

# DL Seminar

## Sequence-to-Sequence

한양대학교 AI Lab  
이 주 홍



Introduction



Seq2Seq



Application



Tutorial



Q&A

- Sequence-to-Sequence (Seq2Seq)란 무엇인가?
- input = Sequence  
output = Sequence
- 여기서 Sequence는 문장(sequence of words), 신호, 영상 등 시계열 데이터
- Seq2Seq는 RNN의 확장판 중 하나라고 볼 수 있음.
- 기계번역(Machine Translation), 대화시스템(Dialog System; Chatbot) 등에서 좋은 성과를 보임.

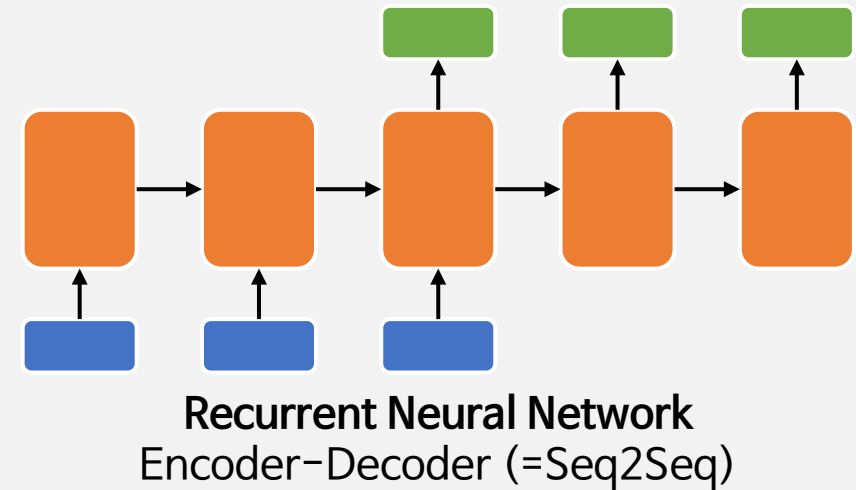
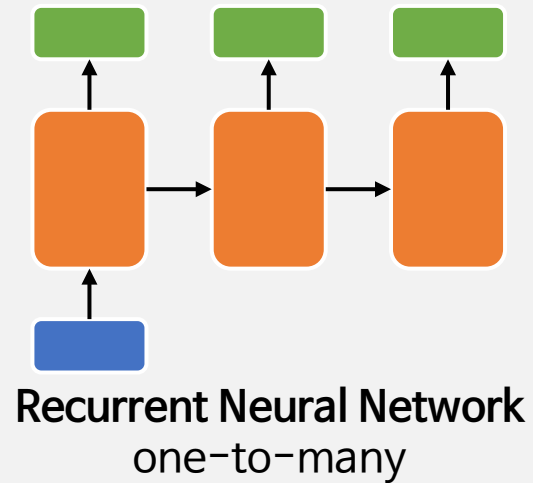
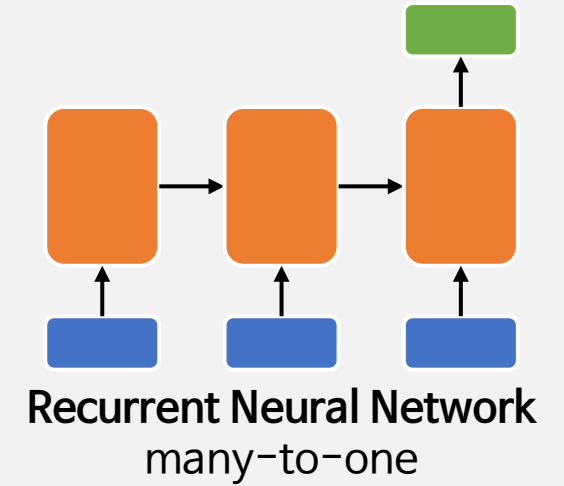
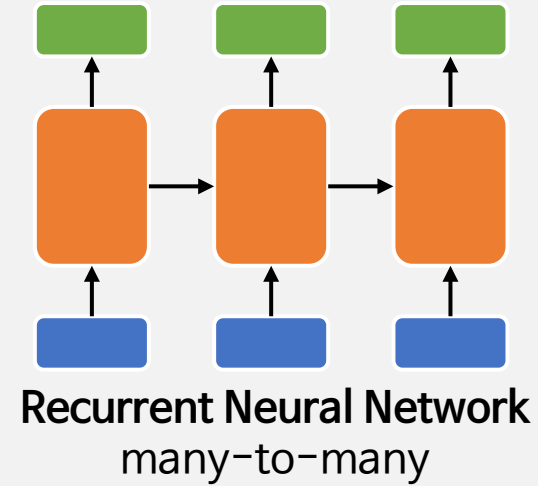
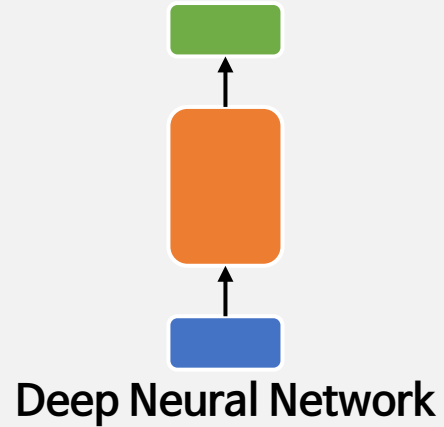
 Introduction

 Seq2Seq

 Application

 Tutorial

 Q&A



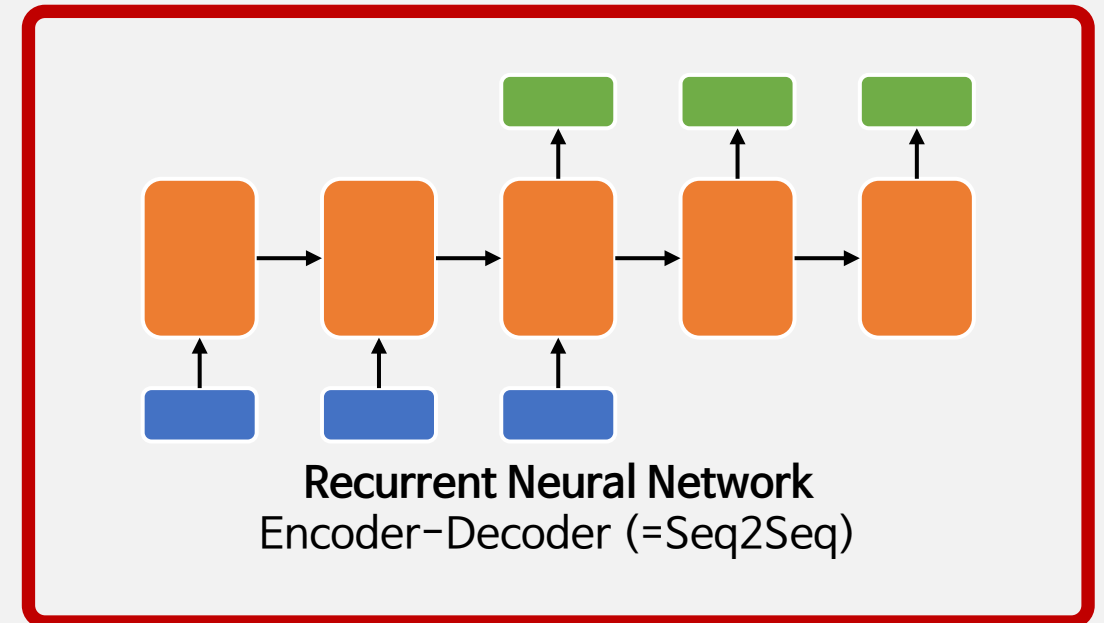
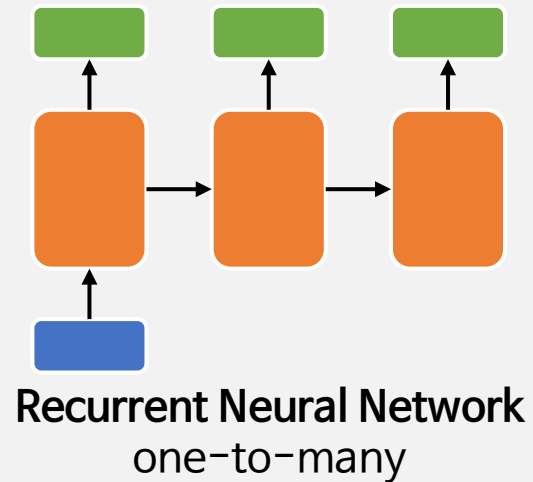
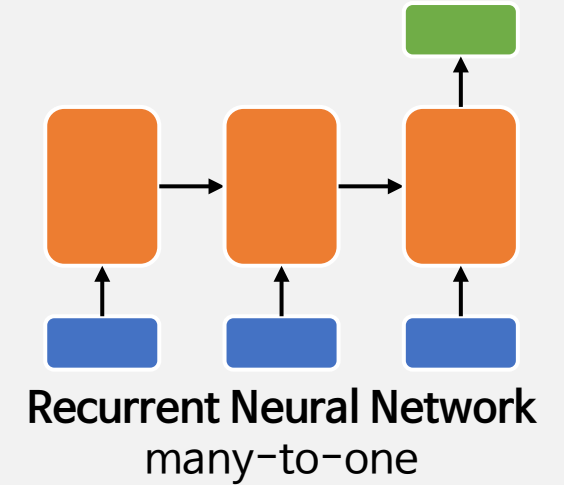
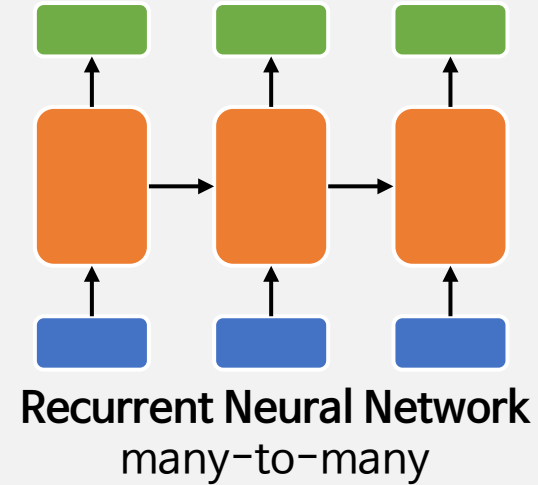
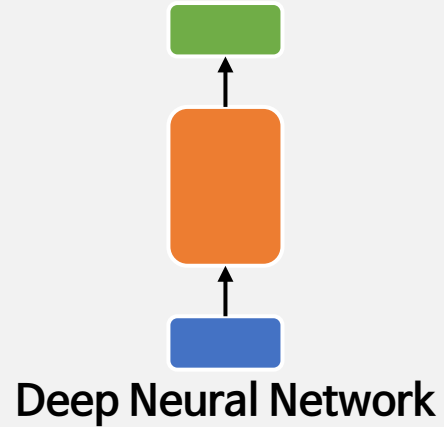
 Introduction

 Seq2Seq

 Application

 Tutorial

 Q&A





# Seq2Seq



Introduction



Seq2Seq



Application



Tutorial



Q&A

- Seq2Seq는 완전히 새로운 모델이 아님
- RNN 두 개(Encoder, Decoder)를 연결시킨 것!
- 아래의 두 논문이 가장 대표적인 Seq2Seq 논문
- **Learning Phrase Representations using RNN Encoder - Decoder for Statistical Machine Translation** (EMNLP 2014), K Cho et al.
- **Sequence to Sequence Learning with Neural Networks** (NIPS 2014), I Sutskever et al.

# Seq2Seq Architecture

Introduction

Seq2Seq

Application

Tutorial

Q&A

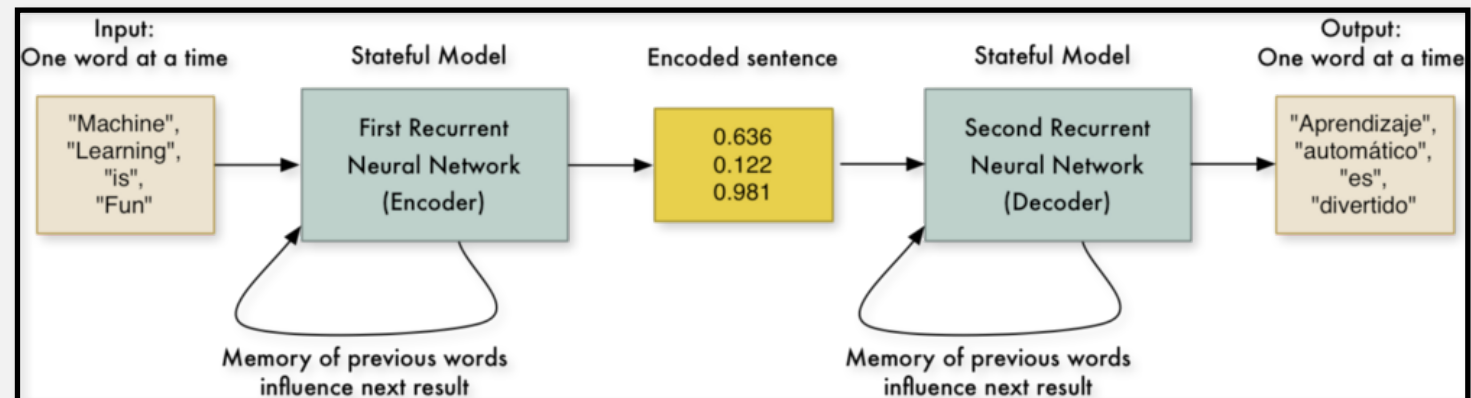
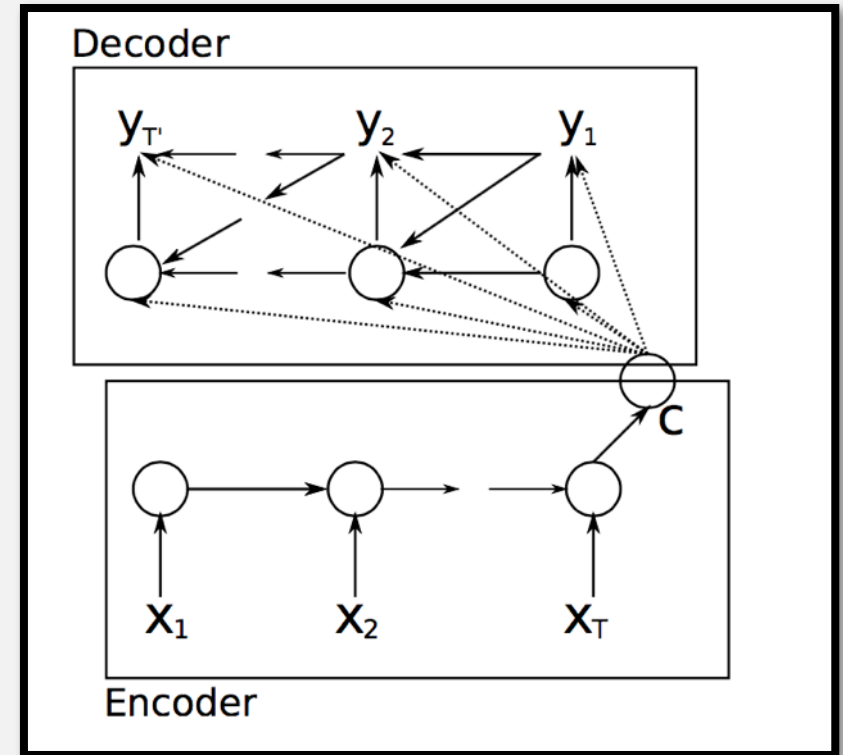
- Encoder는 일반 RNN을 사용

$$\mathbf{h}_{\langle t \rangle} = f(\mathbf{h}_{\langle t-1 \rangle}, x_t),$$

- Decoder는 Encoding된 문장인 C를 같이 봄

$$\mathbf{h}_{\langle t \rangle} = f(\mathbf{h}_{\langle t-1 \rangle}, y_{t-1}, \mathbf{c}),$$

- 각 hidden state를 계산할 때,  
input 문장을 함축적으로 담고 있는 C를 참고하여  
output을 만들어 낸다고 볼 수 있음.





# Seq2Seq Architecture



Introduction



Seq2Seq



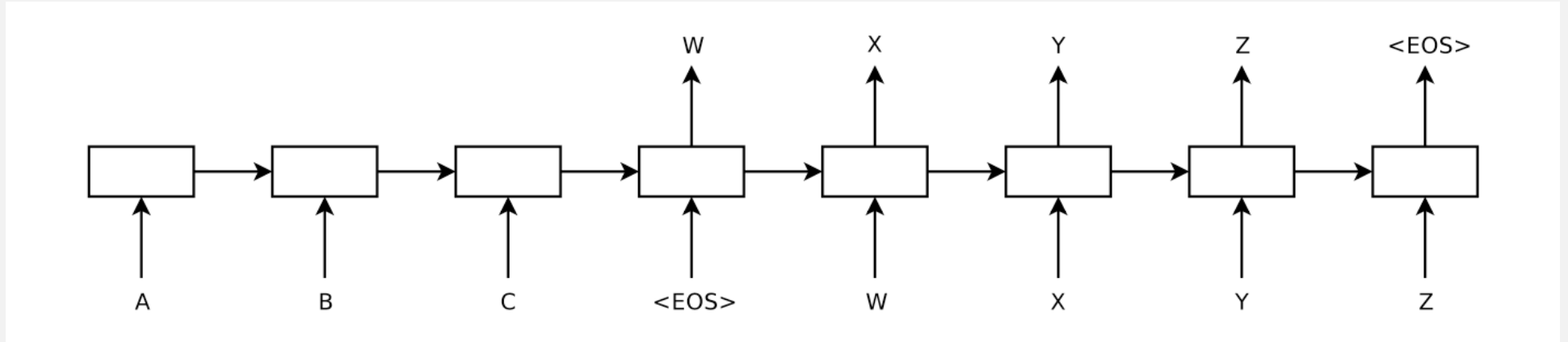
Application



Tutorial



Q&A



- 이 그림과 함께 보면 이해하기 쉬울 듯함

# Seq2Seq Architecture



Introduction



Seq2Seq



Application

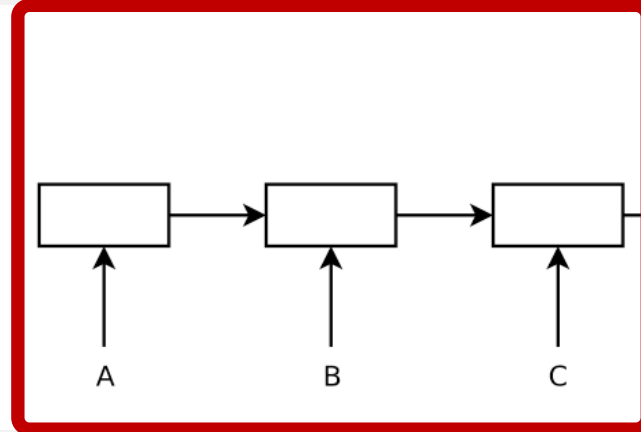


Tutorial

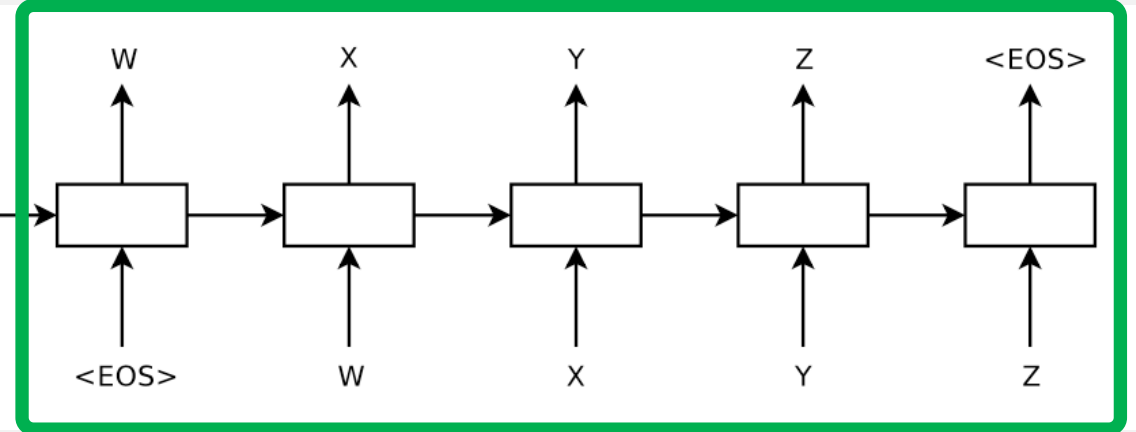


Q&A

Encoder



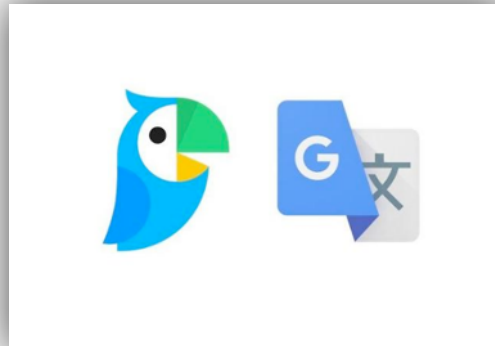
Decoder



- 이 그림과 함께 보면 이해하기 쉬울 듯함
- 아래의 3가지 데이터가 필요
  - Encoder input  $\rightarrow$  input  $x$
  - Decoder input
  - Decoder output  $\rightarrow$  target  $y$
- 위의 그림으로보면,
  - $\langle A, B, C \rangle$ 가 Encoder input
  - $\langle \langle \text{EOS} \rangle, W, X, Y, Z \rangle$ 가 Decoder input
  - $\langle W, X, Y, Z, \langle \text{EOS} \rangle \rangle$ 가 Decoder output
- 실제로는  $\langle A, B, C \rangle$ 와  $\langle W, X, Y, Z \rangle$ , 2개의 input/output sequence pair가 있으면 됨



- 기계 번역에서 Seq2Seq가 상당히 좋음
- 순서를 따져보면,  
Seq2Seq라는 model이 원래 있었는데 이를 번역에 적용시킨 게 아니고,  
**기계번역 문제를 풀기 위해 Seq2Seq를 고안**해낸 것으로 보임
- input (한국어) = 안녕, 나는 주홍이야.
- output (영어) = Hello, I am Joohong.



## 기계 번역(Machine Translation)

**Rule-based MT**  
Dictionary, 문법 기반의 번역 (Parser, Analyzer, Generator, Transfer Lexicon)

**Statistical MT**  
이미 번역된 문서들을 바탕으로한 통계 기반의 번역

**Hybrid MT**  
Rule-based와 Statistical을 함께 사용

**Neural MT**  
Deep Learning을 활용한 번역

© NBT All Rights Reserved. NBT



Introduction



Seq2Seq



Application



Tutorial



Q&A



안녕하세요.

반갑습니다.

.....



대화 시스템, 챗봇(chatter-bot) !



Introduction



Seq2Seq



Application



Tutorial



Q&A

- 준비물: 대화 코퍼스

ex) 안녕.  
안녕하세요.  
너는 이름이 뭐야?  
나는 주홍이야.

- Input = 질문, Output = 대답

- 모든 대화가 질문-대답 형식으로 진행되지 않는다는 점을 고려하면서,  
검사검사 데이터의 크기도 늘이기 위해 아래와 같이 instance 생성 (like sliding window)

x1 = 안녕.

y1 = 안녕하세요.

x2 = 안녕하세요.

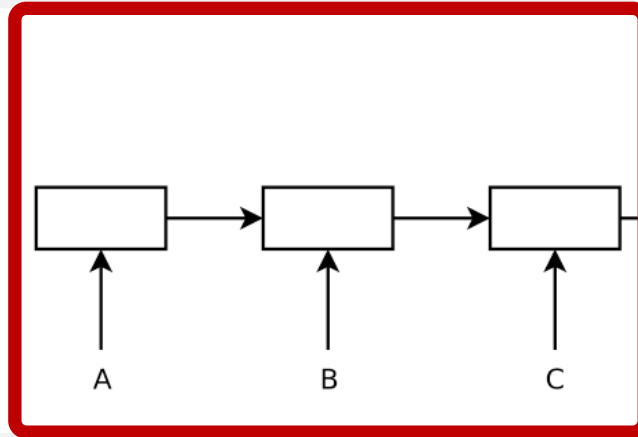
y2 = 너는 이름이 뭐야?

x3 = 너는 이름이 뭐야?

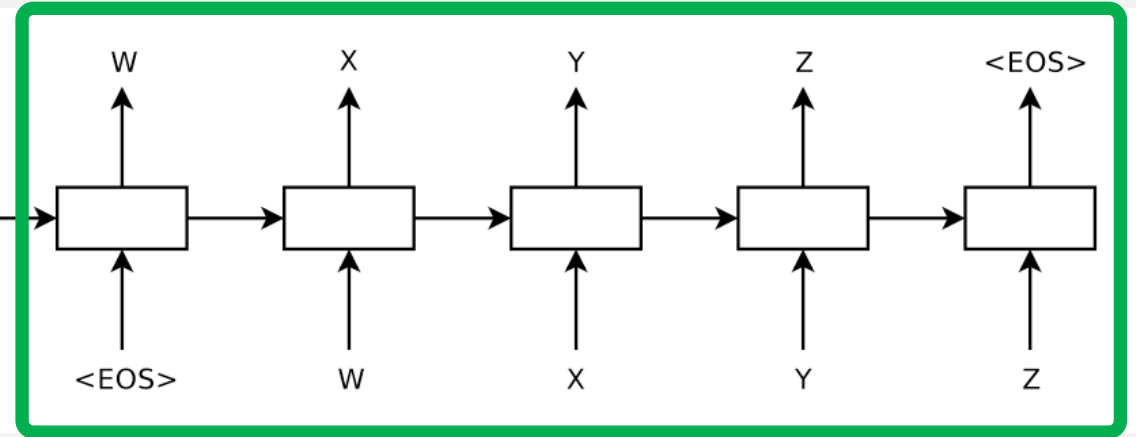
y3 = 나는 주홍이야.

...

Encoder



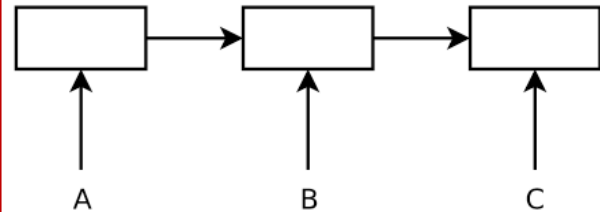
Decoder



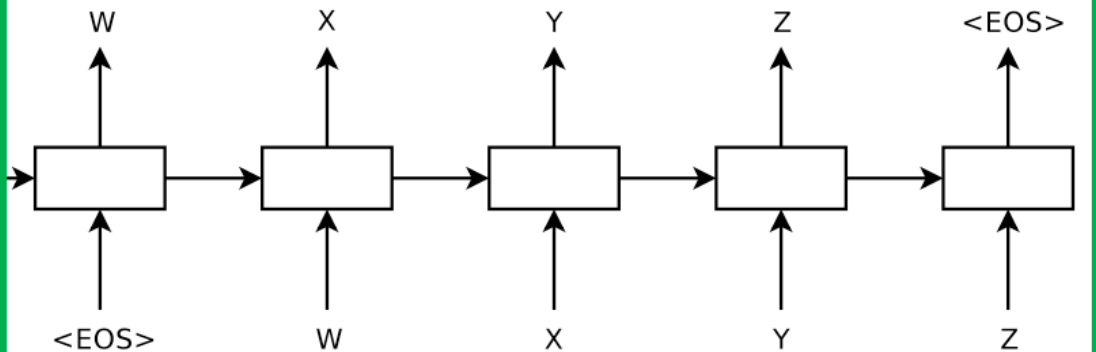
너의	이름은	뭐야	?	<PAD>	<PAD>	<PAD>
나는	주홍	이야	.	<PAD>	<PAD>	<PAD>
오늘	미세	먼지가	정말	많은	걸	?

- Input의 크기가 정해져있기에 (fixed length), 최대 길이에 맞춰 Padding한다.
- Encoder와 Decoder의 length는 달라도 된다.

Encoder



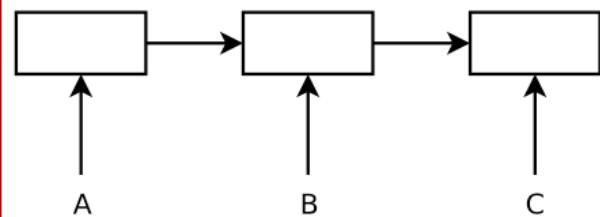
Decoder



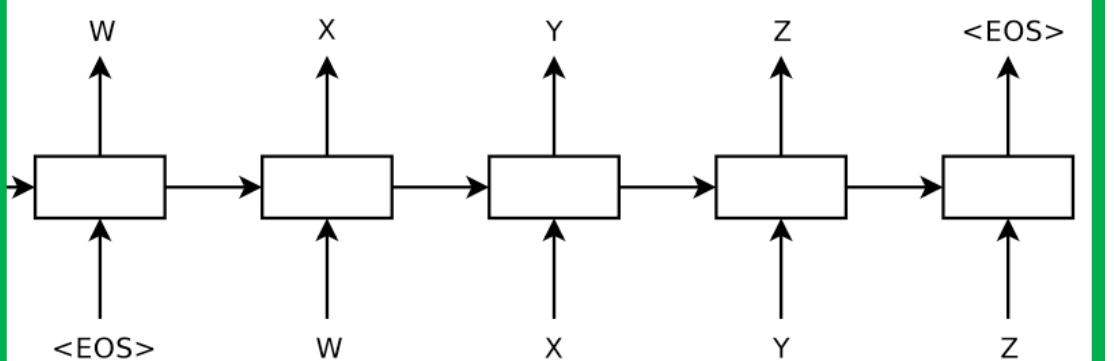
너의	이름은	뭐야	?	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>
나는	주홍	이야	.	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>
오늘	미세	먼지가	정말	~	~	~	~	~	~	했다	.

- 문장의 길이의 차이가 너무 커서 <PAD>가 너무 많아지면 어떻게 하지...?

Encoder



Decoder



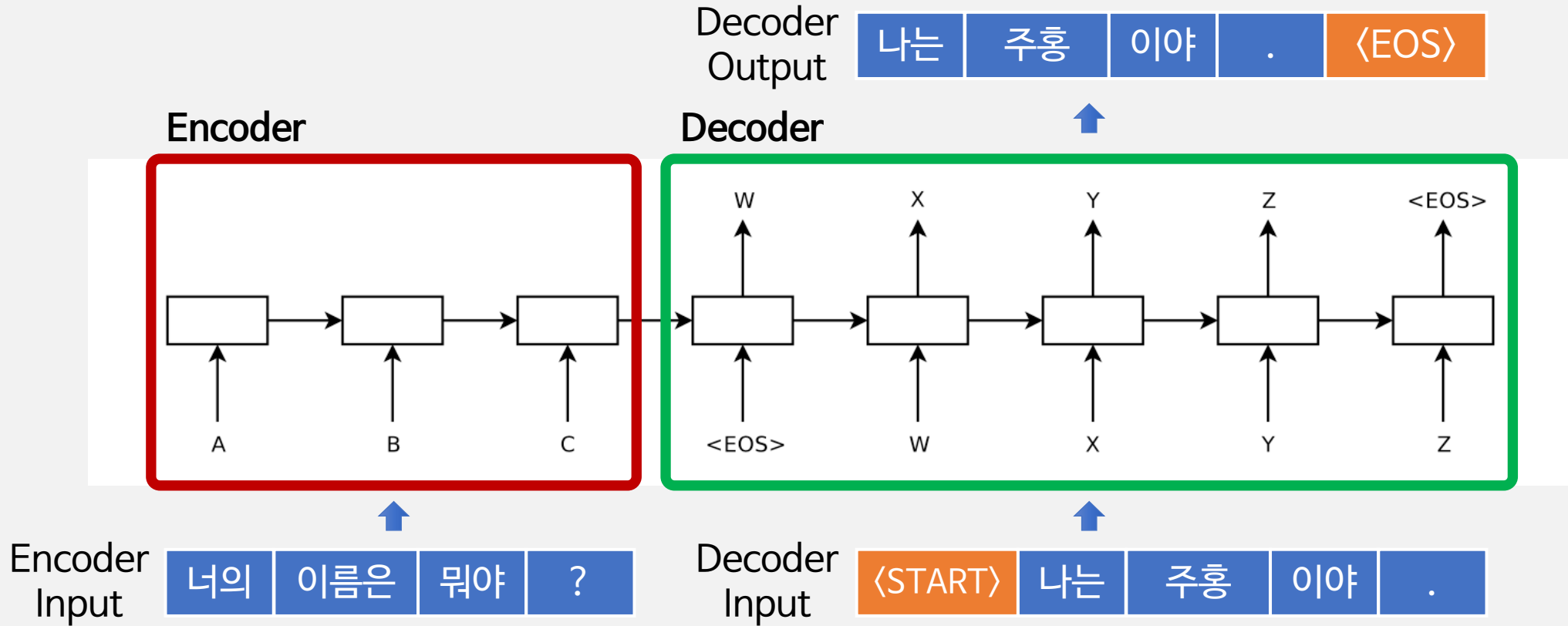
bucket  
≤ 4

너의	이름은	뭐야	?	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>
나는	주홍	이야	.	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>	<PAD>

bucket  
≤ 8

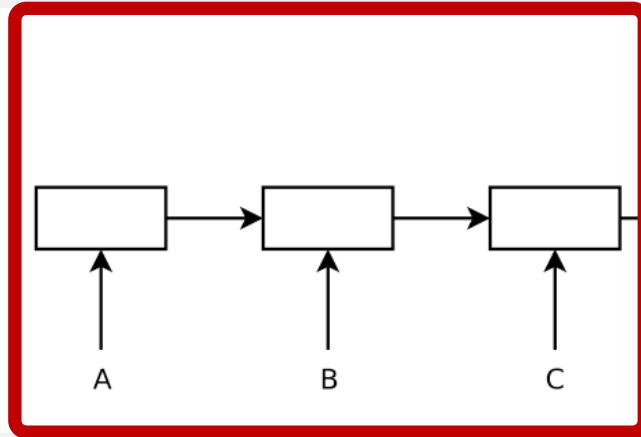
너	정말	나쁜	~	~	아이	구나	.	<PAD>	<PAD>	<PAD>	<PAD>
오늘	미세	먼지가	정말	~	~	~	~	~	~	했다	.

- 길이에 따라서 구간 별 bucket을 만들어서 각각 학습 !



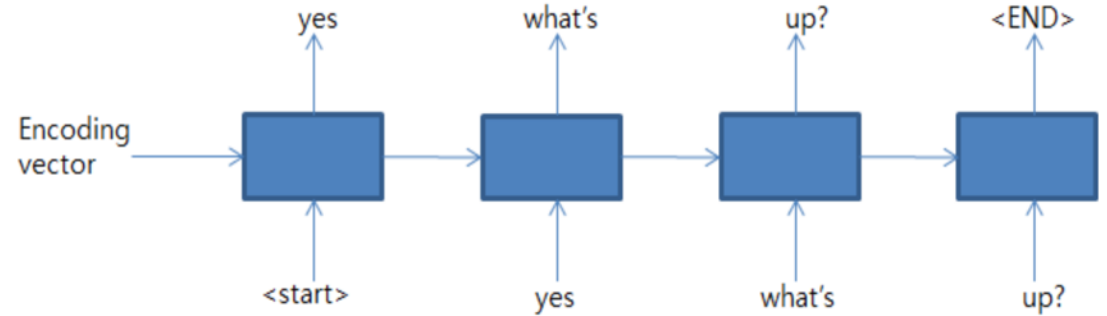
- Decoder의 input / output에는 <START>, <EOS> 등의 토큰으로 시작과 끝을 표시한다.

## Encoder

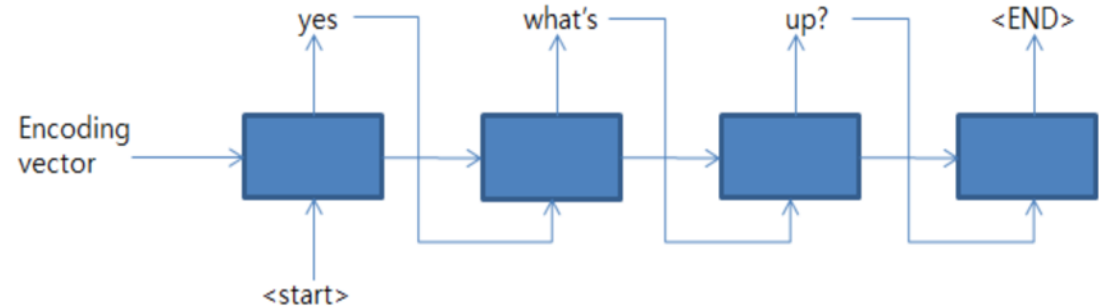


## Decoder

### <TRAIN MODEL>



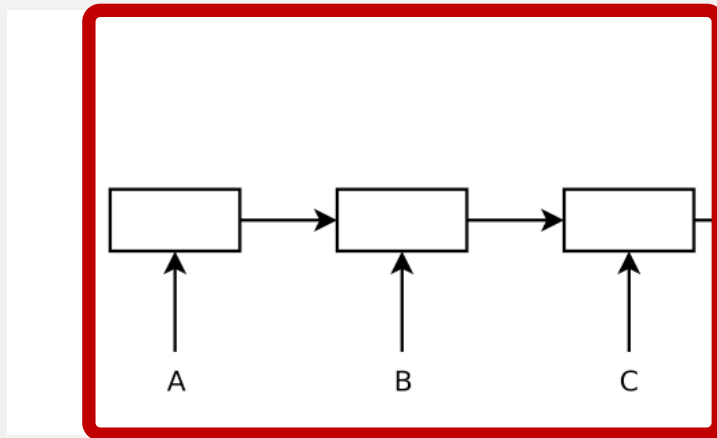
### <TEST MODEL>



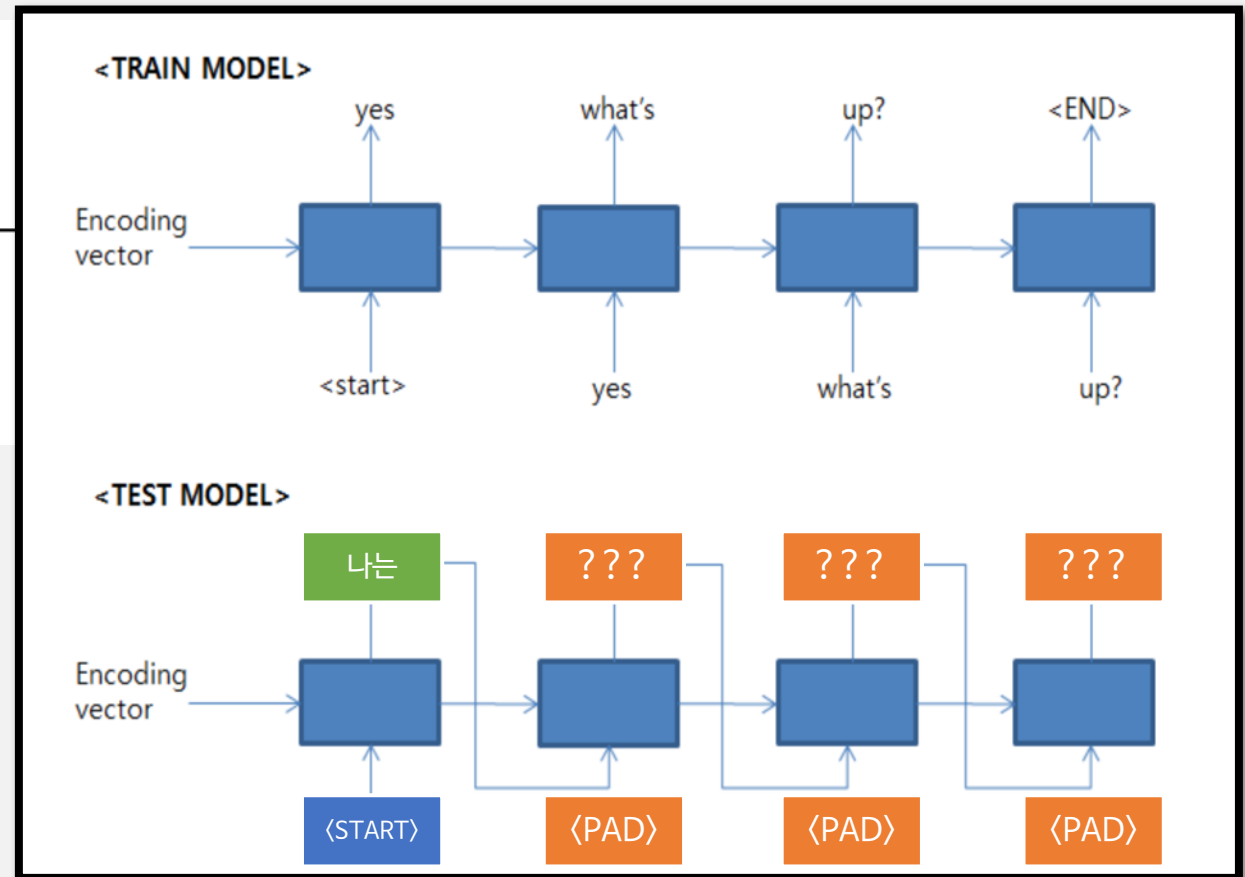
- 어떻게 Output을 바로 Input으로 연결시킬 수 있을까?



## Encoder

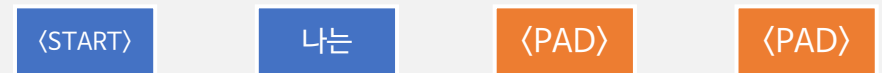


## Decoder

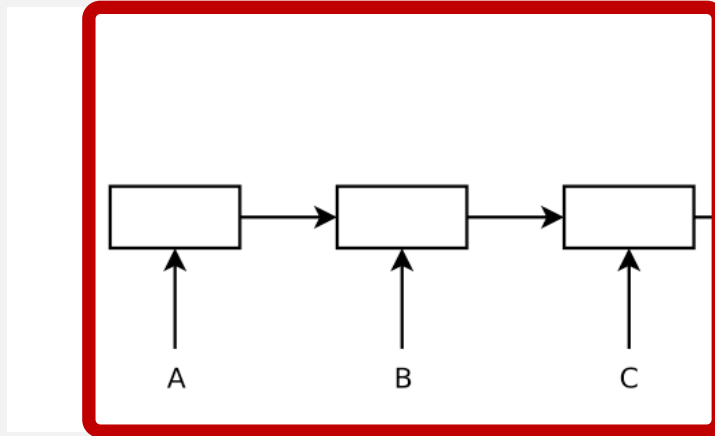


- 〈START〉만 있는 문장을 넣는다.
- 첫번째 output만 가지고와서 〈START〉 뒤에 붙인다.

Next Input Sequence

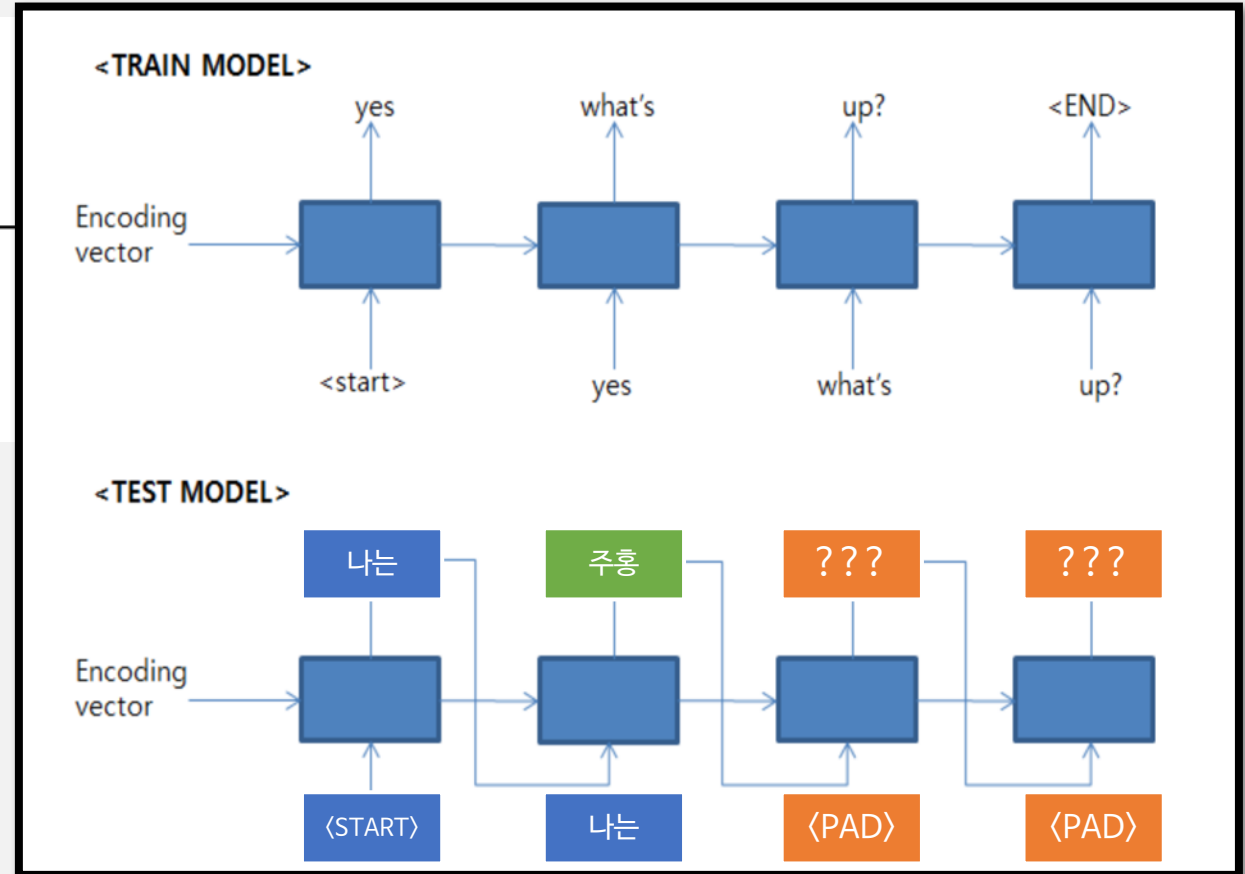


## Encoder

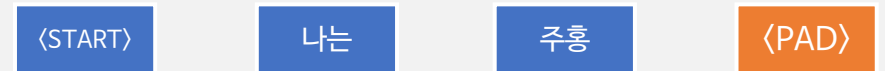


- 첫번째 output을 붙인 문장을 넣고 두번째 output을 취한다.
- 두번째 output을 붙인 문장을 그다음에 넣는다. 반복.

## Decoder



Next Input Sequence



- 학습데이터 증가시킴(100문장)
- Epoch = 100,000번

> 이 새끼가 약을 팔아?  
다들 천하의 아귀가 헛바닥이 이렇게 길어 ? 후달리냐 ?  
> 안녕  
안녕하세요 .

> 너는 누구야?  
다음에 개를 만나요 .  
> 개를 왜 만나?  
내가 그만 . 이 냐 ?  
> 그게 무슨 소리야?  
나는 인공지능 . 나와봐 .  
> 뭐 이자식야?  
필요 , 천만에요 .

> 혹시 장이야?  
확실하지 않으면 승부를 마라 . 안 확인 ? . 둘 까 ? ? ?  
> 후달려? 내 돈 모두를 건다.  
후달려 ? 허허허허허허 . 오냐 , 내 돈 모두하고 내 손모가질 건다  
> 꼭 피를 봐야겠어?  
구라치다 걸리면 피 보는 거 안 배웠냐 ?  
> 젠장 덩벼라  
뭐 이새끼야 !  
> 야! 이 씨발놈 손모가지 찍어라!  
거 중구형 장난이 너무 심한거 아니오

> 안녕하세요.  
안녕 .  
> 반갑습니다.  
안녕하세요 .

> 내가 빙다리 헛바지로 보이냐 ?  
싸늘하다 ?  
> 싸늘하다  
가슴에 날아와 꽂히다 .

> 예림이, 그 패 봐봐, 혹시 장이야?  
패 건들지 마 ! 손모가지 날라가붕게 . 해머 갖고 와 .  
> 이 놈이 어디서 약을 팔어?  
씨발 천하의 아귀가 헛바닥이 왜 이렇게 길어 ? 후달리냐 ?  
> 사쿠라 네?  
사쿠라 야 ?

> 내가 봤어. 이 놈 밀장 빼는 걸 똑똑히 봤다니께!!  
확실하지 않으면 승부를 걸지 마라 이런 거 안 배웠어 ? 뭐해 , 니네 형님 손 안 찍고 ?

> 어이 브라더~  
거기 이거 예사롭지 쥘라 .

> 어이 브라더~  
거기 솜씨가 예사롭지 않습니다 .

> 거기 누구 담배있으면 하나만 쥘라.  
꺼져 . 담배 사람을 소리 거다 !

> 거 죽기 딱 좋은 날씨네.  
살려는 .

> 거 죽기 딱 좋은 날씨네.  
살려는 드릴께

> 너 정말 나쁜 아이구나.  
오빠는 키도 작으면서 .

> 반말 쓰지마!  
키도 작으면서요

> 예쁘면 다냐?  
나는 예쁘면 다야

>

- 학습데이터 증가시킴(200문장)
- Epoch = 200,000번

→ 대부분의 경우에 잘 안되더라...

> 울어라.  
안돼 어디로 너무 많아 .

> 인정?  
준비됐어 ? . 보까 ? . 후달려 . 준비됐어 같이 임마 어 ! 어 어 같이 어 !

> 안녕하세요.  
울어라 .

> 이거 이거 장짜리 아니여?  
안재욱 결혼식 때 왜 안왔어 ?

> 그 놈에 안재욱 안재욱. 지겹지도 않냐?  
오랜만에 술 한 잔 하자 .

> 후... 술 좋지. 지금 나가자.  
이미 늦었어 . 있을 때 잘 해야지 .

> 이런 개새끼가  
안재욱 들으면 때 안돼 ?

> 꺼져 .

> 거기가 어디야?  
! ? ? 왜 해야지 .

> 쿵짜짜  
? 잠겼어 있네 !

> 너 나에게 모욕감을 줬어  
안재욱 어떻게 안돼 안돼

> 너 정말 나쁜 아이구나.  
안재욱 결혼식 안돼 왜 안왔어 ?

> 키도 작으면서.  
그럴까 안돼 직장인이 안돼 올 니네 애길 봤다니께 !

> 키도 작으면서.  
안돼 고향은 안돼 하자 .

> 키도 작으면서요.  
잘 ?

> 그래, 다해라!  
울어라 .

> 울어라  
, 혹시 뻘 ?

> 말 안할 거야?  
안재욱 ? 어디로 가면 그랬어요 ?

> 안재욱 얘기는 왜 자주 해  
안재욱 결혼식 사람이 어떻게 ? !

> 안재욱 결혼식에 꽃했니?  
이미 늦었어 . 있을 때 새끼야 ?

> 너 말을 예쁘게 안하는 구나  
모르는데 어떻게 안돼 !

> 돼! 할 수 있다고!  
왜 결혼식 어떻게 왜 ? ?

> 결혼식 얘기는 왜 자주 하는거야?  
안재욱 서장님과 어떻게 올 . . ?

> 안재욱 결혼식에는 왜 안왔어?  
왜 어떻게 안돼 ?

> 뭐가 안돼?  
그럴까 . 어디로 너무 되지 ?

> 젠장  
, 잠겼어 .

> 밀장빠기냐?  
아부지 ? 잘 꺼 .

> 아부지 얘기 하지마  
준비됐어 ? 그런 왜 유치한 ? 했을 바보가 뭘래 .

> 바보 같은 새끼  
안재욱 ? 가더라도 있을 때 잘 해야지 .

> 맞아. 사람은 있을 때 잘 해야 해  
올 정말 메로나 .

> 메로나 같은 소리 하고 있네.  
지금부터 거기는 사람이 너무 많아 .



Introduction



Seq2Seq



Application

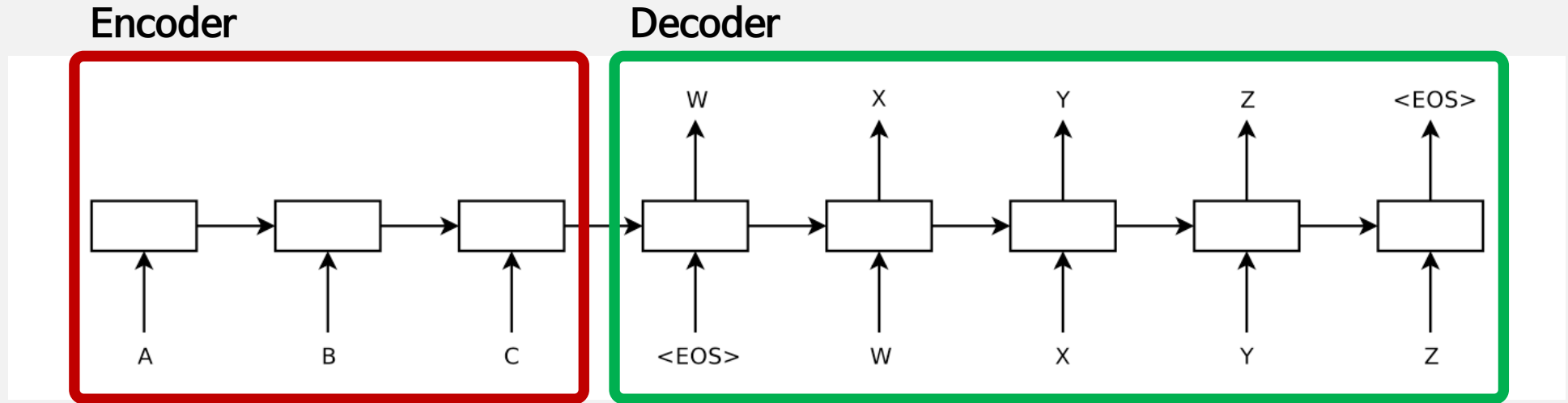


Tutorial



Q&A

- 안타깝지만, Seq2Seq 만으로 챗봇을 만드는 것은 무리임...
  - Attention 등을 적용하고 여러가지 기법을 추가해서 챗봇을 만드려는 노력이 있음
  - 허나, 아직 잘 된 사례는 없음
- 
- 번역과는 다르게 문맥에 대한 의미 파악, 추론 등에 있어서 한계를 보이는 듯함
  - 기존의 rule-based 방법들과 합쳐서 사용하면 보다 나은 결과를 얻을 수 있지 않을까...



Q. 디코딩해서 나온 output vector를 어떻게 symbol(단어)로 바꾸지?

A. output의 차원이 vocabulary 개수만큼 두고, softmax를 붙여서 가장 확률 높은 단어를 선택함.

단어의 개수가 너무 많아서 말이 안되지 않을까? 라는 생각을 했었는데, 실제로 단어의 개수가 너무 많으면 성능이 떨어진다고 함.



# Something-to-Something



Introduction



Seq2Seq



Application



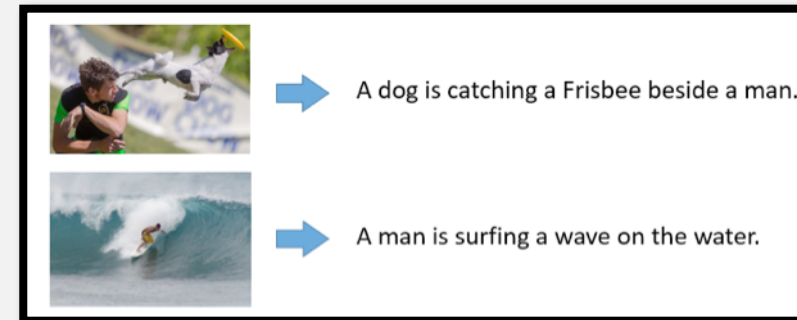
Tutorial



Q&A

- Sth2Sth 모델이 굉장히 많은데 어떤 연관성이 있는지 궁금
- Word2Vec, Sentence2Vec, Paragraph2Vec, Doc2Vec  
→ 모두 Word2Vec와 유사하다고 볼 수 있음. CBOW 형태로 학습

- Image2Text  
→ Encoder를 CNN으로 바꾸면 됨.



- Speech-to-Text / Text-to-Speech





# Thank You

made by. 이 주 흥