

# *Computational Morphology*

Pawan Goyal

CSE, IITKGP

Week 3: Lecture 2

# *Morphology*

Morphology studies the internal structure of words, how words are built up from smaller meaningful units called **morphemes**

Morphology studies the internal structure of words, how words are built up from smaller meaningful units called **morphemes**

*dogs*

- 2 morphemes, 'dog' and 's'
- 's' is a plural marker on nouns

Morphology studies the internal structure of words, how words are built up from smaller meaningful units called **morphemes**

*dogs*

- 2 morphemes, 'dog' and 's'
- 's' is a plural marker on nouns

*unladylike*

3 morphemes

- un- 'not'
- lady 'well-behaved woman'
- -like 'having the characteristic of'

Variants of the same morpheme, but cannot be replaced by one another

## *Example*

- opposite: un-happy, in-comprehensible, im-possible, ir-rational

# Bound and Free Morphemes

## *Bound*

Cannot appear as a word by itself.

-s (*dog-s*), -ly (*quick-ly*), -ed (*walk-ed*)

# Bound and Free Morphemes

## Bound

Cannot appear as a word by itself.

*-s (dog-s), -ly (quick-ly), -ed (walk-ed)*

## Free

Can appear as a word by itself; often can combine with other morphemes too.

*house (house-s), walk (walk-ed), of, the, or*

## Stems and Affixes

- Stems (roots): The core meaning bearing units
- Affixes: Bits and pieces adhering to stems to change their meanings and grammatical functions



# Stems and Affixes

## Stems and Affixes

- Stems (roots): The core meaning bearing units
- Affixes: Bits and pieces adhering to stems to change their meanings and grammatical functions

Mostly, stems are free morphemes and affixes are bound morphemes

# *Types of affixes*

# Types of affixes

- Prefix: un-, anti-, etc (a-, ati-, pra- etc.)  
*un-happy, pre-existing*

# Types of affixes

- Prefix: un-, anti-, etc (a-, ati-, pra- etc.)  
*un-happy, pre-existing*
- Suffix: -ity, -ation, etc (-taa, -ke, -ka etc.)  
*talk-ing, quick-ly*

# Types of affixes

- Prefix: un-, anti-, etc (a-, ati-, pra- etc.)  
*un-happy, pre-existing*
- Suffix: -ity, -ation, etc (-taa, -ke, -ka etc.)  
*talk-ing, quick-ly*
- Infix: 'n' in 'vindatī' (he knows), as contrasted with vid (to know).  
*Philippines: basa 'read' → b-um-asa 'read'*  
*English:*

# Types of affixes

- Prefix: un-, anti-, etc (a-, ati-, pra- etc.)  
*un-happy, pre-existing*
- Suffix: -ity, -ation, etc (-taa, -ke, -ka etc.)  
*talk-ing, quick-ly*
- Infix: 'n' in 'vindatī' (he knows), as contrasted with vid (to know).  
*Philippines: basa 'read' → b-um-asa 'read'*  
*English: abso-bloody-lutely (emphasis)*

# Types of affixes

- Prefix: un-, anti-, etc (a-, ati-, pra- etc.)  
*un-happy, pre-existing*
- Suffix: -ity, -ation, etc (-taa, -ke, -ka etc.)  
*talk-ing, quick-ly*
- Infix: 'n' in 'vindati' (he knows), as contrasted with vid (to know).  
*Philippines: basa 'read' → b-um-asa 'read'*  
*English: abso-bloody-lutely (emphasis)*
- Circumfixes - precedes and follows the stem  
*Dutch: berg 'mountain', ge-berg-te 'mountains'*

# *Content and functional morphemes*

## *Content morphemes*

Carry some semantic content

*car, -able, un-*



# Content and functional morphemes

## *Content morphemes*

Carry some semantic content

*car, -able, un-*

## *Functional morphemes*

Provide grammatical information

*-s (plural), -s (3<sup>rd</sup> singular)*

# *Inflectional and Derivational Morphology*

Two different kind of relationship among words

# *Inflectional and Derivational Morphology*

Two different kind of relationship among words

## *Inflectional morphology*

Grammatical: number, tense, case, gender

Creates new forms of the same word: *bring, brought, brings, bringing*

# *Inflectional and Derivational Morphology*

Two different kind of relationship among words

## *Inflectional morphology*

Grammatical: number, tense, case, gender

Creates new forms of the same word: *bring, brought, brings, bringing*

## *Derivational morphology*

Creates new words by changing part-of-speech: *logic, logical, illogical, illogicality, logician*

# Inflectional and Derivational Morphology

Two different kind of relationship among words

## *Inflectional morphology*

Grammatical: number, tense, case, gender

Creates new forms of the same word: *bring, brought, brings, bringing*

## *Derivational morphology*

Creates new words by changing part-of-speech: *logic, logical, illogical, illogicality, logician*

Fairly systematic but some derivations missing: *sincere - sincerity, scarce - scarcity, curious - curiosity, fierce - fierceness?*

# Morphological processes

## Concatenation

Adding continuous affixes - the most common process:

- hope+less, un+happy, anti+capital+ist+s

# Morphological processes

## Concatenation

Adding continuous affixes - the most common process:

- hope+less, un+happy, anti+capital+ist+s

Often, there are phonological/graphemic changes on morpheme boundaries:

- book + s [s], shoe + s [z]
- happy +er → happier

**Reduplication:** part of the word or the entire word is doubled



**Reduplication:** part of the word or the entire word is doubled

- Nama: 'go' (look), 'go-go' (examine with attention)

**Reduplication:** part of the word or the entire word is doubled

- Nama: 'go' (look), 'go-go' (examine with attention)
- Tagalog: 'basa' (read), 'ba-basa'(will read)

**Reduplication:** part of the word or the entire word is doubled

- Nama: 'go' (look), 'go-go' (examine with attention)
- Tagalog: 'basa' (read), 'ba-basa'(will read)
- Sanskrit: 'pac' (cook), 'papāca' (perfect form, cooked)

**Reduplication:** part of the word or the entire word is doubled

- Nama: 'go' (look), 'go-go' (examine with attention)
- Tagalog: 'basa' (read), 'ba-basa' (will read)
- Sanskrit: 'pac' (cook), 'papāca' (perfect form, cooked)
- Phrasal reduplication (Telugu): *pillavāḍu naḍustū naḍustū paḍi pōyāḍu*  
(The child fell down while walking)

## *Suppletion*

'irregular' relation between the words

*go - went, good - better*

# Morphological processes

## *Suppletion*

'irregular' relation between the words

*go - went, good - better*

## *Morpheme internal changes*

The word changes internally

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

## *Compounding*

Words formed by combining two or more words

Example in English:

- Adj + Adj  $\rightarrow$  Adj: bitter-sweet
- N + N  $\rightarrow$  N: rain-bow
- V + N  $\rightarrow$  V: pick-pocket
- P + V  $\rightarrow$  V: over-do

## *Compounding*

Words formed by combining two or more words

Example in English:

- Adj + Adj  $\rightarrow$  Adj: bitter-sweet
- N + N  $\rightarrow$  N: rain-bow
- V + N  $\rightarrow$  V: pick-pocket
- P + V  $\rightarrow$  V: over-do

## *Particular to languages*

*room-temperature*: Hindi translation?



## *Compounding*

Words formed by combining two or more words

Example in English:

- Adj + Adj  $\rightarrow$  Adj: bitter-sweet
- N + N  $\rightarrow$  N: rain-bow
- V + N  $\rightarrow$  V: pick-pocket
- P + V  $\rightarrow$  V: over-do

## *Particular to languages*

*room-temperature*: Hindi translation?

## Acronyms

*laser*: Light Amplification by Stimulated Emission of Radiation

## Acronyms

*laser*: Light Amplification by Simulated Emission of Radiation

## Blending

Parts of two different words are combined

- breakfast + lunch → brunch
- smoke + fog → smog
- motor + hotel → motel

## *Acronyms*

*laser*: Light Amplification by Simulated Emission of Radiation

## *Blending*

Parts of two different words are combined

- breakfast + lunch → brunch
- smoke + fog → smog
- motor + hotel → motel

## *Clipping*

Longer words are shortened

# Word Formation

## Acronyms

*laser*: Light Amplification by Simulated Emission of Radiation

## Blending

Parts of two different words are combined

- breakfast + lunch → brunch
- smoke + fog → smog
- motor + hotel → motel

## Clipping

Longer words are shortened

*doctor, laboratory, advertisement, dormitory, examination, bicycle, refrigerator*

- Lemmatization: word  $\rightarrow$  lemma  
saw  $\rightarrow$  {see, saw}

# Processing morphology

- Lemmatization: word  $\rightarrow$  lemma  
saw  $\rightarrow$  {see, saw}
- Morphological analysis : word  $\rightarrow$  setOf(lemma +tag)  
saw  $\rightarrow$  { <see, verb.past>, < saw, noun.sg> }

# Processing morphology

- Lemmatization: word  $\rightarrow$  lemma  
saw  $\rightarrow$  {see, saw}
- Morphological analysis : word  $\rightarrow$  setOf(lemma +tag)  
saw  $\rightarrow$  { <see, verb.past>, < saw, noun.sg>}
- Tagging: word  $\rightarrow$  tag, considers context  
Peter saw her  $\rightarrow$  { <see, verb.past>}



- Lemmatization: word  $\rightarrow$  lemma  
saw  $\rightarrow$  {see, saw}
- Morphological analysis : word  $\rightarrow$  setOf(lemma +tag)  
saw  $\rightarrow$  { <see, verb.past>, < saw, noun.sg>}
- Tagging: word  $\rightarrow$  tag, considers context  
Peter saw her  $\rightarrow$  { <see, verb.past>}
- Morpheme segmentation: de-nation-al-iz-ation

- Lemmatization: word  $\rightarrow$  lemma  
saw  $\rightarrow$  {see, saw}
- Morphological analysis : word  $\rightarrow$  setOf(lemma +tag)  
saw  $\rightarrow$  { <see, verb.past>, < saw, noun.sg>}
- Tagging: word  $\rightarrow$  tag, considers context  
Peter saw her  $\rightarrow$  { <see, verb.past>}
- Morpheme segmentation: de-nation-al-iz-ation
- Generation: see + verb.past  $\rightarrow$  saw

# *What are the applications?*

- Text-to-speech synthesis:  
*lead:*

# *What are the applications?*

- Text-to-speech synthesis:  
*lead*: verb or noun?  
*read*:

# What are the applications?

- Text-to-speech synthesis:  
*lead*: verb or noun?  
*read*: present or past?
- Search and information retrieval
- Machine translation, grammar correction

# Morphological Analysis

Input	Morphological Parsed Output
cats	cat +N +PL
cat	cat +N +SG
cities	city +N +PL
geese	goose +N +PL
goose	(goose +N +SG) or (goose +V)
gooses	goose +V +3SG
merging	merge +V +PRES-PART
caught	(catch +V +PAST-PART) or (catch +V +PAST)

# Morphological Analysis

Input	Morphological Parsed Output
cats	cat +N +PL
cat	cat +N +SG
cities	city +N +PL
geese	goose +N +PL
goose	(goose +N +SG) or (goose +V)
gooses	goose +V +3SG
merging	merge +V +PRES-PART
caught	(catch +V +PAST-PART) or (catch +V +PAST)

## Goal

To take input forms like those in the first column and produce output forms like those in the second column.

Output contains stem and additional information; +N for noun, +SG for singular, +PL for plural, +V for verb etc.

# *Issues involved*

boy → boys



# *Issues involved*

boy  $\rightarrow$  boys

fly  $\rightarrow$  flys  $\rightarrow$  flies (y  $\rightarrow$  i rule)

# *Issues involved*

boy → boys

fly → flys → flies (y → i rule)

Toiling → toil

# Issues involved

boy → boys

fly → flys → flies (y → i rule)

Toiling → toil

Duckling → duckl?

# Issues involved

boy → boys

fly → flys → flies (y → i rule)

Toiling → toil

Duckling → duckl?

- Getter → get + er
- Doer → do + er

# Issues involved

boy → boys

fly → flys → flies (y → i rule)

Toiling → toil

Duckling → duckl?

- Getter → get + er
- Doer → do + er
- Beer → be + er?

# Knowledge Required

## *Knowledge of stems or roots*

*Duck* is a possible root, not *duckl*.

*We need a dictionary (lexicon)*

# Knowledge Required

## *Knowledge of stems or roots*

*Duck* is a possible root, not *duckl*.

*We need a dictionary (lexicon)*

## *Morphotactics*

Which class of morphemes follow other classes of morphemes inside the word?

*Ex: plural morpheme follows the noun*

# Knowledge Required

## *Knowledge of stems or roots*

*Duck* is a possible root, not *duckl*.

*We need a dictionary (lexicon)*

## *Morphotactics*

Which class of morphemes follow other classes of morphemes inside the word?

*Ex: plural morpheme follows the noun*

## *Only some endings go on some words*

- *Do+er*: ok
- *Be+er*: not so



# Knowledge Required

## *Knowledge of stems or roots*

*Duck* is a possible root, not *duckl*.

*We need a dictionary (lexicon)*

## *Morphotactics*

Which class of morphemes follow other classes of morphemes inside the word?

*Ex: plural morpheme follows the noun*

## *Only some endings go on some words*

- *Do+er*: ok
- *Be+er*: not so

## *Spelling change rules*

Adjust the surface form using spelling change rules

- Get + er → getter

# *Why can't this be put in a big lexicon?*

- English: just 317,477 forms from 90,196 lexical entries, a ratio of 3.5:1

# *Why can't this be put in a big lexicon?*

- English: just 317,477 forms from 90,196 lexical entries, a ratio of 3.5:1
- Sanskrit: 11 million forms from a lexicon of 170,000 entries, a ratio of 64.7:1

# *Why can't this be put in a big lexicon?*

- English: just 317,477 forms from 90,196 lexical entries, a ratio of 3.5:1
- Sanskrit: 11 million forms from a lexicon of 170,000 entries, a ratio of 64.7:1
- New forms can be created, compounding etc.

*One of the most common methods is finite-state-machines*