Main components

① Lexical Analysis — analyzing text given a lexicon

② Morphology — how words are formed from smaller parts.

③ Syntactical Analysis — applying rules of lang to stuff we found earlier.

④ Semantical Analysis — understanding meaning

⑤ Representation of Semantics — represent the meaning.

⑥ Generation of Meaningful Text.   (optional)

So what is NLP?

→ It's a branch of AI w/ 2 goals

Understand how lang. Operates
(Science goal)

Build systems that analyze & generate language
(Engineering goal)

→ It's NOT computational linguistics
→ aims to MODEL language.

NLP is processing lang from computational pov IN ORDER to build diff. appl^n & tools.

→ Turing test :

no diff?

machine response

human response

Bob

Semantics

↳ all about the meaning of words

Lexical
what words mean
ALONE

Compositional
How meanings get combined.

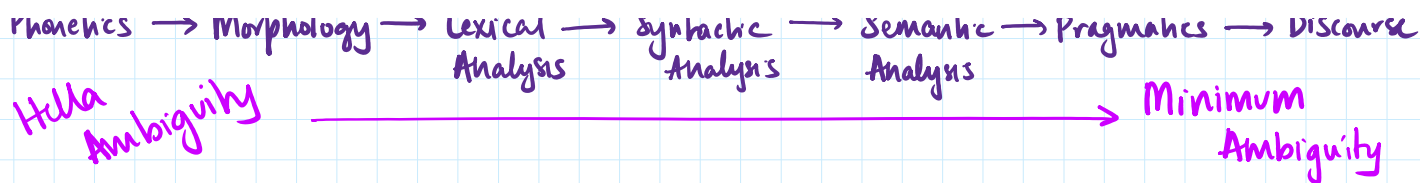→ Implicit meanings — not directly stated by the user.

↓

rely on context + shared knowledge.

How does language processing usually go?

Phonetics → Morphology → Lexical Analysis → Syntactic Analysis → Semantic Analysis → Pragmatics → Discourse
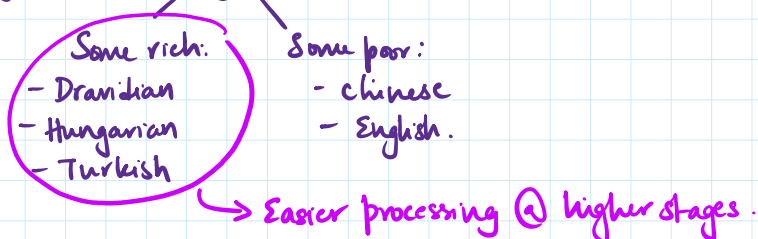
~ Minimum

Namita Achyuthan

Phonetics → Morphology → Lexical → Syntactic → Semantic → Pragmatics → Discourse
                          Analysis   Analysis   Analysis

Hella Ambiguity ————————————————————————→ **Minimum Ambiguity**

① Phonetics —— speech processing —— ambiguity?
                                    HOMO PHONES

② Morphology —— disambiguate by (rules) → form words from ETCs

   ⟨ Some rich: Some poor:
     - Dravidian   - chinese
     - Hungarian   - English.
     - Turkish ⟩
            → Easier processing @ higher stages.

③ Lexical analysis —— disambiguate by referring to dict.
                  Eg: Dog
              1. Disambiguate as POS
                      Dog — (n)
                      Dog — (v)
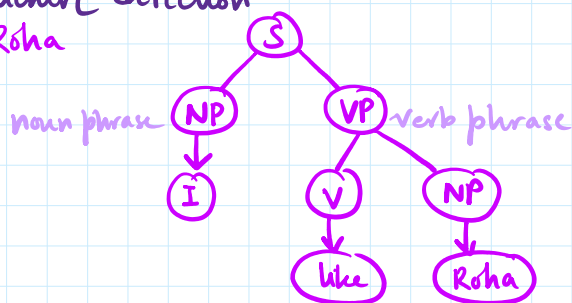   2. Disambiguation step
                      Dog — animal
                      Dog — bad person
                      B*tch — -"-

④ Syntactic analysis:

      A. disambiguate by structure detection
             Eg: I like Roha
                                        (S)
           noun phrase (NP)   (VP) verb phrase
                      | /   \
                     (I)   (V)   (NP)
                                  | |
                               (like)   (Roha)

      B. disambiguate by scope (region of influence)
             Eg: I saw a boy w/ a telescope

   me   boy telescope   me   telescope   boy.

   Possible soln: Resolution
              ↓   ↘
          Anaphora   Cataphora

Namita Achyuthan

Anaphora resolution is the process of identifying what a pronoun or a noun phrase refers to in a sentence. For example:

- **Sentence**: "The empress has not arrived but she should be here any moment."
- **Anaphor**: "she"
- **Antecedent**: "The empress"

In this case, "she" refers back to "The empress." Anaphora resolution helps in understanding that both "she" and "The empress" refer to the same entity.

Cataphora resolution is the opposite of anaphora resolution. It occurs when a pronoun or a noun phrase refers to something mentioned later in the text. For example:

- **Sentence**: "Because she was going to the post office, Madhu was asked to post a parcel."
- **Cataphor**: "she"
- **Antecedent**: "Madhu"

Here, "she" refers to "Madhu," which is mentioned later in the sentence. Cataphora resolution helps in understanding that "she" and "Madhu" refer to the same entity, even though "she" appears first.

⑤ Semantic Analysis — knowledge representation in terms of

predicate calculus   Semantic nets   frames   Conceptual dependencies.

Challenge: ambiguity in semantic role labelling.

⑥ Pragmatics — super hard.

Pragmatics is concerned with the ways in which language is used in context and how the context influences the interpretation of meaning. It goes beyond the literal meaning of words to understand the speaker's intentions and the social aspects of communication.

→ User intention
\+
→ World knowledge disambiguation

⑦ Discourse
↳ processing a sequence of sentences
↓
needs reasoning + appl^n of world knowledge.

We essentially form a bunch of hypotheses + discard as we get new evidence.

Components of NLP Systems.

I/P text string → [Parser → Semantic Analyzer → Generator] → O/P

Parser ↕ Lexicon

Semantic Analyzer ↕ Knowledge Base