

# 단어 시퀀스에서의 확률 할당

## A. 단어 시퀀스의 확률

하나의 단어를  $w$ , 단어 시퀀스를 대문자  $W$ 라고 한다면,  $n$ 개의 단어가 등장하는 단어 시퀀스  $W$ 의 확률은 다음과 같습니다.

$$P(W) = P(w_1, w_2, w_3, w_4, w_5, \dots, w_n)$$

## B. 다음 단어 등장 확률

이제 다음 단어 등장 확률을 식으로 표현해보겠습니다.  $n-1$ 개의 단어가 나열된 상태에서  $n$ 번째 단어의 확률은 다음과 같습니다.

$$P(w_n | w_1, \dots, w_{n-1})$$

|의 기호는 조건부 확률(conditional probability)을 의미합니다.

# 통계적 언어모델(SLM)

- 딥러닝의 발전 이전에도 있었던 개념, 전통적 언어 모델
- 문장의 확률을 구하기 위해 **조건부 확률** 사용
- 이때의 확률값은 **카운트에 기반해 계산**

$$P(w_1, w_2, w_3, w_4, w_5, \dots w_n) = \prod_{n=1}^n P(w_n | w_1, \dots, w_{n-1})$$