

# Unigram, Bigram & Trigram

- > Unigram -> One word.
- > Bigram -> Two words
- > Trigram -> Three words
- > <s> I am here </s>
- > <s> Who am I </s>
- > <s> I would like to know </s>
- > Estimating Bigrams:

$$P(I | <s>) = \frac{2}{3} = 66.67\%$$

$$P(</s> | here) = 1 = 100\%$$

$$P(would | I) = \frac{1}{3} = 33.33\%$$

$$P(here | am) = \frac{1}{2} = 50\%$$

$$P(know | like) = \frac{0}{1} = 0\%$$

-> Practical Issue:

Everything in  $b_2$  space

Everything in log space

$$\log (P_1 \times P_2 \times P_3 \times P_4) = \log P_1 + \log P_2 + \log P_3 + \log P_4$$

$$\Pi \approx \Sigma$$

- > Avoids underflow.
- > MORPHOLOGY:
  - dogs -> 2 morphemes (dog + s)
- > Variants of same morpheme  
eg: un-happy
- > Bound & Free Morpheme:
  - ↳ Bound: cannot appear as a word itself.
    - s (eat-s)
    - ly (quick-ly)
    - ed (walk-ed)

↳ Free: Independent

house (house-s)

→ Content & Functional Morpheme:

→ Content: Semantic Meaning

eg: car, -able, -un

→ Functional: Provides Grammatical Info

eg: -s (plural)

walk-ing      walking

drive-er      driver

Verb

Noun

~ Inflectional  
Morphology

↳ Derivational  
Morphology

→ Inflectional: Numbers, Tense, Case, Gender

-> Inflectional : Numbers, Tense, Case, Gender  
eg: bring, brought, brings, bringing.

-> Derivational : Change Part-of-speech.  
eg: logic, logical, illogical

-> Morphological Process:

-> Concatenation : Combining affixes.

eg: un+happy, anti+capital+ist+s

-> Phonological / Grapheme Changes:

-> books +s [s]

-> shoe +s [z]

-> happy +er -> happier (happy er)  
happy

happi

-> Suppletion: Irregularities

go - went , good - better

-> Morpheme Internal Changes:

sing - sang - sung , man - men , goose - geese

-> Word Formation:

↳ Compounding:

-> Adj + Adj -> Adj : Bitter-Sweet

-> N + N -> N : Rain-bow

-> V + N -> V : pick-pocket

-> P + V -> V : over-do