# CS224N : Lecture 14 - Neural Language Generation

## Before T5

### Overview

- Transformer to T5
  1. Seq2Seq : `Neural Machine Translation`
  2. Attention+ Seq2Seq : `Neural Machine Translation`
  3. Transformer : `Neural Machine Translation`
  4. GPT-1 : `Task Agnostic : Pretraining + Finetuning`
  5. BERT : `Task Agnostic : Pretraining + Finetuning`
  6. GPT-2 : `Task Agnostic : Zeroshot task transfer`
  7. XLNet : `Task Agnostic`
  8. RoBERTa : `Task Agnostic`
  9. MASS : `Task Agnostic`
  10. BART : `Task Agnostic`
  11. MT-DNN : `Task Agnostic`
  12. T5 : `Task Agnostic`

### Pretraining's Objective

- Auto Encoding
- Auto Regressive

## T5 Model

## T5

> “Unified framework that converts every language problem into a text-to-text format”

## What is text to text?

- Text 형태로 주어진 문제에 대한 text 정답 찾기

## Transfer learning in NLP

| | BERT style model | T5 |
|---|---|---|
| Structure | Encoder-only | Encoder-Decoder structure |
| (특화) Task | Classification & Span Prediction | All NLP task |
| Env | Single prediction per input token & Single prediction for an entire input sequence | 모든 NLP task에서 동일한 model, loss.hyperparameter 사용 가능 (장점) |

## T5 논문’s Points

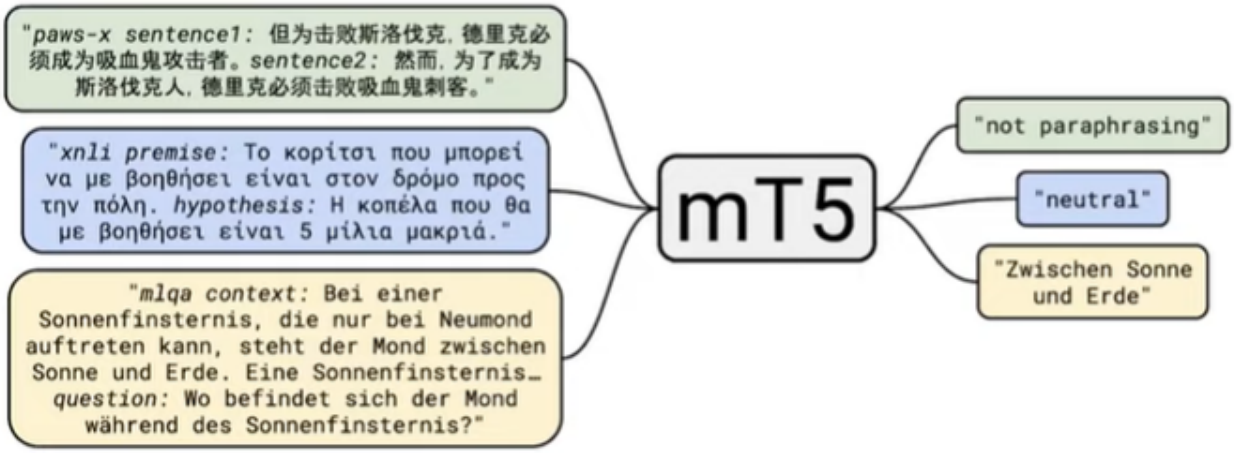
1. Model Architecture
   - 성능 차이 :
     ⇒ `Basic Transformer` > `Encoder, Decoder only`
2. Pretraining Objectives
   - 가장 효율적인 방법 :
     ⇒ Pretraining에서 noising된 input → Denoising하며 단어를 예측
3. Unlabeled datasets
   - Domain specific data → task에 도움
   - But, 데이터의 크기가 작은 경우 → overfitting
4. Training strategies
   - Multi-task learning

    - Unsupervised pre-training과 비슷한 성능
    - 학습 시 task별 적절한 proportion 필요
5. Scaling
    - 작은 모델을 큰 데이터로 학습하는 것이 효과적
6. Pushing the limits
    - 110억 개 파라미터 모델 훈련 → SOTA
    - 1 trilon개가 넘는 token에 대해 훈련 진행

# Other topics

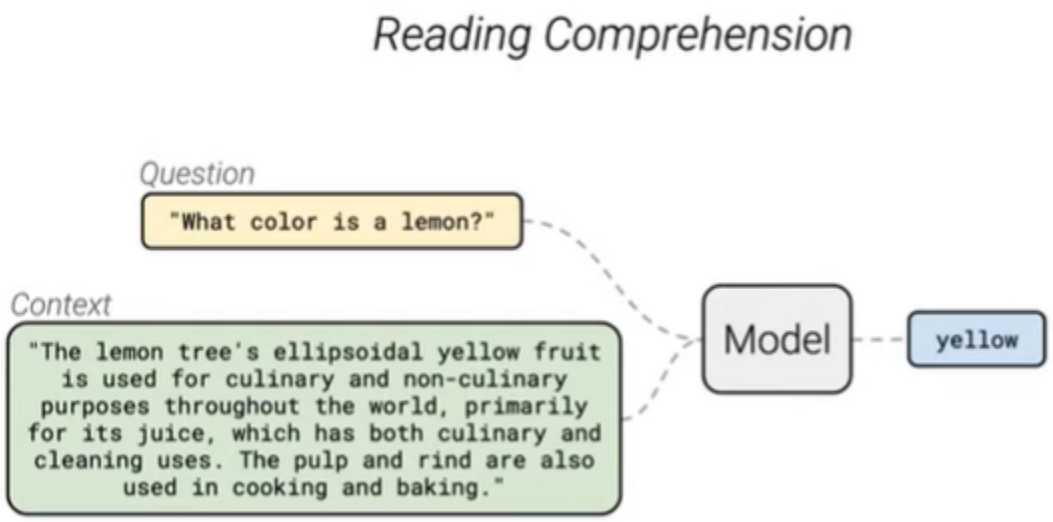
## Multilingual T5

- T5 model trained on multilingual corpus



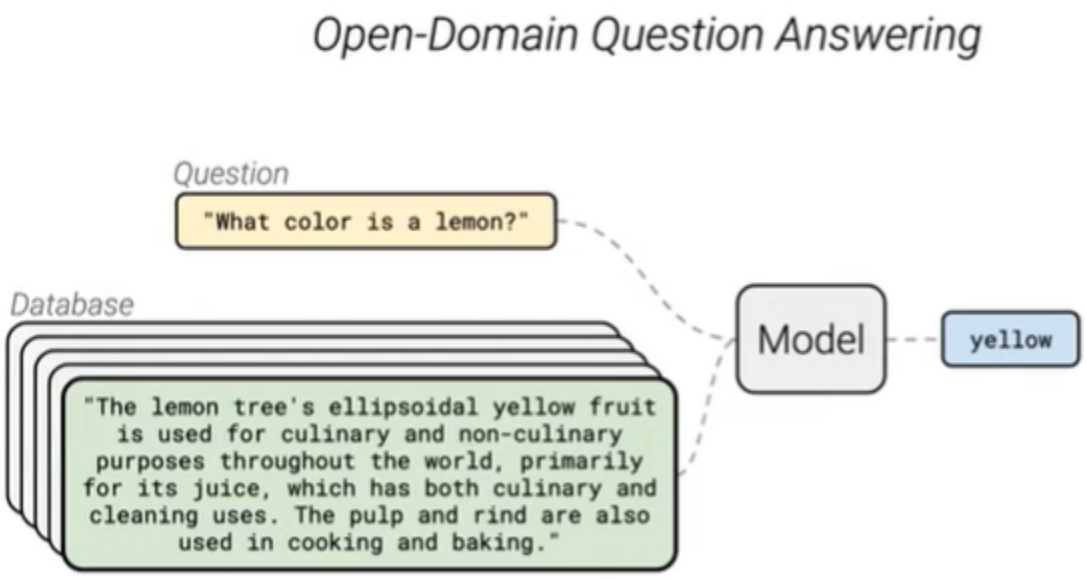

## Reading comprehension

- Given a paragraph or an article
- Asked a question about the paragraph or article, it has to extract the answer




## Open-Domain Question Answering

- Given a question
- Access to a large external DB of knowledge
  - Find the article in the DB → Extract the answer




## Closed-Book Question Answering

- Given the question

- Does not have access to an external knowledge source
  - Only uses the knowledge picked up during pre-training