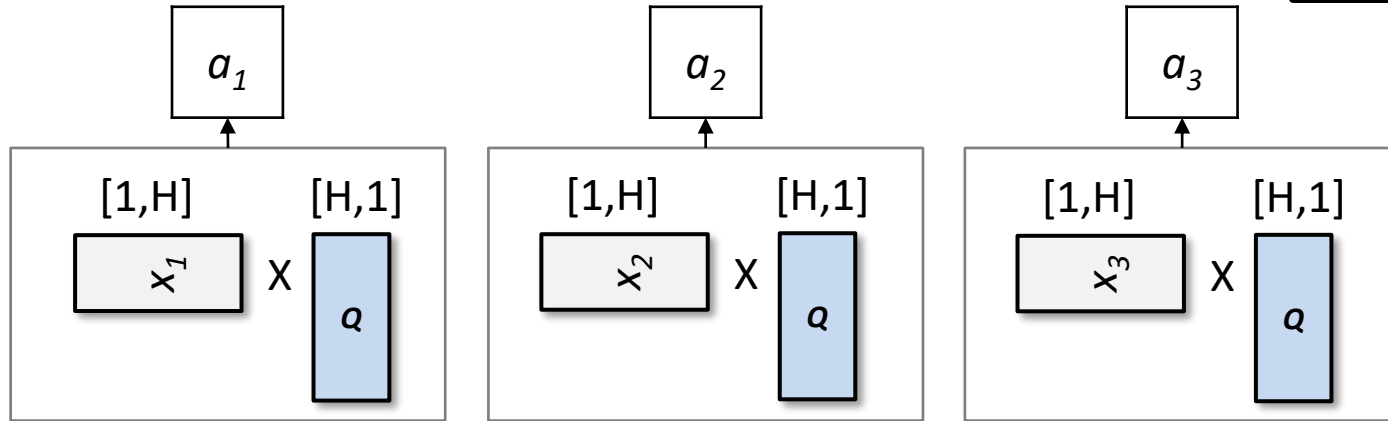
$$\text{score} = qx_i^T$$



- For loop over multiple items N time
- It conducts N dot operation
- In optimized code, we do not use this approach

Dot method | Matrix Multiplication Approach

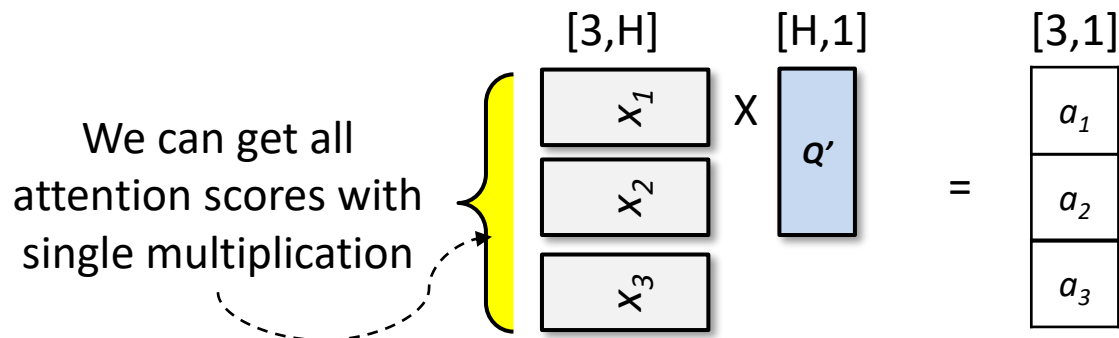
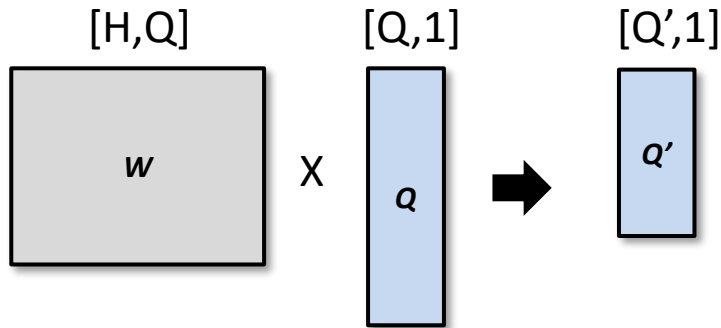
$$\text{score} = qx_i^T$$



- Replace for-loop with single multiplication
- It conducts 1 dot operation
- You can find more optimized operation – bmm (batch mat-mul)

General method | matrix multiplication approach

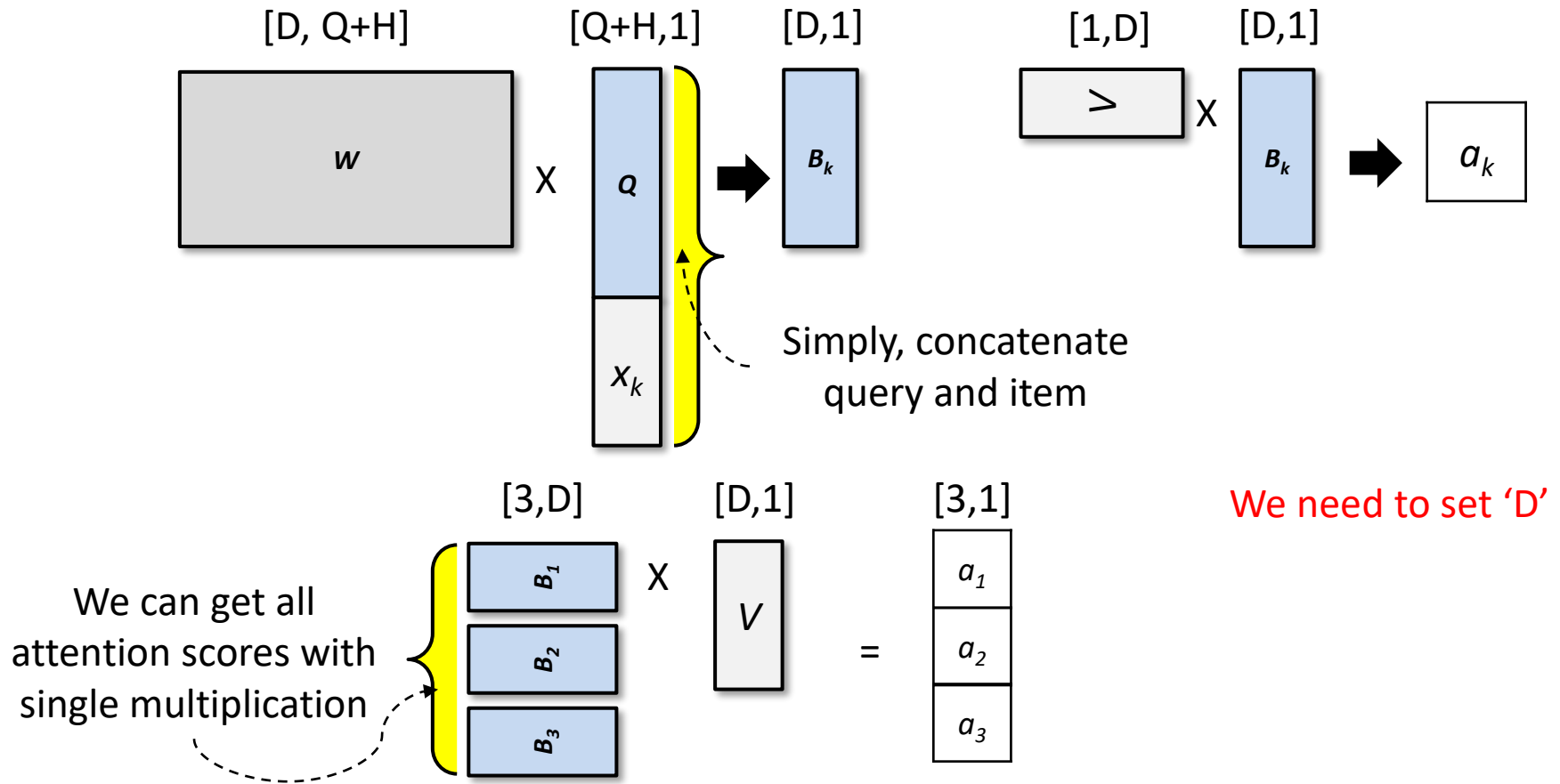
$$\text{score} = cWx_i^T$$



- It is working even when **$\text{dim}(Q) \neq \text{dim}(H)$**
- Convert query to be same shape of input item
- It conducts 1 dot operation
- You can find more optimized operation – bmm (batch mat-mul)
- It is more generalized version of 'dot' method

Concat method | matrix multiplication approach

$$\text{score} = v^T \tanh(W[x_i; c])$$



- It is working even when **$\dim(Q) \neq \dim(H)$**
- Blend query and item by concatenation
- Use additional parameter V to make concated vector to be scalar
- You can find more optimized operation – bmm (batch mat-mul)

Case Study

ATTENTION FOR ENCODING- DECODING

Attention Modeling

:: Korean → English translation

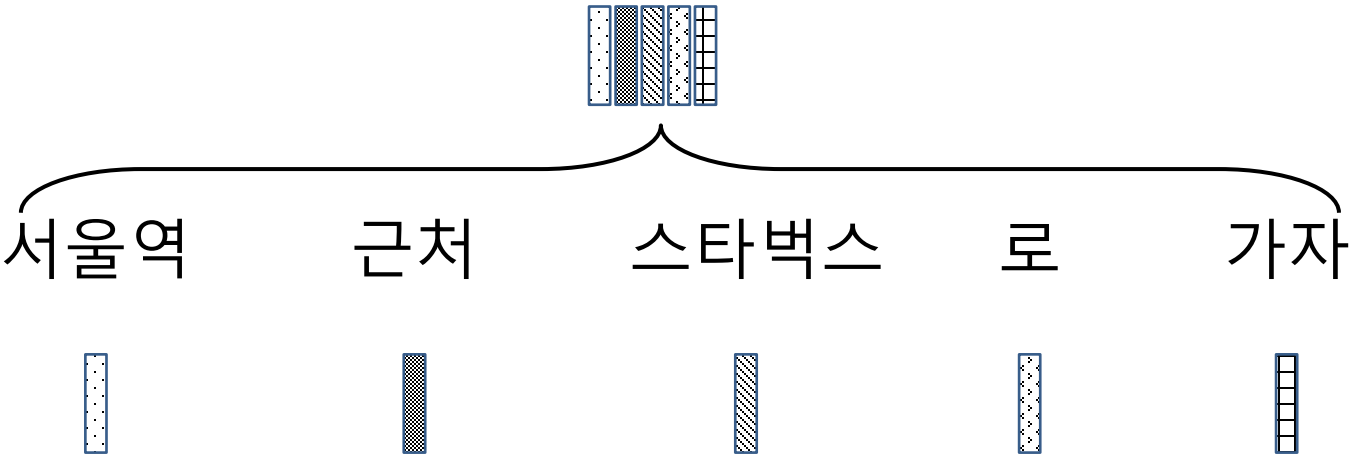
Let's go to Starbucks near Seoul station

서울역 근처 스타벅스 로 가자

Attention Modeling

Encoding

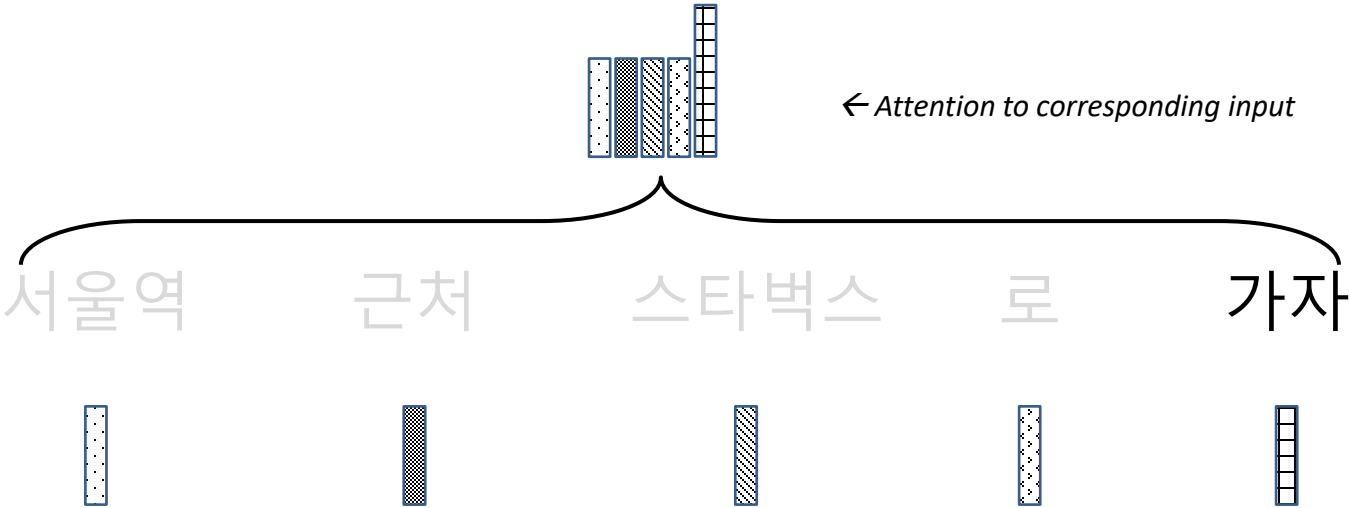
Let's go to Starbucks near Seoul station



Attention Modeling

Decoding

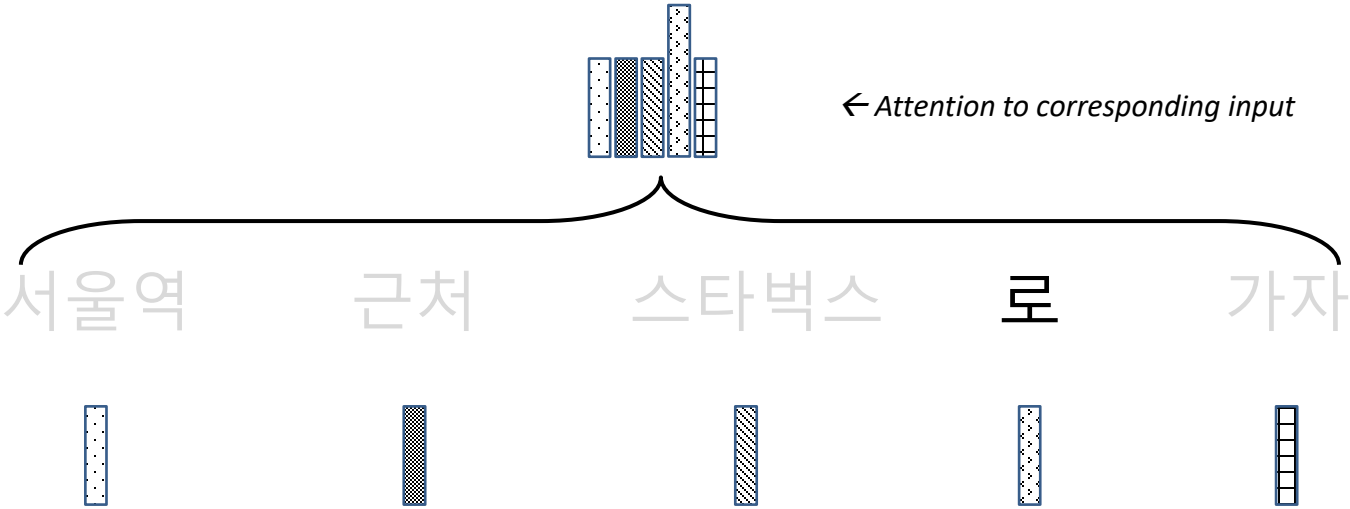
Let's go to Starbucks near Seoul station



Attention Modeling

Decoding

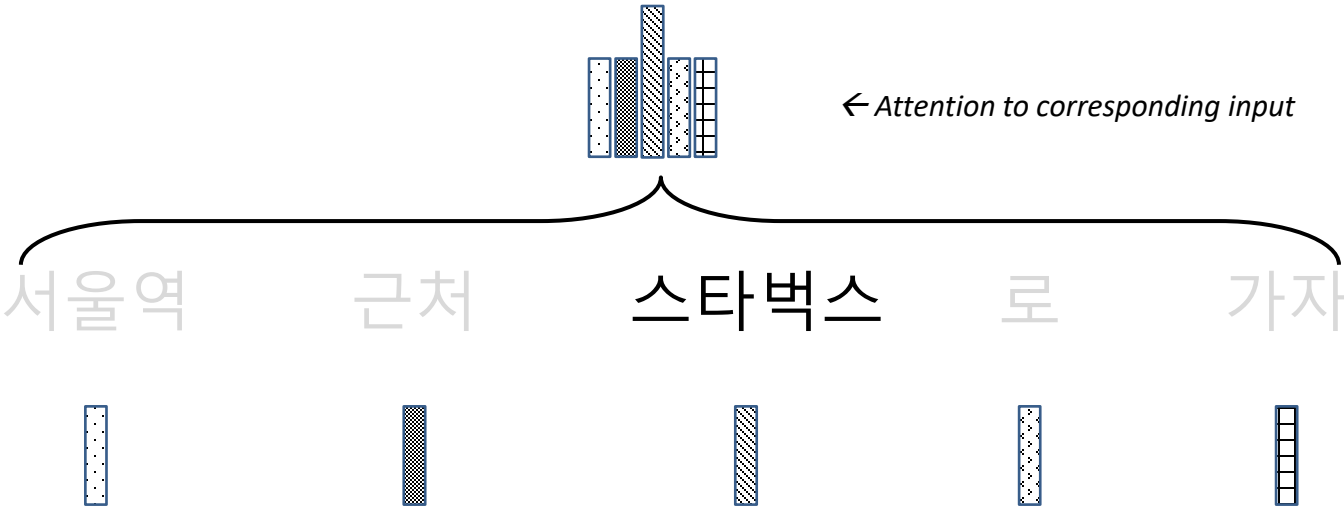
Let's go to Starbucks near Seoul station



Attention Modeling

Decoding

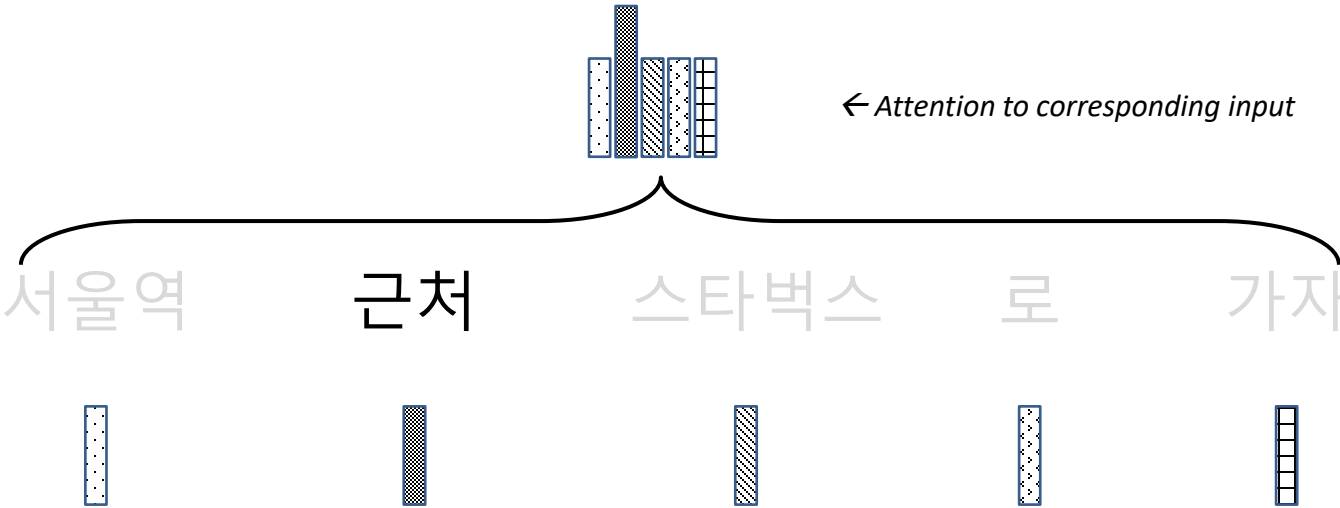
Let's go to Starbucks near Seoul station



Attention Modeling

Decoding

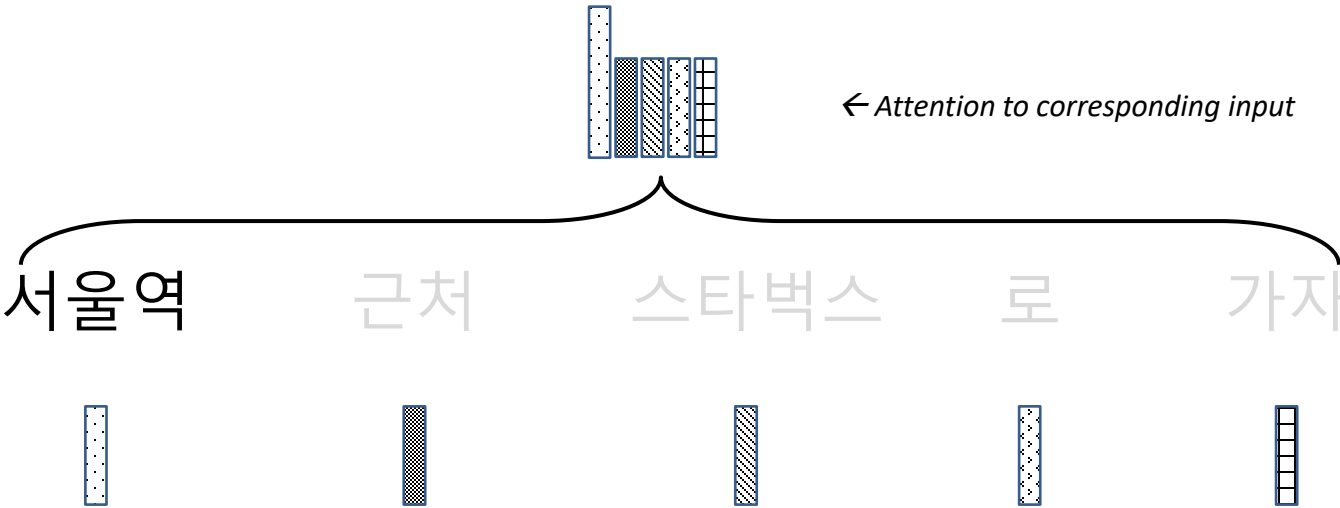
Let's go to Starbucks near Seoul station



Attention Modeling

Decoding

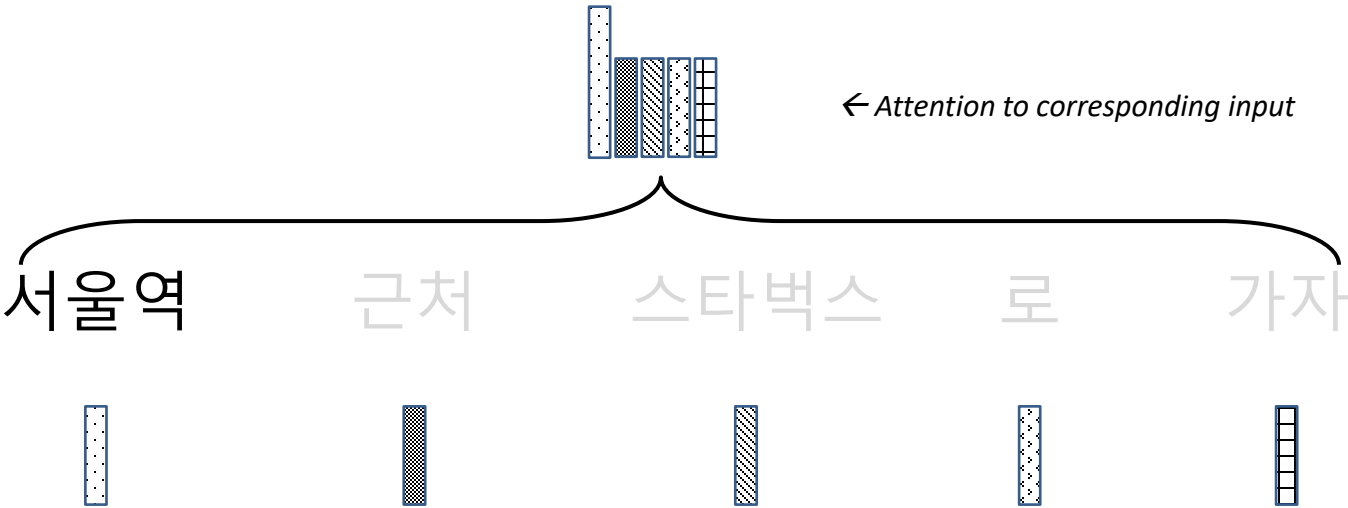
Let's go to Starbucks near Seoul station



Attention Modeling

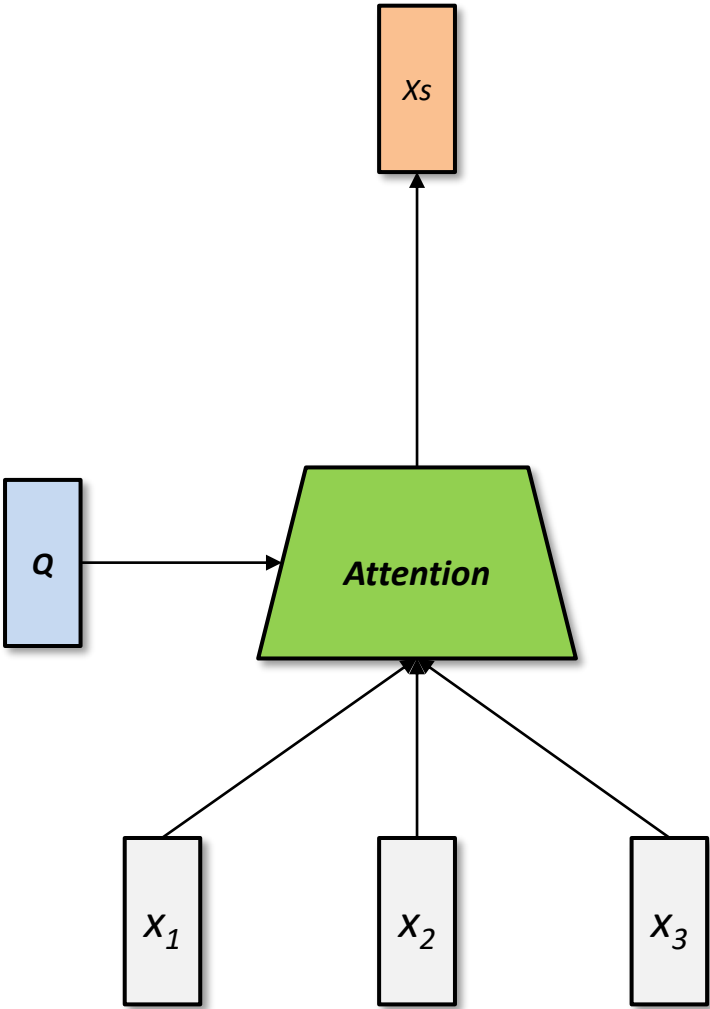
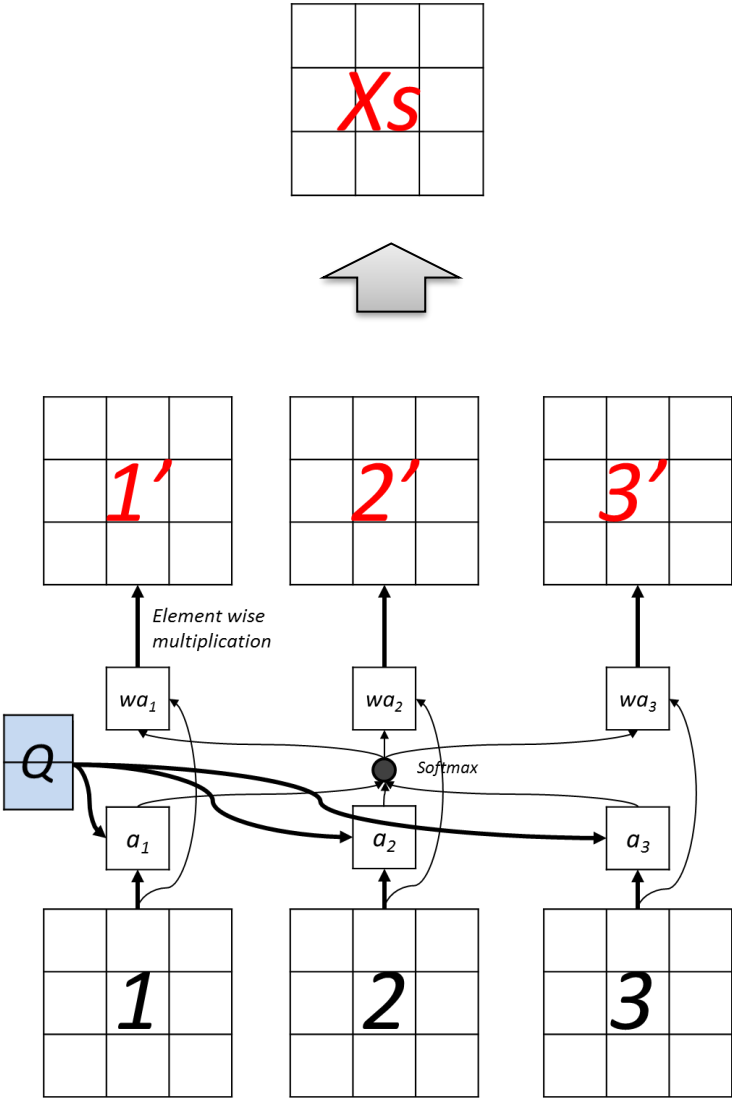
Decoding

Let's go to Starbucks near Seoul station

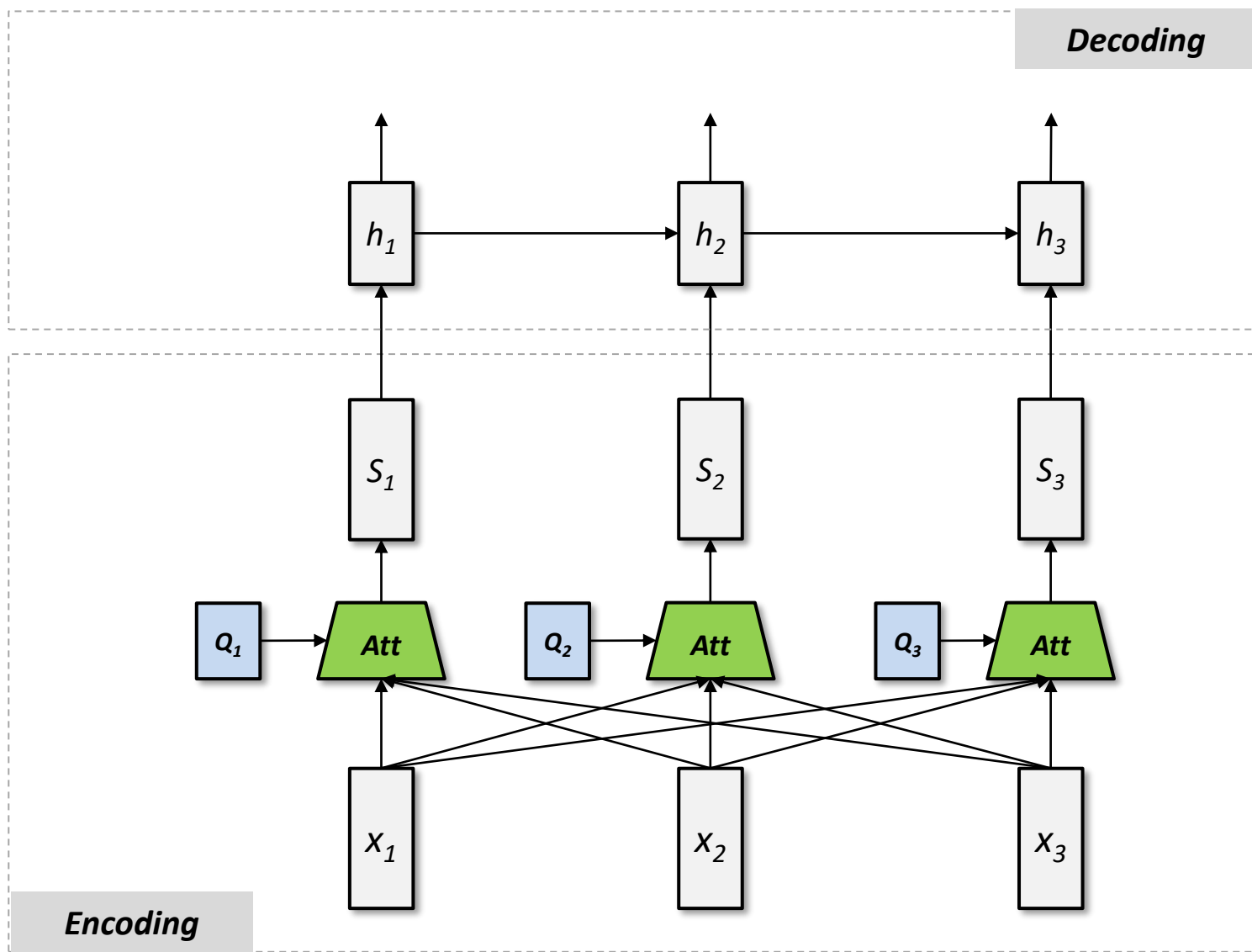


:: Decode more correct terms by dynamically summarizing the encoded results.

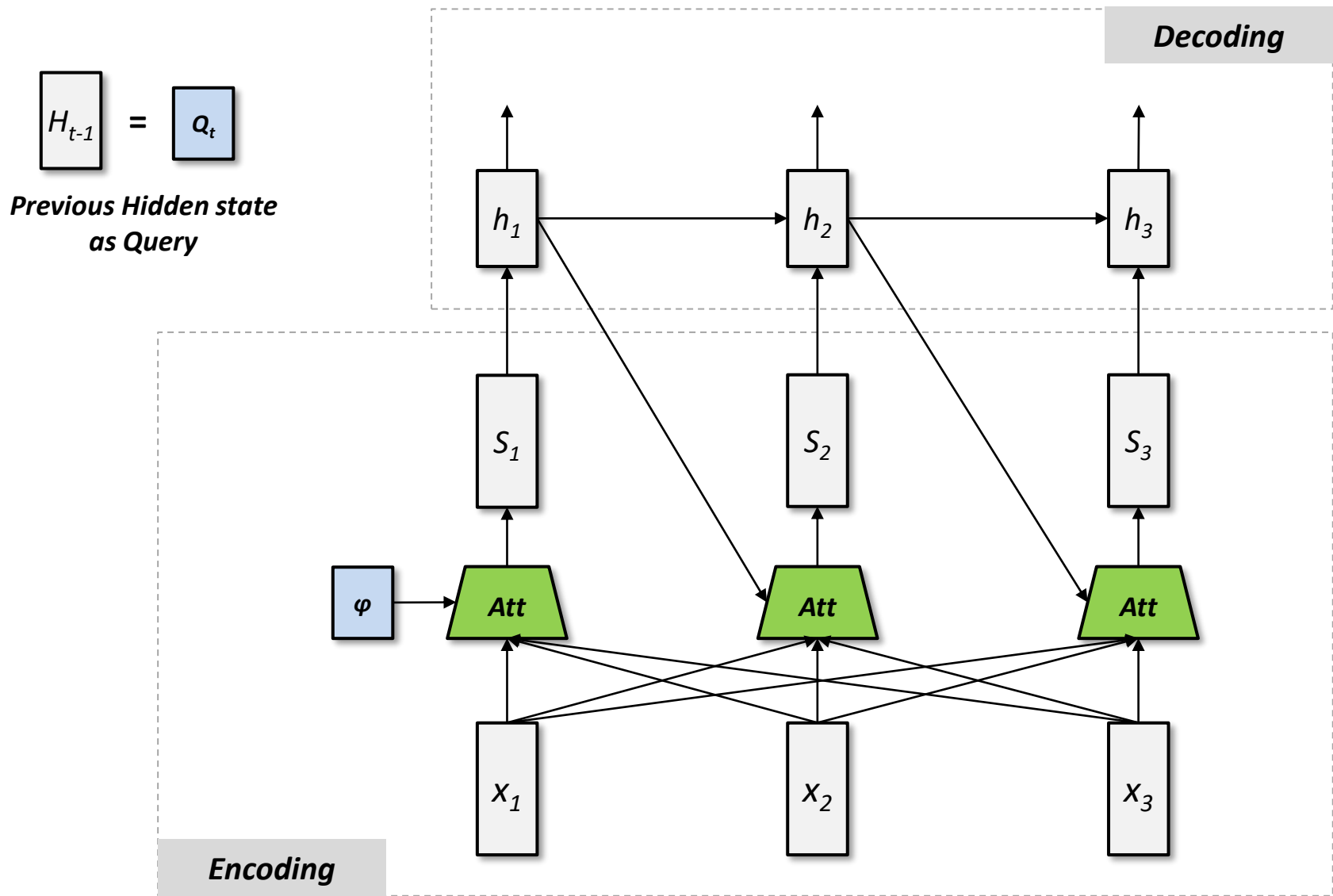
[Review] Attention With Query – Simple View



Encoding + Attention Decoding With Query



Encoding + Attention Decoding With Query (from prev. RNN output)



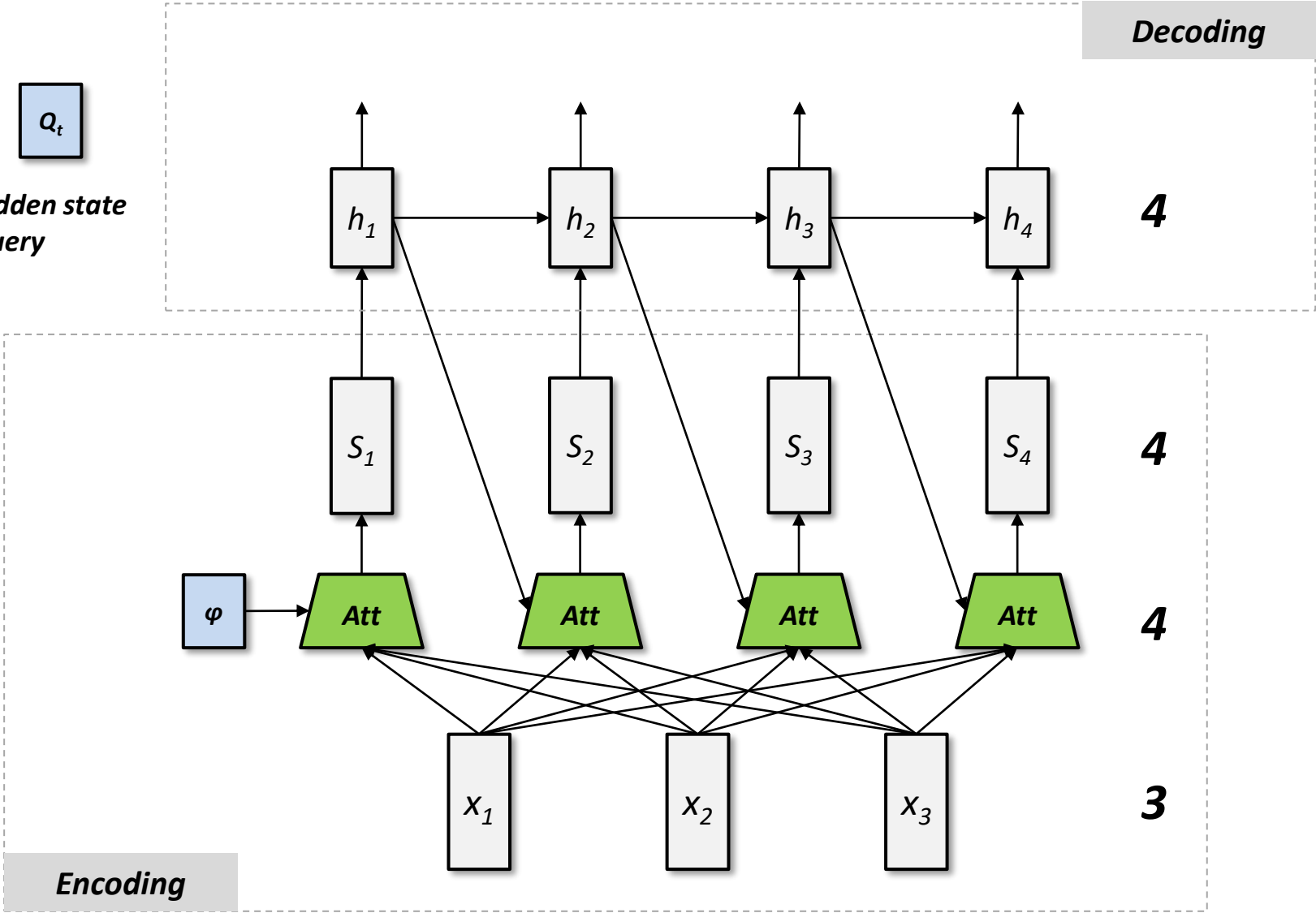
[Question]

**The number of input items should same
the number of output items?**

No!

Encoding + Attention Decoding With Query (from prev. RNN output)

H_{t-1} = Q_t
Previous Hidden state
as Query



Len(Output Sequence) = Number of Attention Module

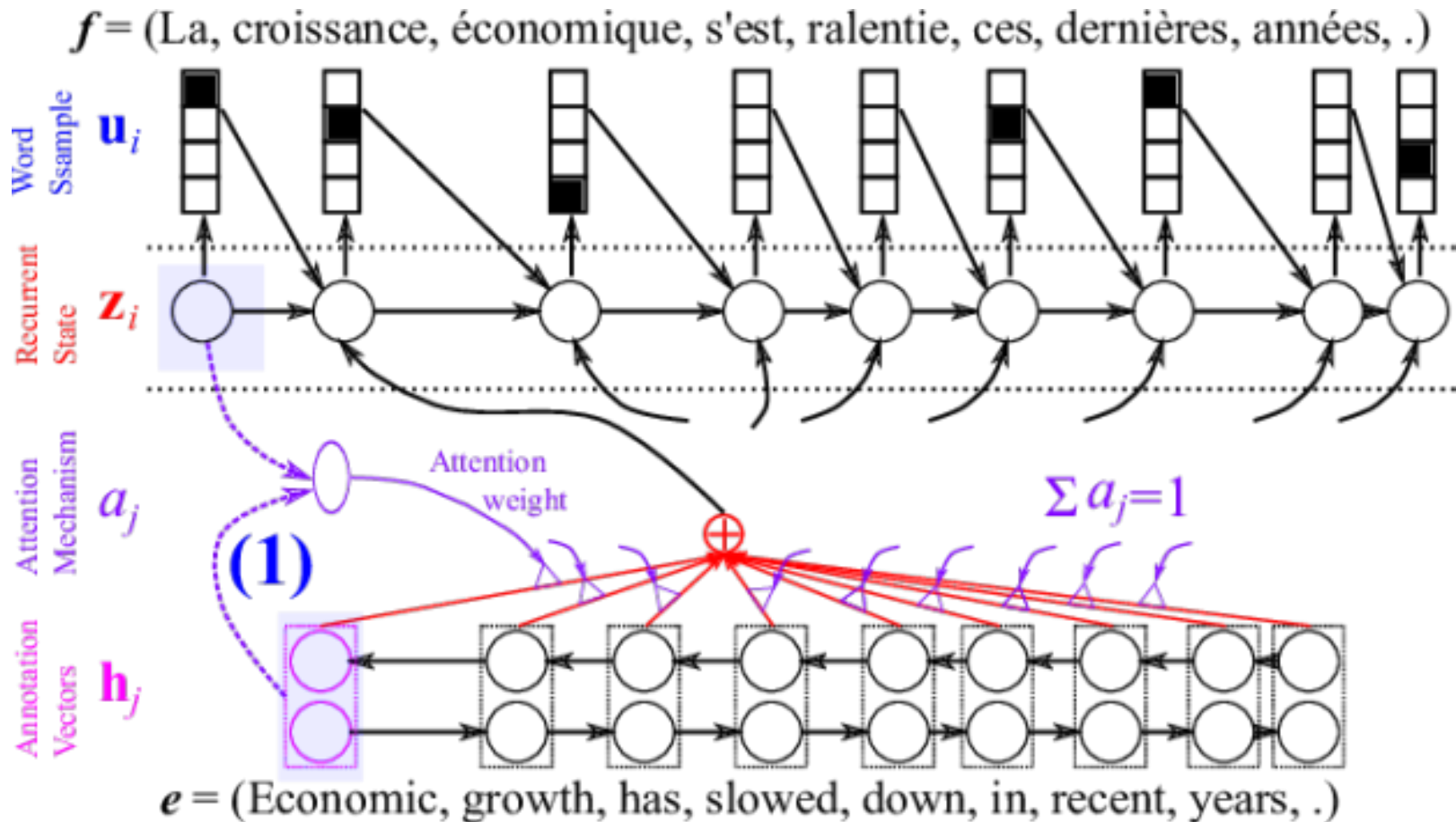
SANGKEUN JUNG

Attention Mechanism

=

Global, Selective, Dynamic
Sequence Summarization(Blending)

Case | Attention Modeling

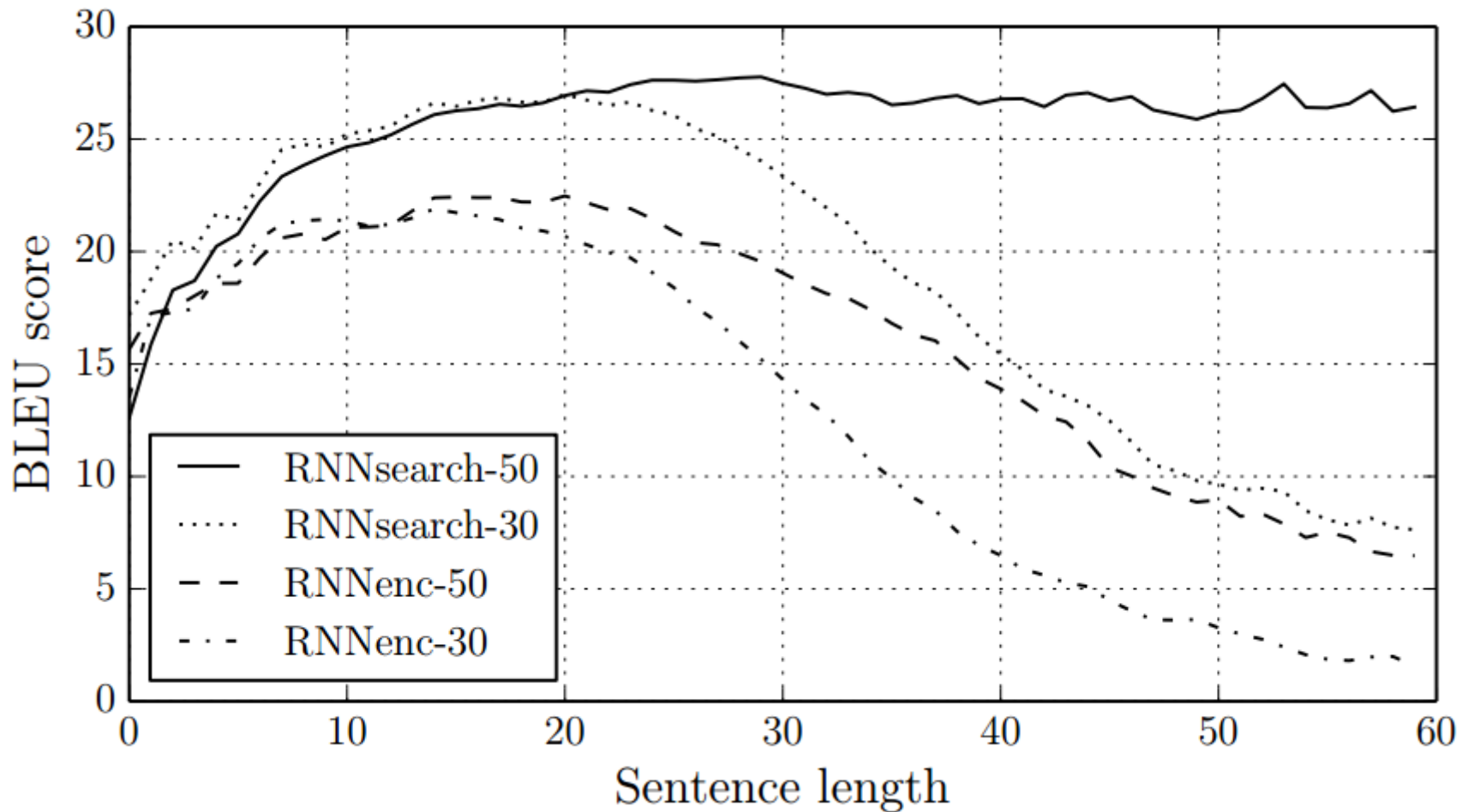


- Bidirectional RNN for Encoding
- Attention Modeling

Bahdanau et al. (2014)

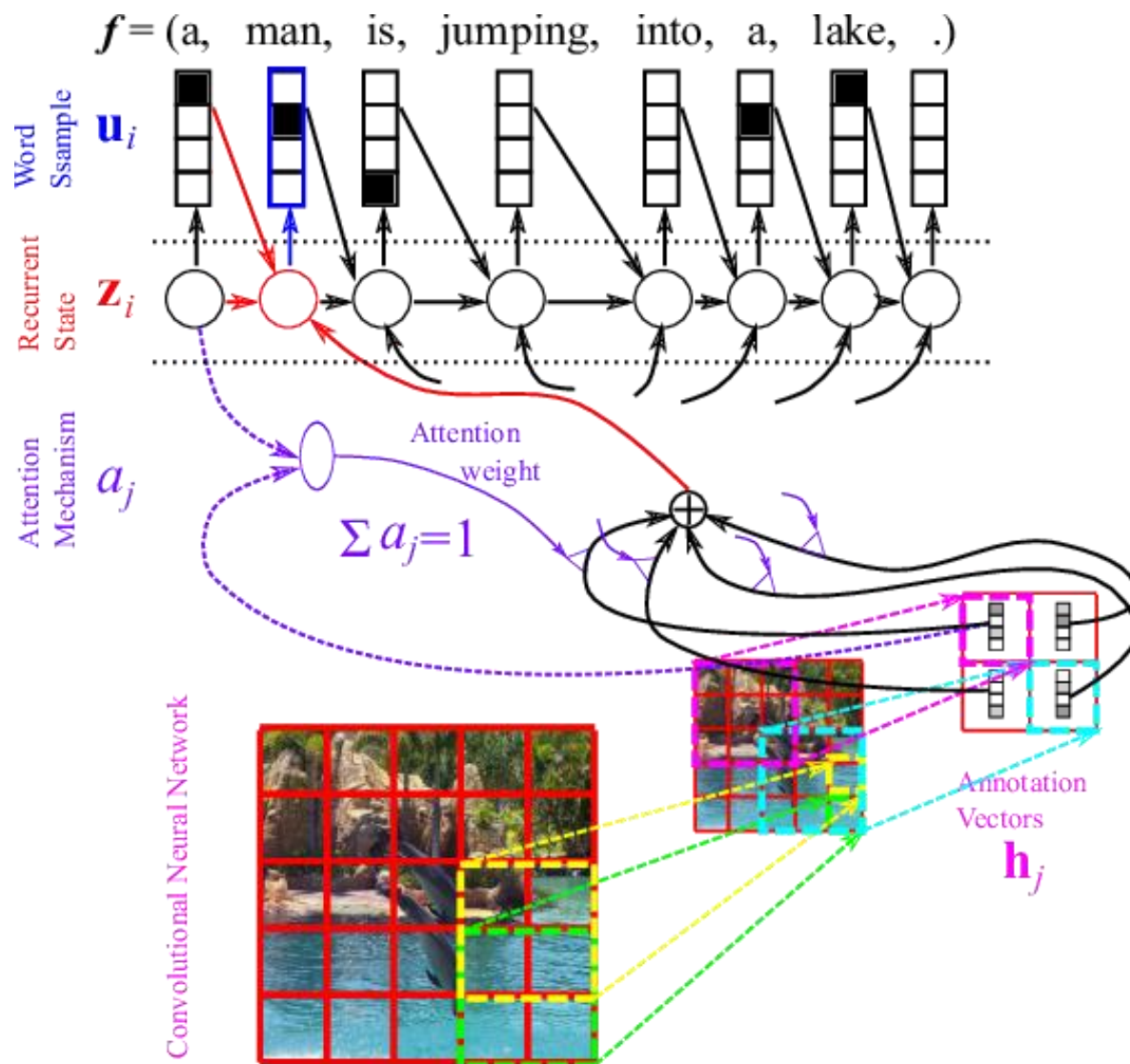
<http://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-3/>

Case | Performance – Attention Modeling @ Machine Translation



:: 선별적으로 가중치가 적용된 Encoding 이 적용됨으로서, 긴 문장에서도 번역 성능이 떨어지지 않는다.

Case | Attention Modeling for Image2Text



<http://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-3/>

Xu et al. (2015)

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Attention Modeling for Image2Text

Encoder / Decoder 에서 Text Sequence Encoding 을
Image Sequence Encoding 으로 교체만 해도 작동함

X



Y

A woman is throwing a frisbee in a park.

X



Y

A little girl sitting on a bed with
a teddy bear.

Xu et al. (2015)

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Copyright, Sangkeun Jung, All Rights Reserved.