

# Introduction to Language Modeling

Ki Hyun Kim

[nlp.with.deep.learning@gmail.com](mailto:nlp.with.deep.learning@gmail.com)

# Introduction

- 언어모델(Language Model, LM)은 문장의 확률을 나타낸 모델
  - 문장 자체의 출현 확률을 예측 하거나,
  - 이전 단어들이 주어졌을 때 다음 단어를 예측하기 위한 모델
- 버스 정류장에서 방금 버스를 ○○○.
  - ① 사랑해
  - ② 고양이
  - ③ 놓쳤다
  - ④ 사고남

# Introduction

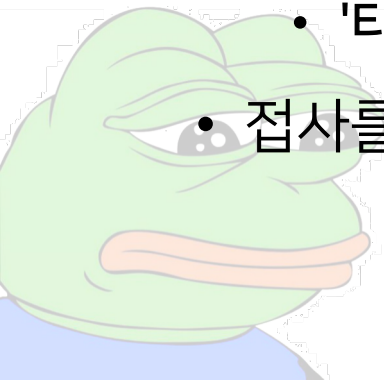
- 우리의 머릿속에는 단어와 단어 사이의 확률이 우리도 모르게 학습되어 있음
  - 대화를 하다가 정확하게 듣지 못하여도 대화에 지장이 없음
- 많은 문장들을 수집하여, 단어와 단어 사이의 출현 빈도를 세어 확률을 계산
- 궁극적인 목표는 우리가 일상 생활에서 사용하는 언어의 문장 분포를 정확하게 모델링 하는 것
  - 특정 분야(domain)의 문장의 분포를 파악하기 위해서 해당 분야의 말뭉치를 수집하기도

# Again, Korean is Hell

- 단어와 단어 사이의 확률을 계산하는데 **불리**하게 작용
  - ① 단어의 어순이 중요하지 않기 때문에
  - ② 또는 생략 가능하기 때문에
- Example
  - ① 나는 학교에 갑니다 버스를 타고 .
  - ② 나는 버스를 타고 학교에 갑니다 .
  - ③ 버스를 타고 나는 학교에 갑니다 .
  - ④ (나는) 버스를 타고 학교에 갑니다 .

- **확률이 퍼지는 현상**

- '타고' 다음에 나타날 수 있는 단어들은 '.', '학교에', '나는' 3개이기 때문
  - 접사를 따로 분리해주지 않으면 어휘의 수가 기하급수적으로 늘어나 희소성이 더욱 높아짐



# Applications

- Natural Language Generation

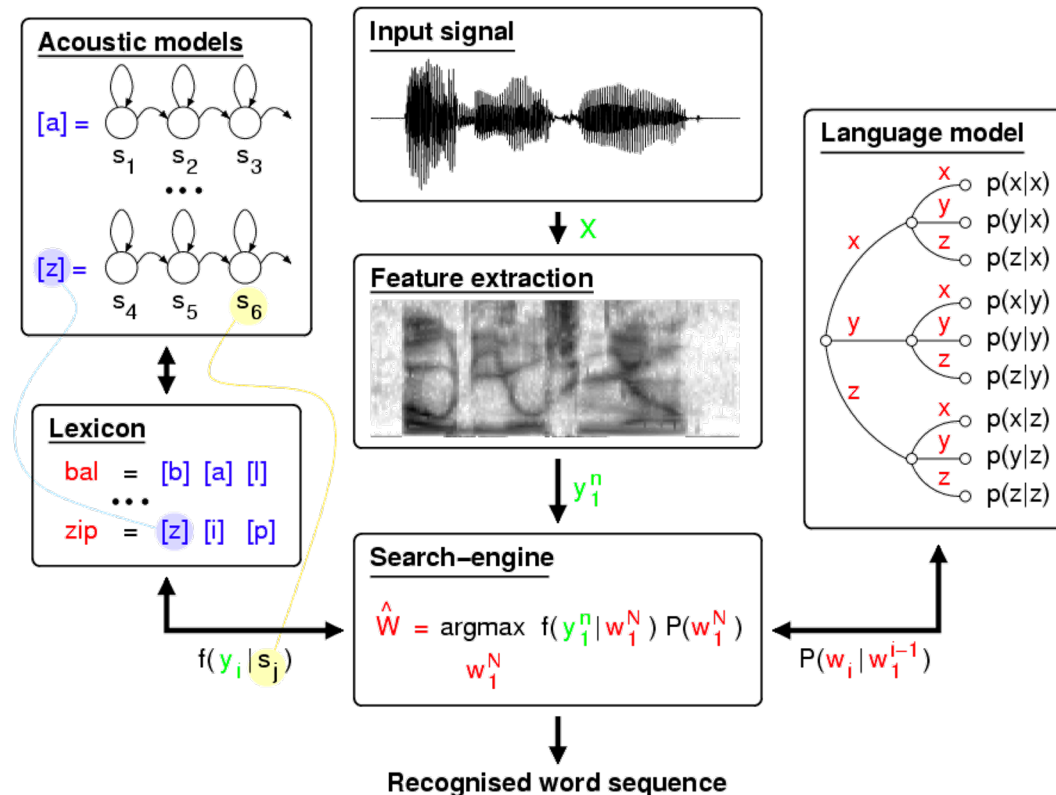
| Task                          | Description   |
|-------------------------------|---|
| Speech Recognition            | Acoustic Model과 결합하여, 인식된 phone(음소)의 sequence에 대해서 좀 더 높은 확률을 갖는 sequence로 보완 |
| Machine Translation           | 번역 모델과 결합하여, 번역 된 결과 문장을 자연스럽게 만듦   |
| Optical Character Recognition | 인식된 character candidate sequence에 대해서 좀 더 높은 확률을 갖는 sequence를 선택하도록 도움        |
| Other NLG Tasks               | 뉴스 기사 생성, chat-bot 등  |
| Other...                      | 검색어 자동 완성 등...  |

# Applications

- Automatic Speech Recognition (ASR)
- Objective

$$y_{1:n} = \operatorname{argmax} P(y|x; \theta)$$

where  $x$  is audio signal, and  $y_{1:n}$  is word sequence,  $y_{1:n} = \{y_1, y_2, \dots, y_n\}$ .



# Automatic Speech Recognition (ASR)

$$y_{1:n} = \operatorname{argmax} P(y|x; \theta)$$

where  $x$  is audio signal, and  $y_{1:n}$  is word sequence,  $y_{1:n} = \{y_1, y_2, \dots, y_n\}$ .

- By Bayes Theorem,

$$\begin{aligned}\hat{y} &= \operatorname{argmax}_{y \in \mathcal{Y}} P(y|x) \\ &= \operatorname{argmax}_{y \in \mathcal{Y}} \frac{P(x|y)P(y)}{P(x)} \\ &= \operatorname{argmax}_{y \in \mathcal{Y}} \underbrace{P(x|y)}_{\text{AM}} \underbrace{P(y)}_{\text{LM}}\end{aligned}$$