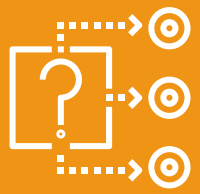
LEVELS OF NATURAL LANGUAGE PROCESSING

WOA7013

Purpose



To understand the levels of NLP.

Introduction

- A simple sentence consists a Subject followed with Predicate.
- A word in a sentence acts as **Part of Speech (POS).**
- For an example, for English sentence, the parts of speech are:
  - ***Nouns***
  - ***Pronouns***
  - ***Adjectives***
  - ***Verb***
  - ***Adverb***
  - ***Prepositions***
  - ***Conjunctions***
  - ***Interjections***.
- Most of us understand both written and spoken language, but reading is learned much later, compared to spoken language.

NLP Aspects

SYNTAX

The syntax describes the form of the language. It is usually specified by a grammar.

SEMANTIC

The semantics provides the meaning of the utterances or sentences of the language.

PRAGMATIC

The pragmatic component explains how the utterances relate to the world. To understand language, an agent should consider more than the sentence; it has to take into account the context of the sentence, the state of the world, the goals of the speaker and the listener.

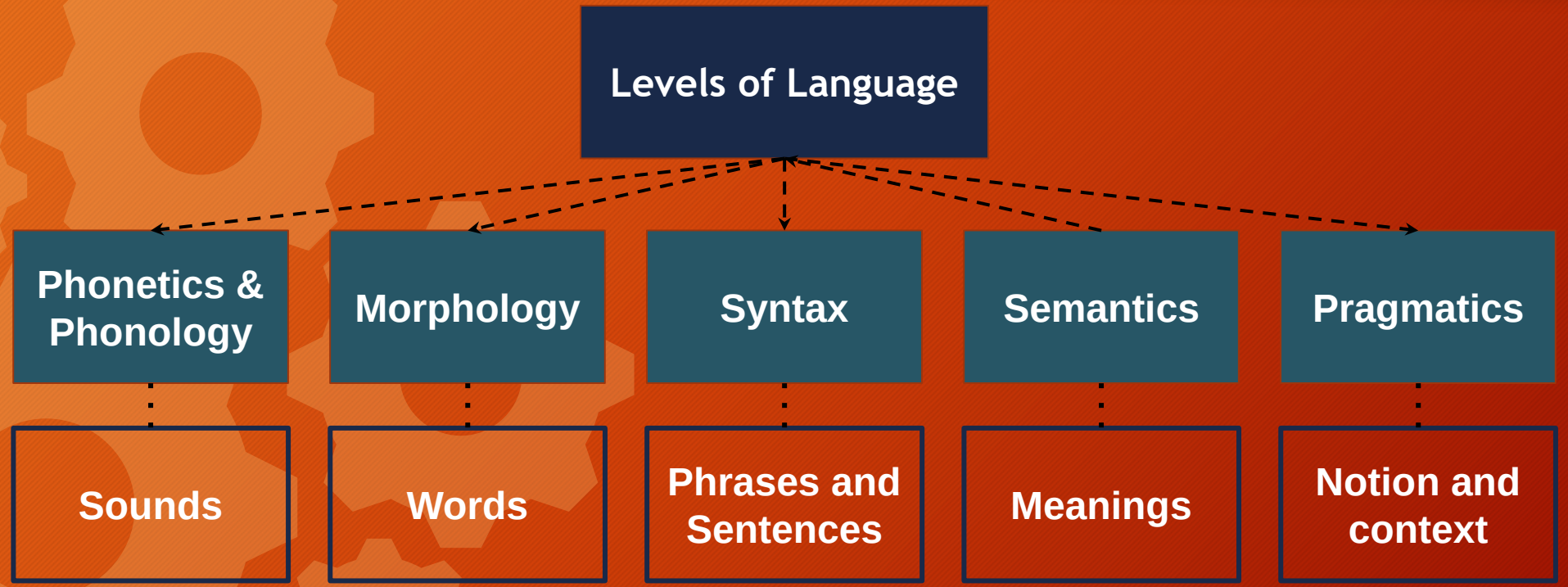
To understand the difference among these levels, consider the following sentences:

1. *This course is about Natural Language Processing.*
2. *The green frogs sleep soundly.*
3. *Colorless green ideas sleep furiously.*
4. *Furiously sleep ideas green colorless*

Examples

1. *This course is about Natural Language Processing.*
   - ✓ syntactically, semantically, and pragmatically well formed
2. *The green frogs sleep soundly.*
   - ✓ syntactically and semantically well formed, but it would sound very strange; it is thus not pragmatically well formed for that context
3. *Colorless green ideas sleep furiously.*
   - ✓ syntactically well formed, but it is semantically nonsensical
4. *Furiously sleep ideas green colorless.*
   - ✓ syntactically ill formed; it does not make any sense - syntactically, semantically, or pragmatically

Levels of NLP



Source:
https://www.uni-due.de/SHE/REV_Levels_Chart.htm

Phonology: 3 types of rules used in phonological analysis

Phonetic rules

For sounds within words;

The difference between these words: write / right /

Phonemic rules

For variations of pronunciation when words are spoken together.

Phoneme : speech sound that helps us constructing meaning

- */r/ : rubble double, Hubble, fubble, wubble.*
- */u/ : rubble rabble, rebel, Ribble, robble...*

Prosodic rules

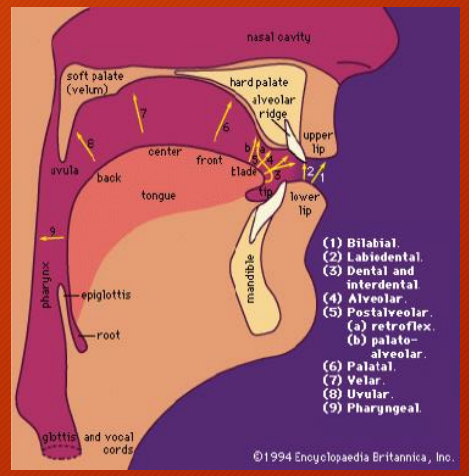
For fluctuation in stress and intonation across a sentence.

Articulotory phonetics

- Concerned with the production of speech sounds.
- Articulation refers to the way we position our lips and tongue against our teeth and the roof of our mouth to make the various sounds that go together to form words.

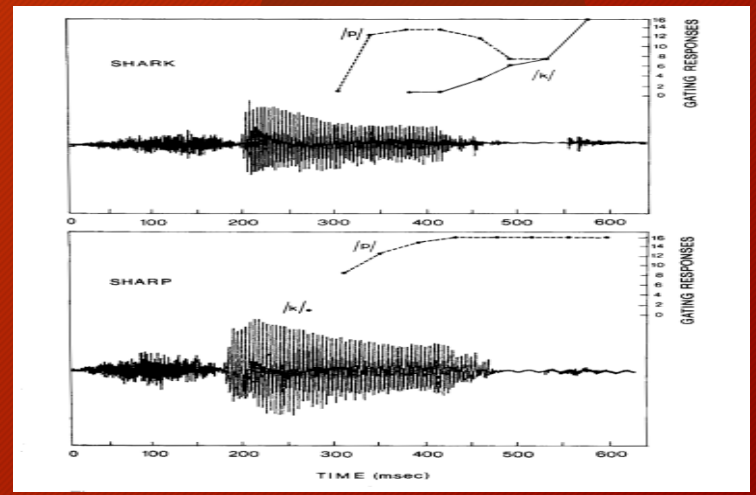
Acoustics phonetics

- Deals with the transmission and physical properties of speech sounds or sound waves (frequency and harmonics)



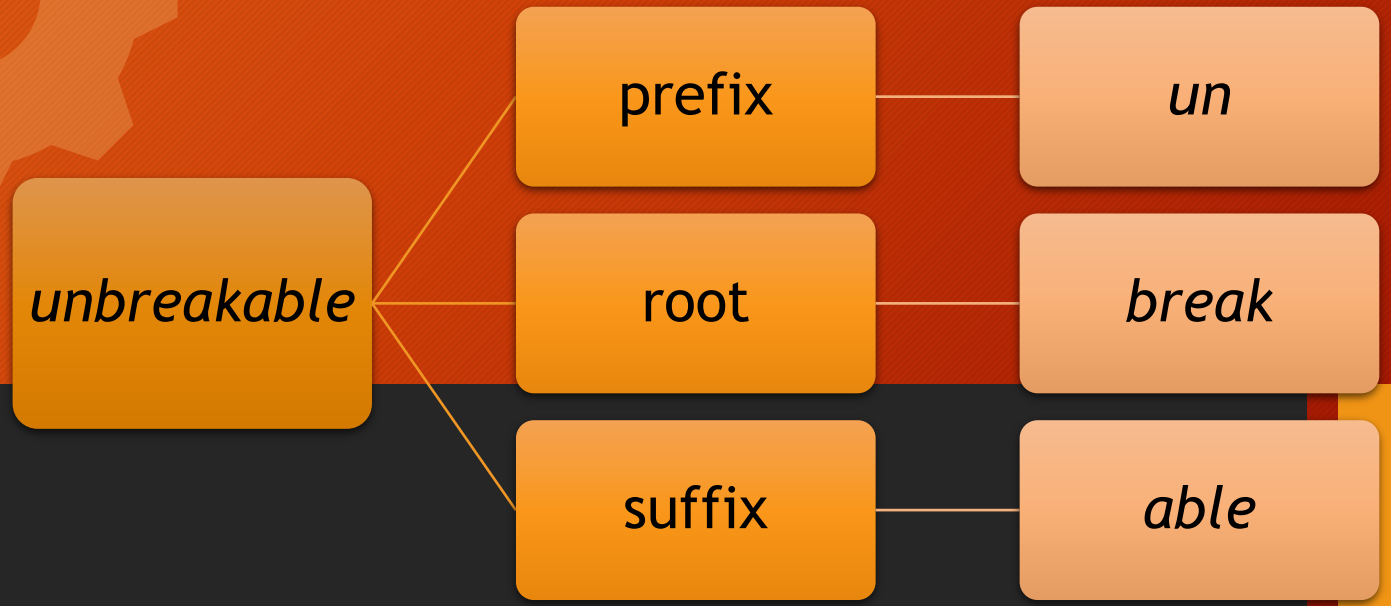
Auditory phonetics

- The study of perception of speech sounds.



Morphology

- Concerns how words are constructed from more basic meaning units called morphemes - the smallest units of meaning in a language.
  Producing and recognizing variations of individual words
  The way words break down into component parts that carry meaning (like sg / pl). For e.g. :



- Morphology is the identification, analysis and description of structure of words.
- Study the structure of words - Morphological analysis:
  - Inflection

    duck + *s* = $[_{N}$duck] + $[_{plural}$ s]

    duck + *s* = $[_{V}$duck] + $[_{3rd\ person}$ s]

  - Spelling changes:

    Drop → Dropping

    Hide → Hiding

  - Derivation

    kind (adj) → kindness (n)

Morphology Methods

Lemmatisation

- process of grouping together the different inflected forms of a word so they can be analysed as a single item
- Need to determine the part of speech of a word in a sentence (requiring grammar knowledge)

Stemming

- operates on a single word without knowledge of the context
- cannot discriminate between words which have different meanings depending on part of speech
- easier to implement and run faster, reduced accuracy may not matter for some applications

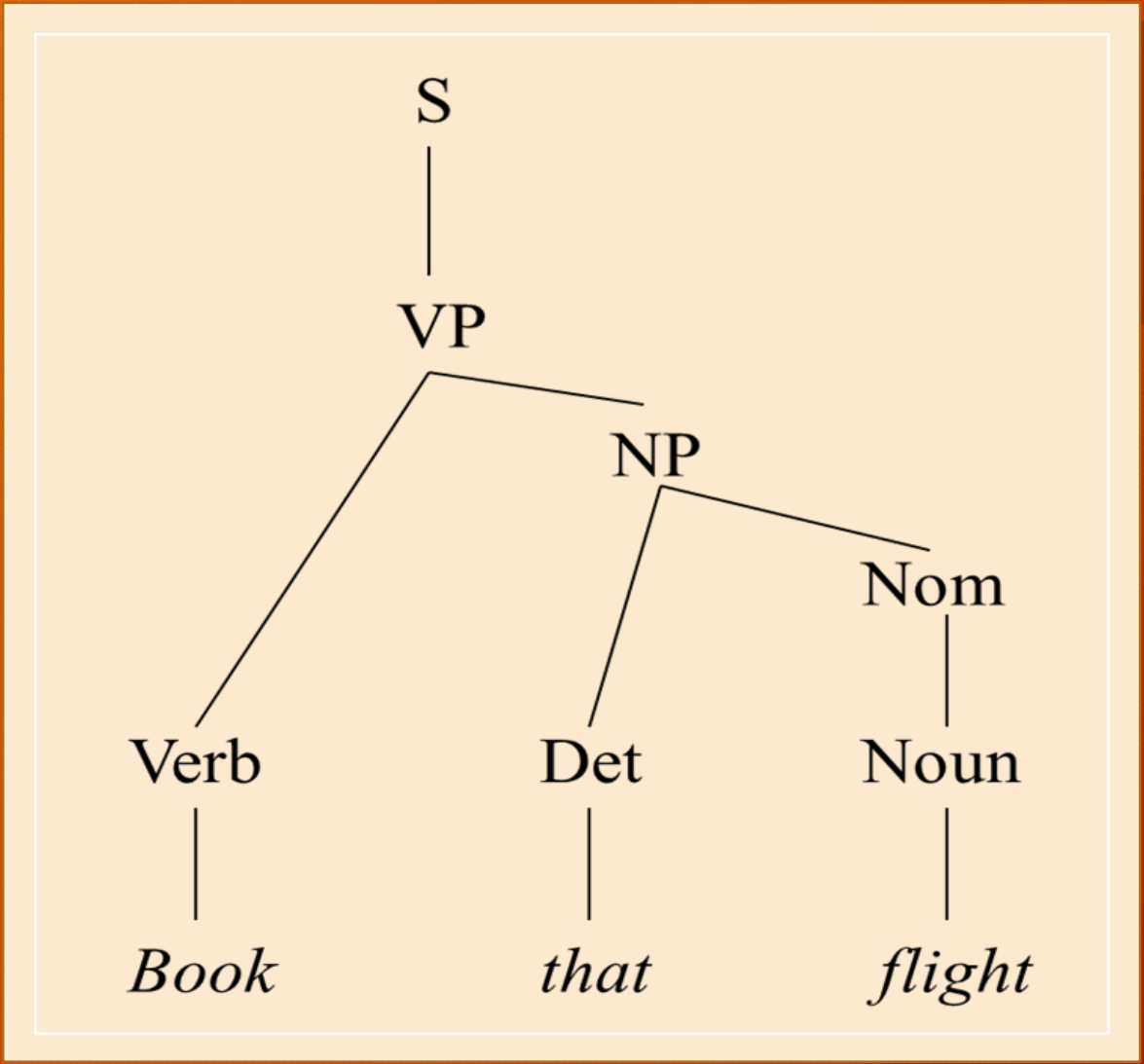
Syntax

- Structure of language
- Languages have structure:
  - not all sequences of words over the given alphabet are valid
  - when a sequence of words is valid (grammatical), a natural structure can be induced on it.
- Grammatical arrangement of words in a sentence to show its relationship to one another in the sentence.
- Concerns how words can be put together to form correct sentences.

John hit Bill
Bill was hit by John
Bill hit John
Bill, John hit
Who John hit was Bill

- Describes the constituent structure of NL expressions
  - *(Bill (was hit) (by John))*
- Grammars are used to describe the syntax of a language
- Syntax is represented by Parse Tree.
  - represents the syntactic structure of a string according to some formal grammar.
  - the interior nodes are labeled by nonterminals of the grammar, while the leaf nodes are labeled by terminals of the grammar.
- Parser – a procedure that compares the grammar against input sentences to produce a parsed structure called parse tree.
- E.g. A sentence: *Book that flight*.

- Example of the syntax tree for the sentence: *Book that flight*.



- Concerns what words mean and how these meanings combine in sentences to form sentence meanings. Semantic processing determines the possible meanings of a sentence by focusing on the interactions among word-level meanings in the sentence.

Semantic

- Meanings of words / phrases / sentences / text.
- Lexical semantics: word meaning
  - Example:
    - John picked up a bad cold.
    - John picked up a large rock.
  - What is the meaning of this word: *to get?*
-

1. come to have or hold; receive.
2. succeed in attaining, achieving, or experiencing; obtain.
3. experience, suffer, or be afflicted with.
4. move in order to pick up, deal with, or bring.
5. bring or come into a specified state or condition.
6. catch, apprehend, or thwart.
7. come or go eventually or with some difficulty.
8. move or come into a
9. specified position or state

Pragmatic

- Concerns with how language is used, that is, meaning in context.
- This requires much world knowledge, including the understanding of *intentions*, *plans*, and *goals*.
- E.g., if someone says "*the door is open*", then it is necessary to know which door "*the door*" refers to.
- In this case, we need to know the intention of the speaker:
  - Could be a pure statement
  - Could be an explanation how the cat got in
  - Could be a request to the person addressed to close the door.

Concerns how sentences are used in different situations and how use affects the interpretation of the sentence.

- Scene 1: KL Central
  - You : KLCC?
  - Passerby: Downstairs, LRT Station.
- Scene 2: Ticket Counter, LRT Station
  - You : KLCC?
  - Clerk: RM4.50.

Scene 3: Information Booth, LRT Station

You : KLCC?

Clerk: 4:20, Platform 5.

Scene 4: On the train

You : KLCC?

Passenger: No, I'm going to Bangsar.

For e.g.: A play in one act **"Going to a shopping Mall - KLCC"**

Knowledge about the kind of actions that speakers intend by their use of sentences

Stages in Natural Language Processing

- Lexical / Morphological Analysis

- Syntactic Analysis

- Semantic Analysis

- Pragmatic Analysis