Lecture 9: Words, morphology, and parts of speech

Zhizheng Wu

Agenda

- Recap
- Words
- Morphology: Internal structure of words
- Parts of speech

Byte-pair encoding

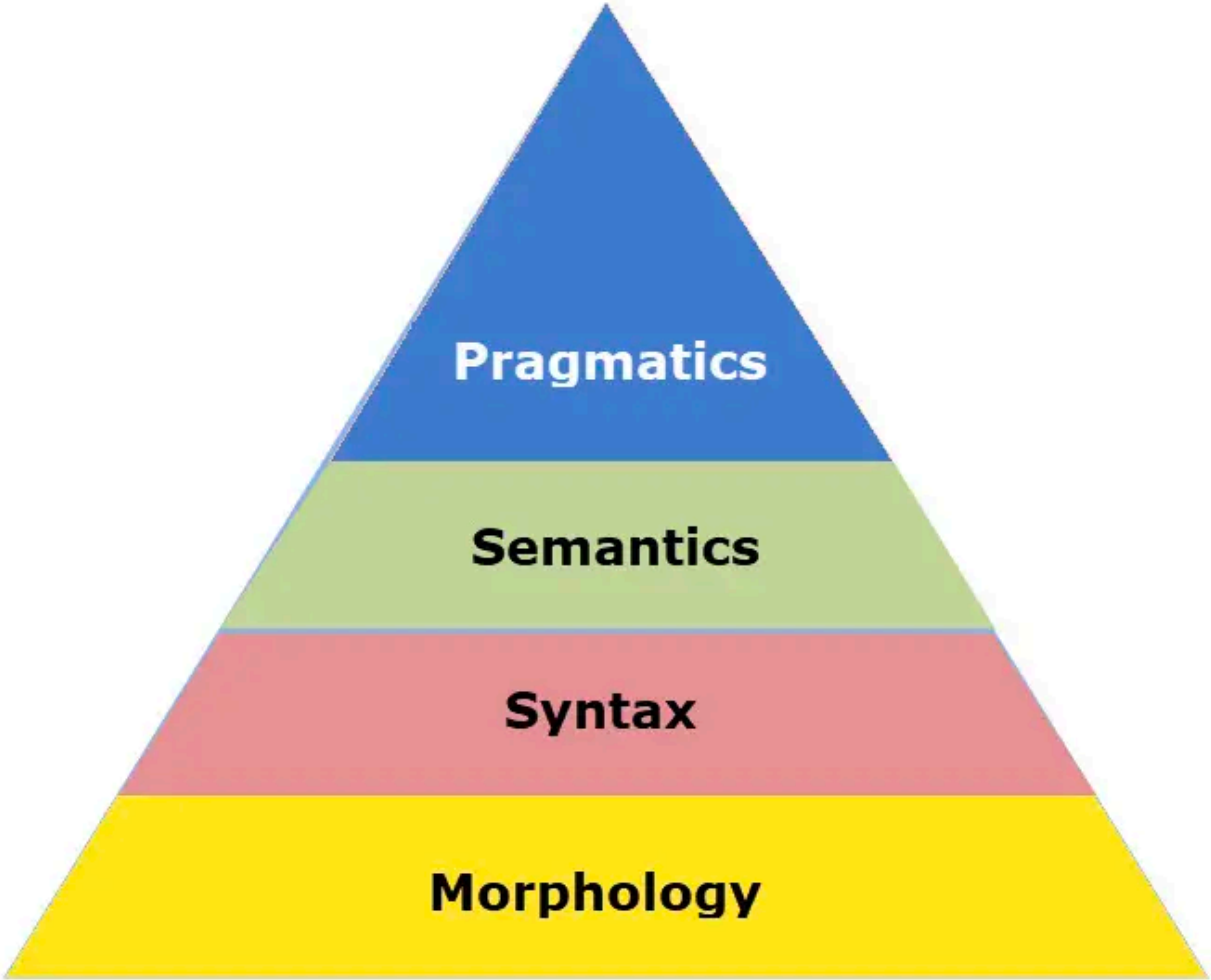
- Originally proposed for lossless data compression

aaabdaaabac	
aaabdaaabac	Replace aa with Z
ZabdZabac	Replace ab with Y
ZabdZabac	Replace ab with Y
ZYdZYac	
…	

Edit distance table

	M	O	N	K	E	Y
M	0	1	2	3	4	5
O	1	0	1	2	3	4
N	2	1	0	1	2	3
E	3	2	1	2	1	2
Y	4	3	2	3	2	1

words
vocabulary
students
text
academic
learn
love
word
economy
learning
collocations
ask
focus
choose
ones
five
boost
good
important
think
see
come
skills
AWL
everywhere
teach
texts
study
first
look
strategies
Love
need
also
used
use
items
time
however
comprehension
engagement
definitions
across
quick
go
practice
money
uses
studies
expect
highlighted
understand
connotations
follow
every
included
lesson
tax
may
mean
frequently
grammatical
class
List
frequent
board
quickly
primary
give
back
sum
list
note
reading
etc
level
read
single
keep
journals
associations
find
Coxhead
many
school
stay
move
sentences
cost
three
features
alternative
asked
pronunciation
generally
Covering
antonyms
view
talk
long
side
chit-chat
story
chosen
thing
much
impossible
display
critics
samples
free
day
needs
combinations
aims
today
carefully
keeping
Middleton
Wedding
paste
times
tool
new
sums
get
Complete
discuss
Choose
big
light
example
learners
knowledge
interesting
following
began
using



Pragmatics

Semantics

Syntax

Morphology

Natural Language Processing Pyramid

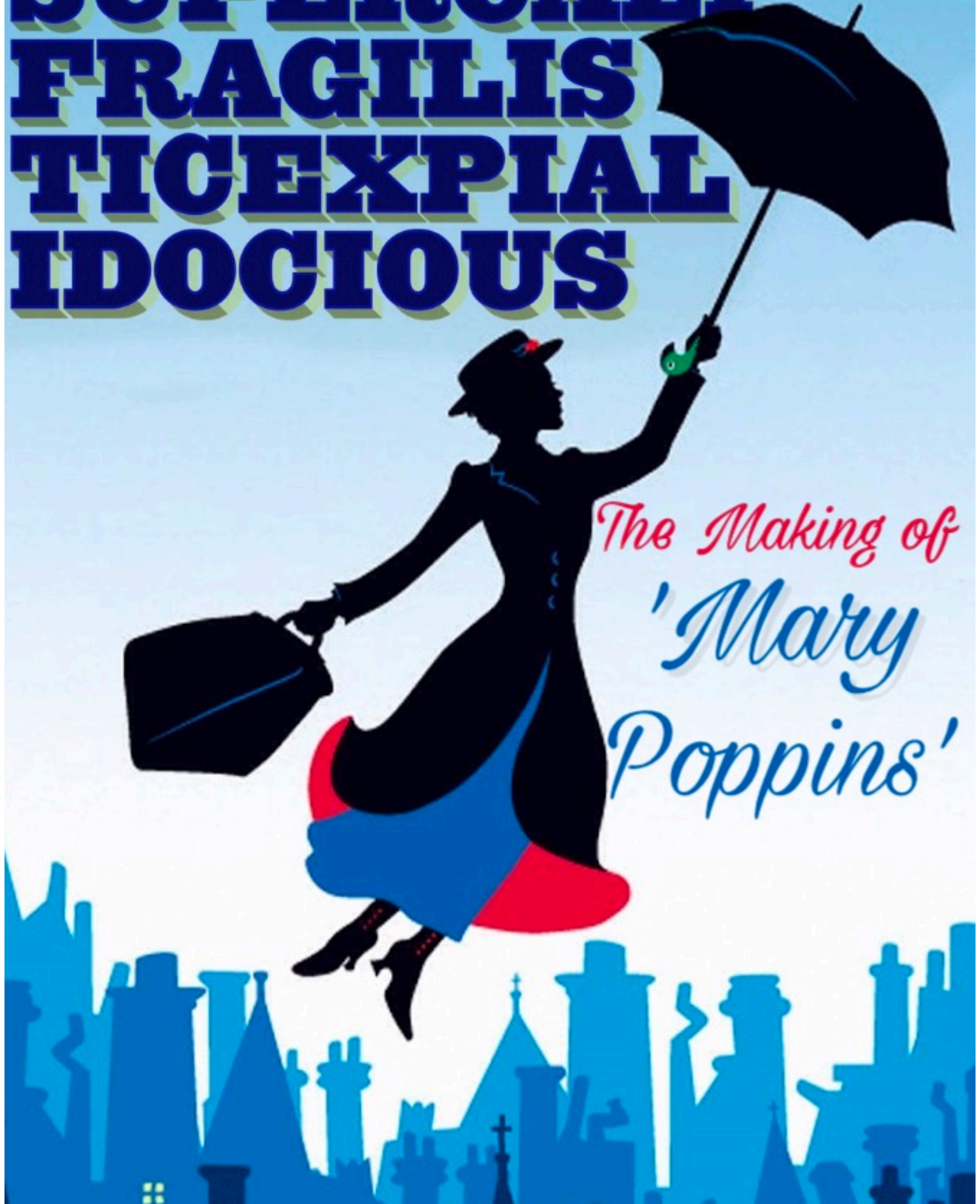
Word

- Words are at the interface between phonology, syntax and semantics

- Words are not atoms
  - They have internal structure

supercalifragilisticexpialidocious

SUPERCALIFRAGILISTICEXPIALIDOCIOUS



The Making of 'Mary Poppins'

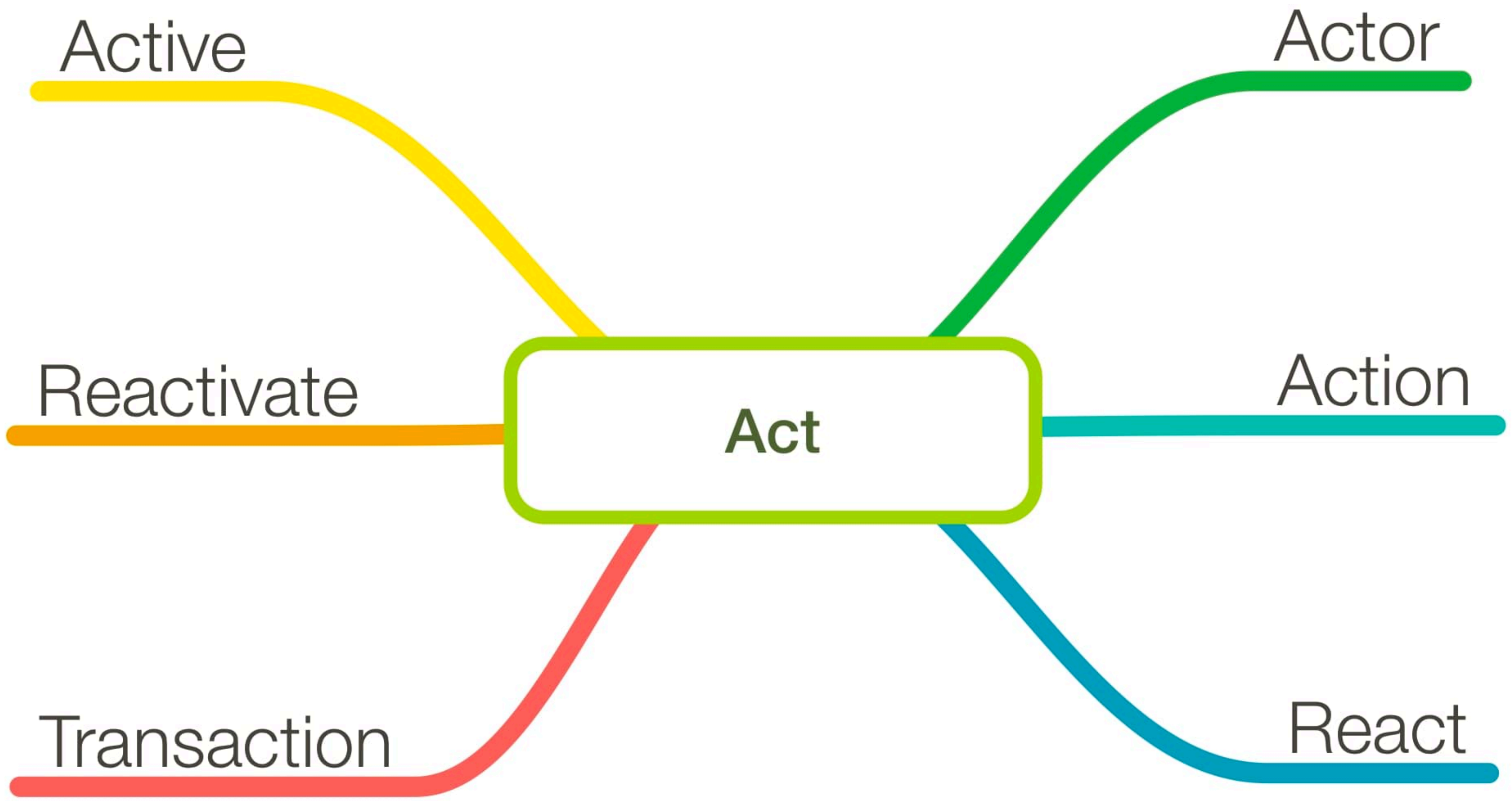
Linguistic morphology

- The study of how morphemes join together to form words
- Morphemes are the ***minimal*** units of linguistic form and ***meaning***

supercalifragilisticexpialidocious

Why do we need to learn morphology?

- The creation of new words
- The modification of existing words. We create new words out of old ones all the time



Differences between Words and Morphemes

- Another difference between words and morphemes is that between two words, we can usually insert some other words, while between two morphemes we can't

  - She has arrive-d.
  - She has already arrive-d.
  - She has arrive-d already.
  - *She has arrive-already-d.

Differences between Words and Morphemes

- Whitespace is not always a good test for the word/morpheme distinction in English. Compound nouns are often spelled with whitespace between their components, yet they are a single word

  - Picture frame
  - Swim team

Chinese example

- In classic Chinese, usually each character is a word and also a morpheme
- Most words in modern Standard Chinese (i.e. Mandarin) are compounds and most roots are bound

难易相成

难和易是相互转化的

Category

1 morpheme

Neither cat nor gory has nothing to do with the meaning of category in English

Categorize

2 morphemes

Category + ize

Categorized

3 morphemes

Category + ize + ed

Overestimating

3 morphemes

over + estimate + ing

Word has structures

supercalifragilisticexpialidocious

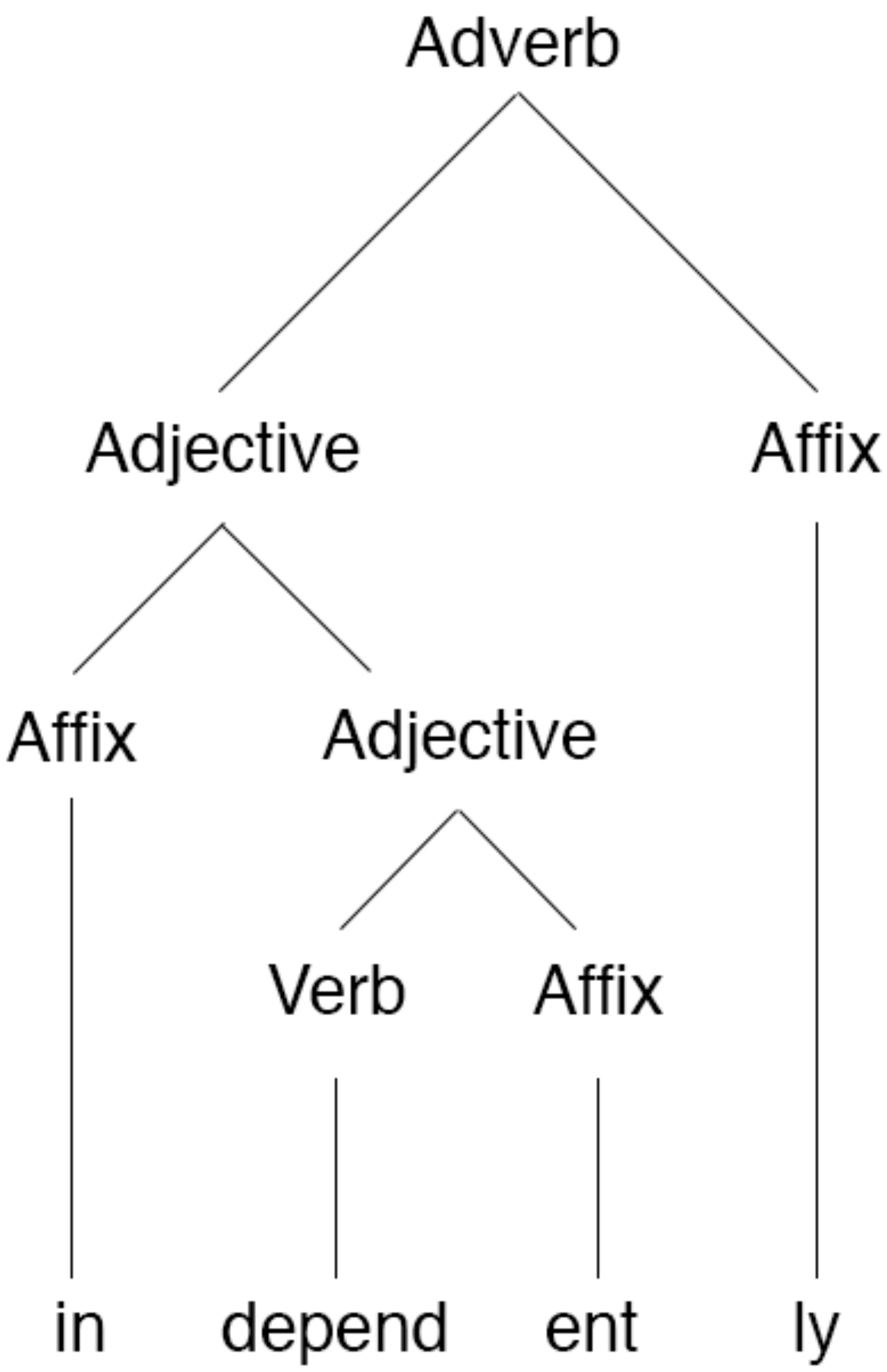
supercalifragilisticexpialidocious

super- "above", cali- "beauty", fragilistic- "delicate", expiali- "to atone", and -docious "educable",

Morpheme: Root

- Root
  - The central morphemes in words, which carry the main meaning

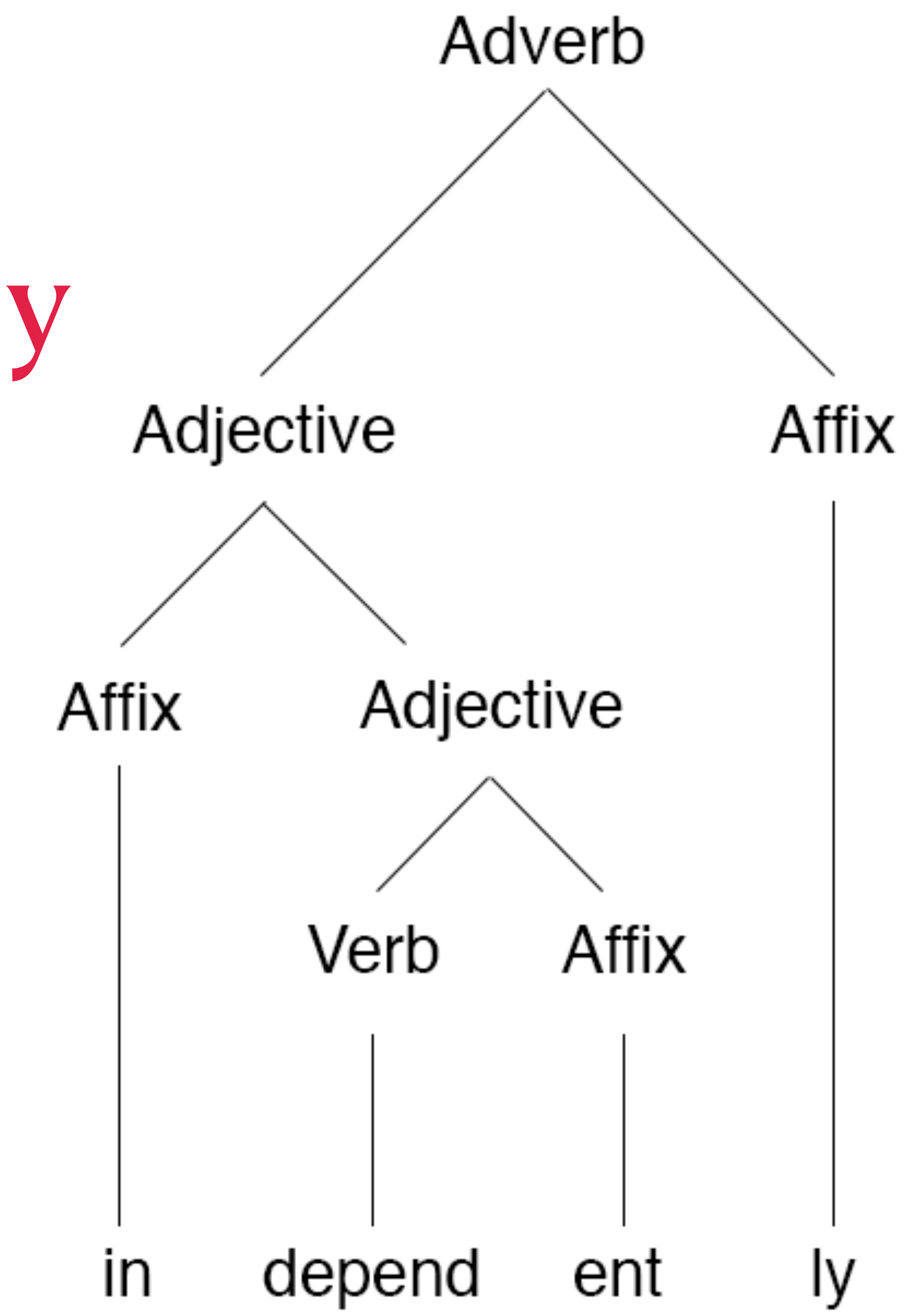
Inde***pend***ently



Morpheme: Affixes

- Affixes
  - Prefixes
    - de-pend, in-correct
  - Suffixes
    - depend-ent, love-ly
  - Infixes (not common)
    - edu-ma-cation
  - Circumfixes

Inde*pend*ently



Edu-ma-cation

Used in a sarcastic sense, or in dialogue, suggesting lack of education on the part of the speaker

Nonconcatenative morphology

- Umlaut
  - Foot : feet
  - Tooth : teeth
- Ablaut
  - Sing, sang, sung
- Root-and-pattern or templatic morphology
  - Common in Arabic, Hebrew, and other Afroasiatic languages
  - Roots made of consonants, into which vowels are shoved
- Infixation
  - Gr-um-adwet

Functional differences in morphology

- Inflectional morphology
  - Adds information to a word consistent with its context within a sentence
    - Student -> students
    - Sleep -> sleeping
    - Listen -> listening
- Derivational morphology
  - Creates new words with new meanings (and often with new parts of speech)
    - Sing -> singer
    - Beauty -> beautiful

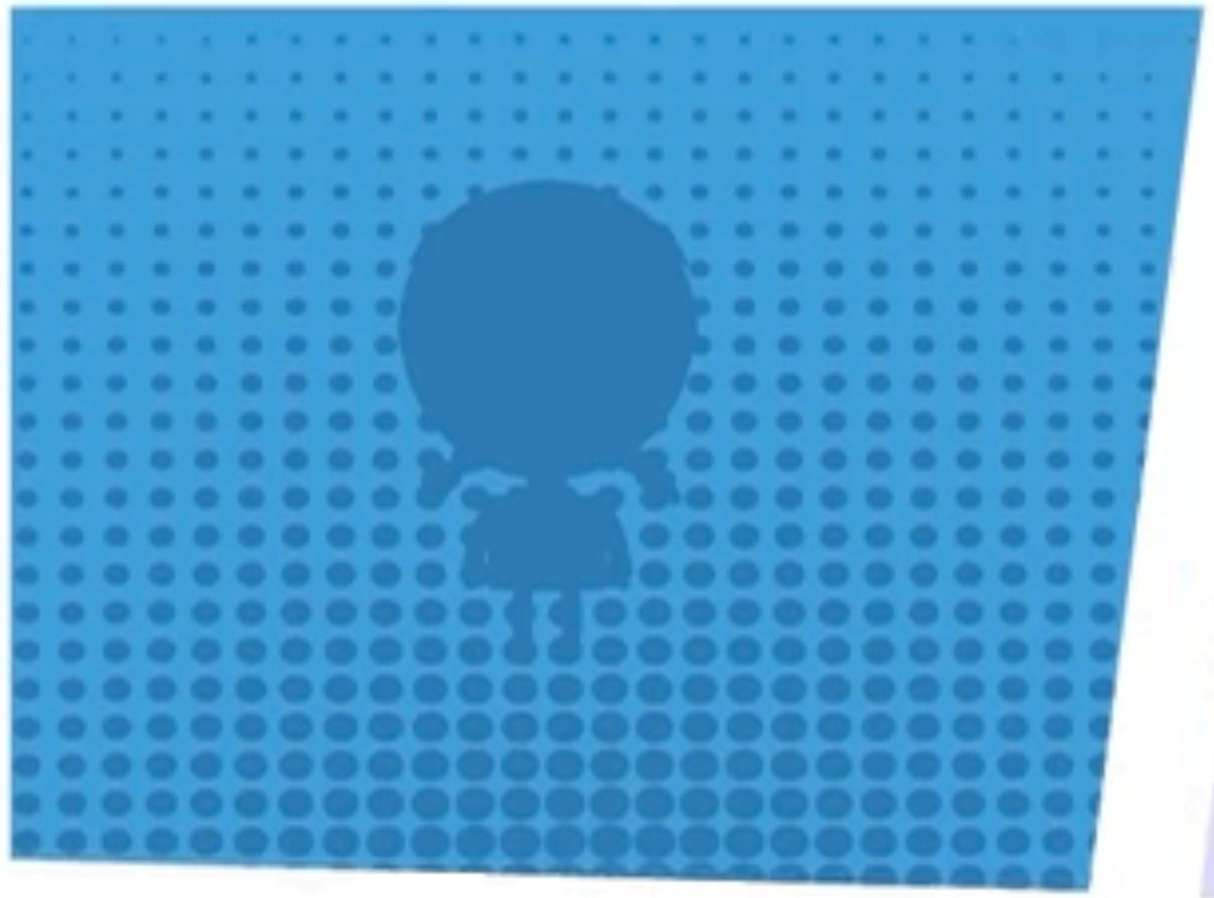
Weirdness of morphology

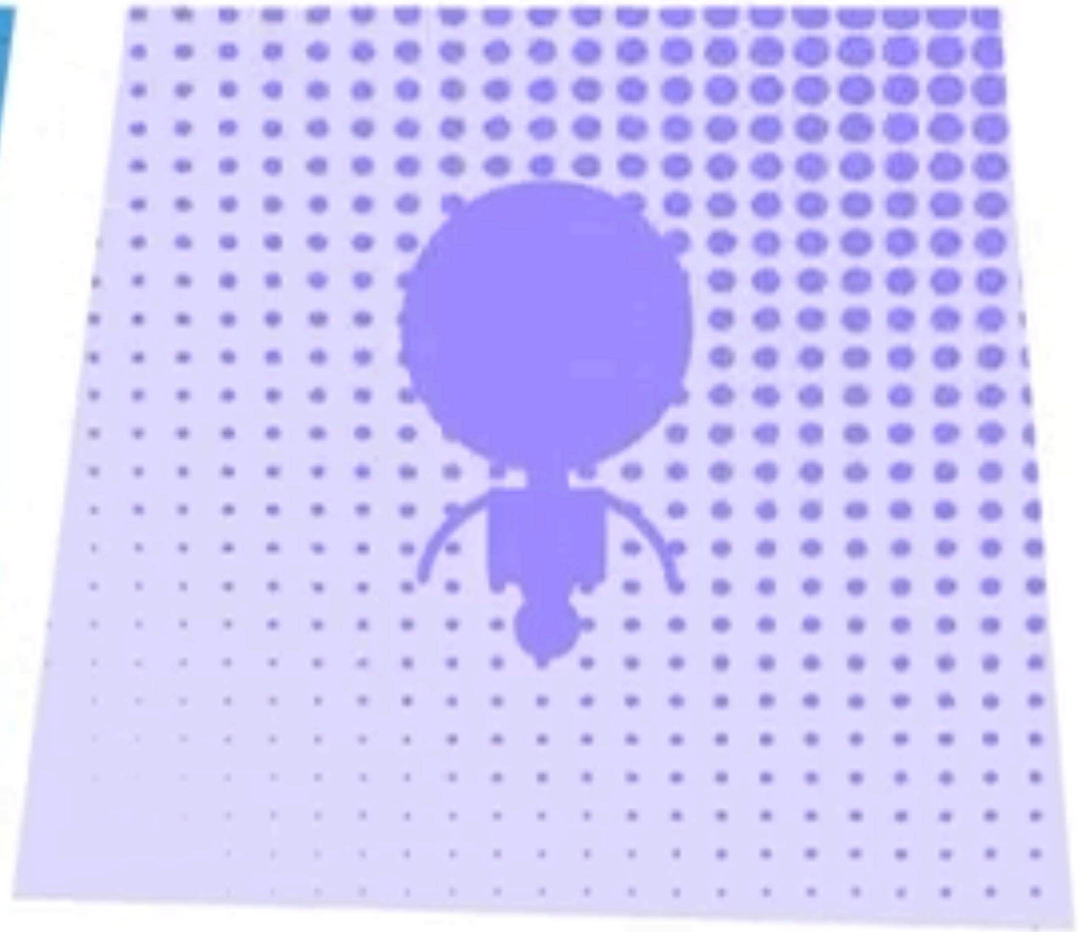
- Anything that a language does with morphology, it usually can also do more straightforwardly with syntax.

- Example: Plural marking
  - Dog -> dog-s

  - more than one dog

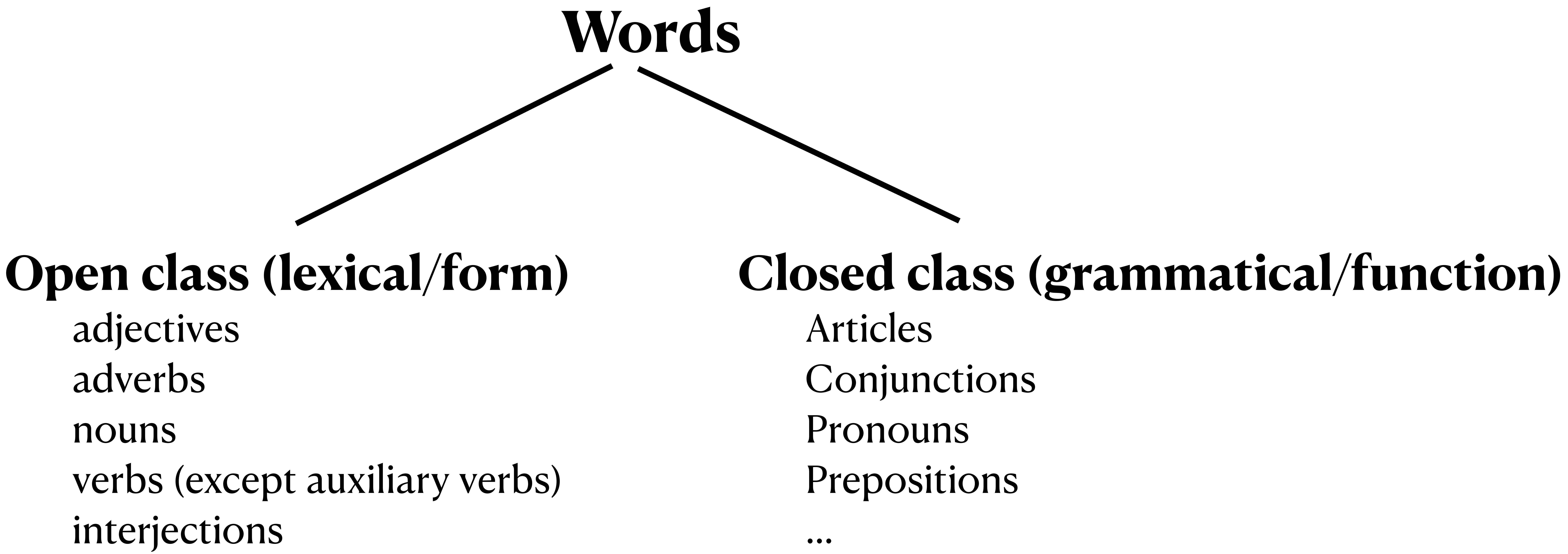
Parts of speech





Parts of Speech League, Parts of Speech League, Parts of Speech League!

Two classes of words



Words

Open class (lexical/form)

- adjectives
- adverbs
- nouns
- verbs (except auxiliary verbs)
- interjections

Closed class (grammatical/function)

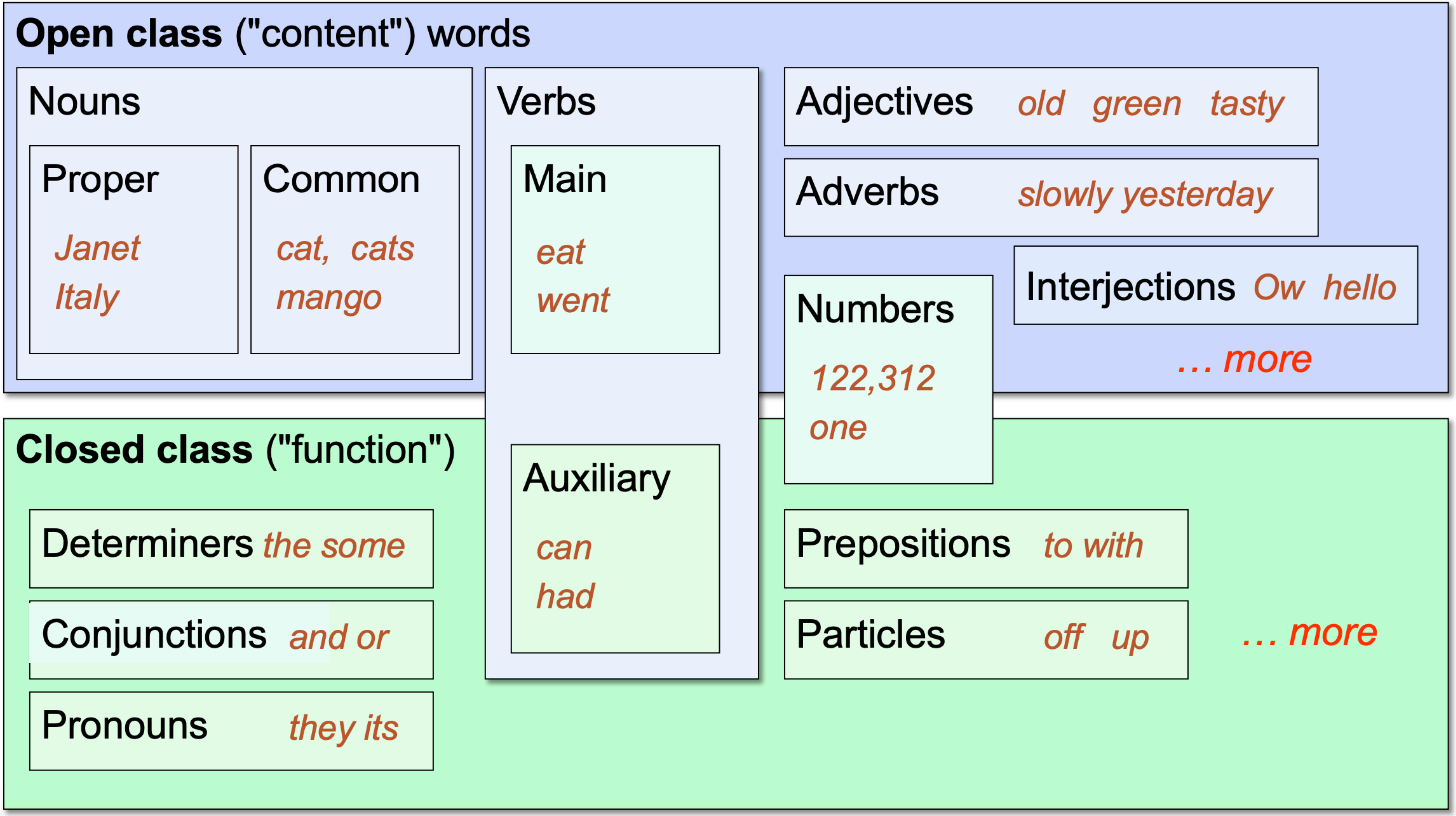
- Articles
- Conjunctions
- Pronouns
- Prepositions
- …

Two classes of words: Open class

- Open class
  - Usually content words: Nouns, Verbs, Adjectives, Adverbs
    - Plus interjections: oh, ouch, uh-huh, …
  - New nouns and verbs like iPhone

Two classes of words: Closed class

- Closed class
  - Relatively fixed membership
  - Usually function words: short, frequent words with grammatical function
    - Determiners: a, an, the
    - Pronouns: she, he, I
    - Prepositions: on, under, over, …



Open class ("content") words

Nouns

Proper *Janet Italy*

Common *cat, cats mango*

Verbs

Main *eat went*

Auxiliary *can had*

Adjectives *old green tasty*

Adverbs *slowly yesterday*

Numbers *122,312 one*

Interjections *Ow hello*

… more

Closed class ("function")

Determiners *the some*

Conjunctions *and or*

Pronouns *they its*

Prepositions *to with*

Particles *off up*

… more

Words are ambiguous

- A word can have more than one possible part-of-speech

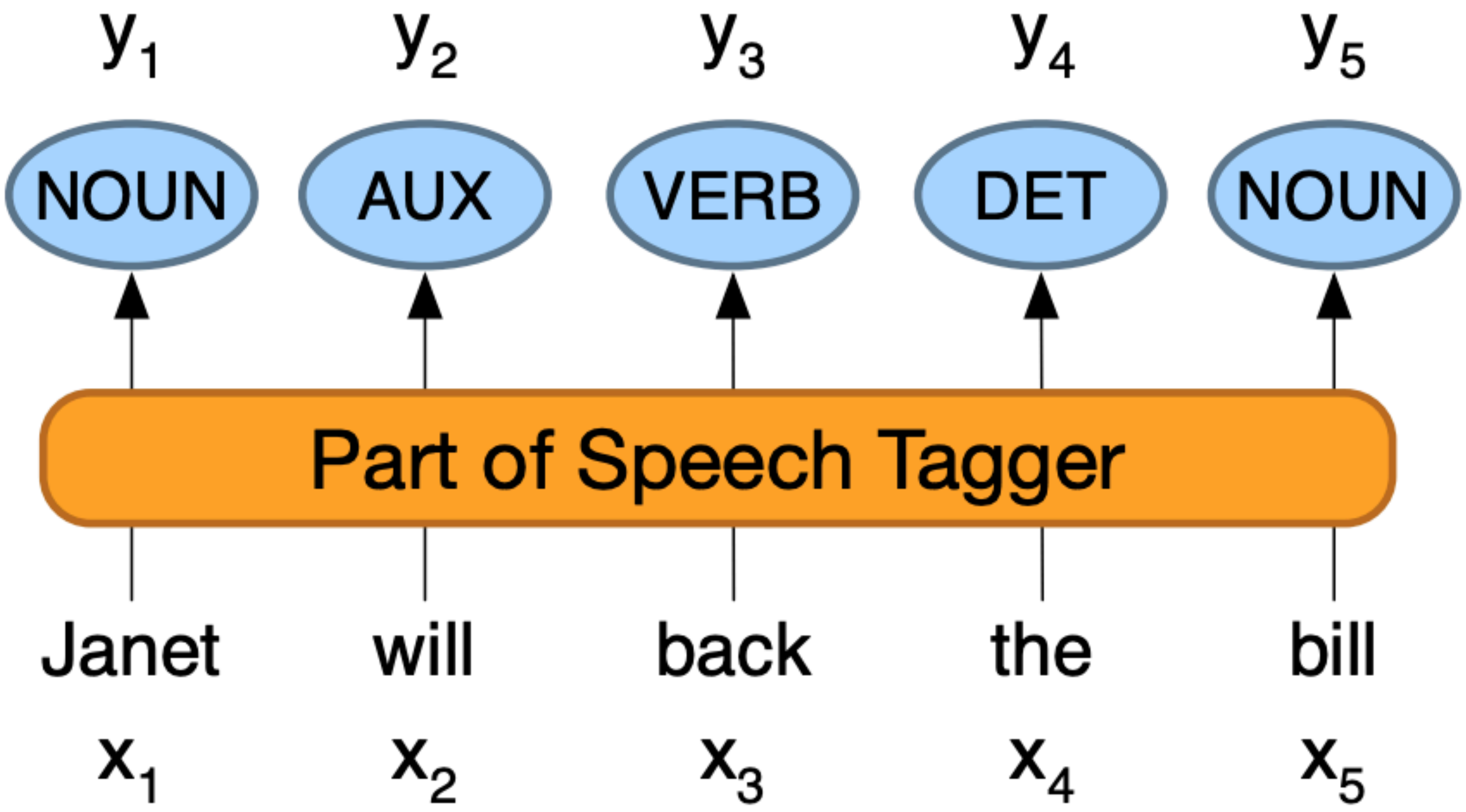
  - She is reading a ***book*** about airplane
  - She will ***book*** a flight for you

Part-of-speech tagging is a disambiguation process

Verb or Noun? Verb or Noun?

She is **reading** a book about **Reading**

POS tagging as a sequence labeling task



Summary

- Words are at the interface between phonology, syntax and semantics

- Words have internal structures, and morphemes are the ***minimal*** units of linguistic form and ***meaning***

- A word can have more than one possible part-of-speech
  - Words can grouped into open and closed classes
  - Part-of-speech tagging is a disambiguation process

Readings

- Chapter 8: Sequence Labeling for Parts of Speech and Named Entities
  - https://web.stanford.edu/~jurafsky/slp3/