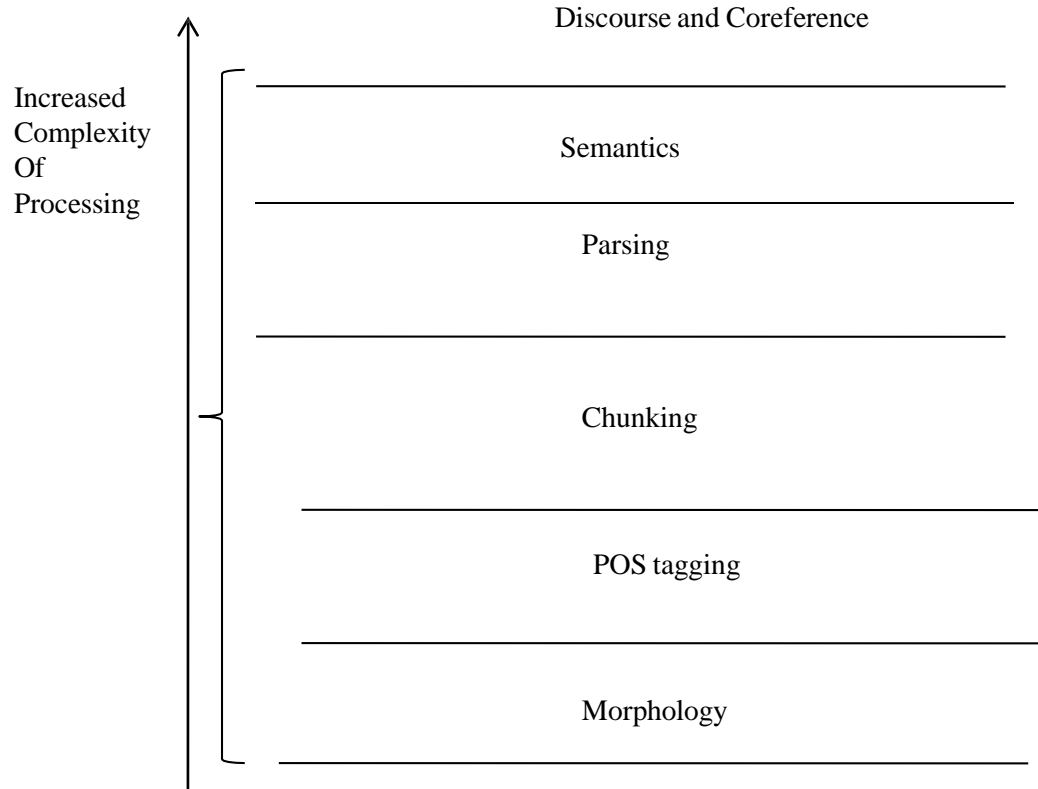


Layers in NLP

Layers of Language Processing



Different Levels of Language Analysis

1. Phonetic and phonological knowledge

- Words are related to sound

2. Morphological Knowledge

- Word are constructed from morphemes

3. Syntactic Knowledge

- Words put together to form correct sentence
- Structural role played by each word

4. Semantic Knowledge

- What word means
- Context independent meaning- meaning the sentence has regardless of the context in which it is used

Different Levels of Language Analysis

5. Pragmatic Knowledge

- How sentences are used in different situations and how use affects the interpretation of the sentence

6. Discourse Knowledge

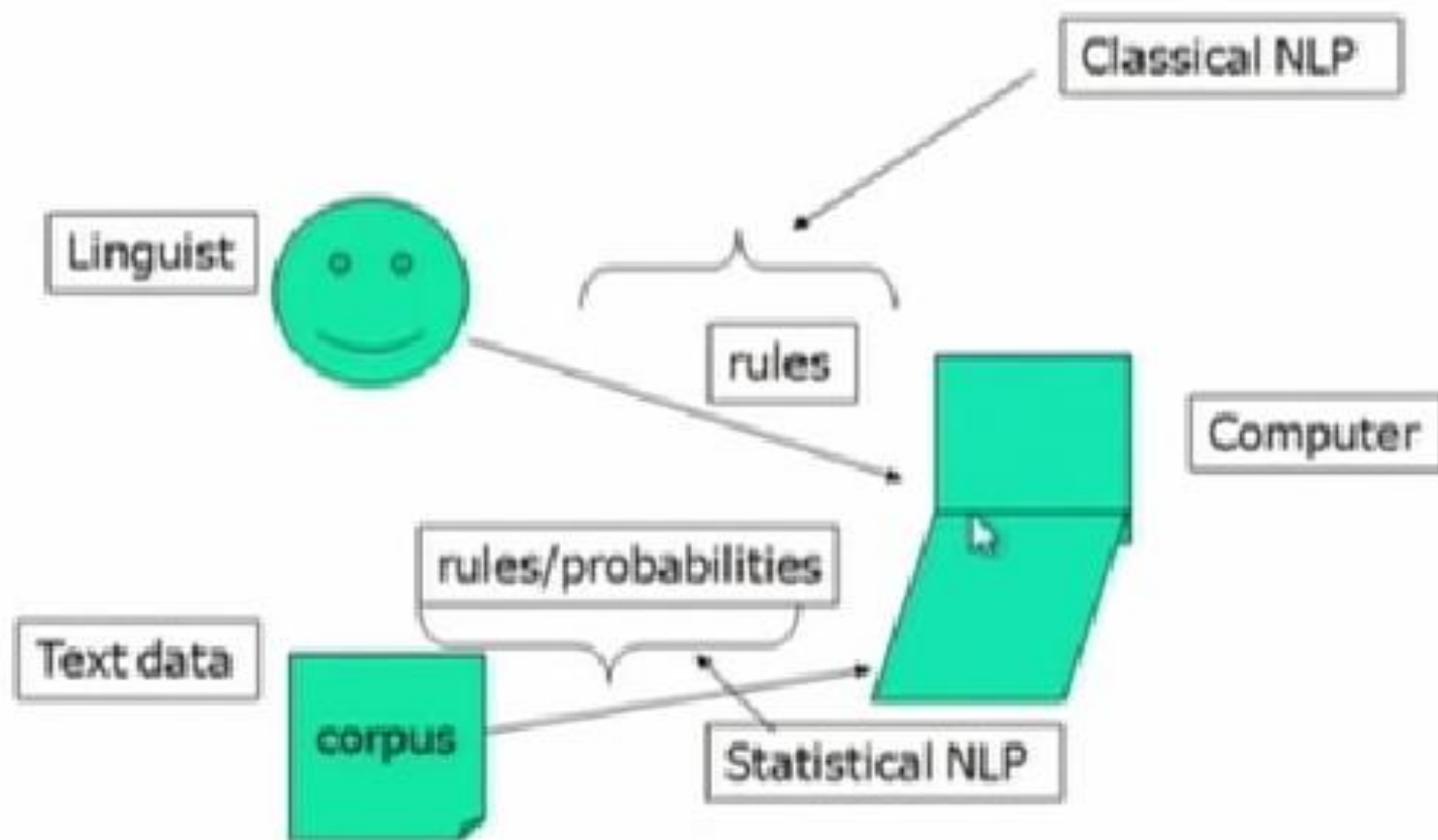
- How the immediately preceding sentences affects the interpretation of next sentence

7. World Knowledge

- Includes general language about the structure of the world that language users must have in order to maintain a conversation.

Classical NLP v/s Statistical NLP

Classical NLP and Statistical NLP



Classical NLP and Statistical NLP

- In classical NLP, the rules or knowledge comes from a Linguist.
- In statistical NLP, the rules and the probabilities are learned from the data or the corpus.
- The textual data provides the machine rules and probability values, by the application of some machine learning techniques.

Why NLP is hard?

Why NLP is hard?

- 1) Lexical Ambiguity
- 2) Structural Ambiguity
- 3) Language Imprecision and Vagueness
- 4) Ambiguity is pervasive
- 5) Ambiguity is explosive

Lexical Ambiguity

Example: Will Will will Will's will?

- i. The first **Will** is a modal verb (should/would/can)
- ii. The second **Will** is name of a person
- iii. The third **will** is a verb
- iv. The fourth **Will** is name of a person(same/diff)
- v. The fifth **will** is a noun

Lexical Ambiguity

Example:

Rose rose to put rose roes on her rows of roses

- i. First **rose** is name of person
- ii. Second **rose** is verb in past tense
- iii. Third **rose** is an adjective
- iv. **Roes** is a seafood
- v. Last **roses** is flowers noun

Roes and rose ambiguity whether roes is rose or
roes

Lexical Ambiguity

Example:

**Buffalo buffalo Buffalo buffalo buffalo buffalo
Buffalo buffalo**

Buffalo used in 3 senses:

- City in US
- Verb (bully)
- Animal

Structural Ambiguity

Different interpretations of the same sentence

Example 1:

The man saw the boy with the telescope

What is ambiguous :

- whether telescope is with the boy
- whether telescope is with the man

Structural Ambiguity

Example 2:

Hole found in the room wall; police are looking into

What is ambiguous: what are the police looking into-

- whether the police looking into the hole
- whether the police looking into the matter

Structural Ambiguity

Example 3:

Flying planes can be dangerous

What is ambiguous :

- whether flying is dangerous
- whether flying planes(together) is dangerous

Language Imprecision and Vagueness

- Lacking exactness and accuracy of detail, uncertain

Example 1:

It is very warm here.

Example 2:

Q: Did your mother call your aunt last night?

A: I'm sure she must have.

Ambiguity is pervasive

Example : I made her duck

What are **different meanings** of words:

Made -> cook / make / create

Duck -> animal (noun) / lower head(verb)

What are different interpretations:

Ambiguity is pervasive

Example : I made her duck

What are different interpretations for different meanings of a word: **made** - cook/make

- 1) I cooked a duck for her.
- 2) I cooked a duck belonging to her.
- 3) I made the duck she owns.
- 4) I made her lower her head.
- 5) I waved my magic card that turned her into a duck.

Ambiguity is pervasive

Example : I made her duck

Suppose, we go to phonetics

i) I aid her duck

ii) I'm eight or duck

Ambiguity is explosive

Example- Structural Ambiguity

I saw the man with the telescope

2 ways of parsing :

- PP attached to verb
- PP attached with noun

Number of parse trees generated - 02

Ambiguity is explosive

Example- Structural Ambiguity

I saw the man with the telescope

I saw the man with the telescope – 02 parses

I saw the man on the hill with the telescope – 5 parses

I saw the man on the will in Pune with the telescope –
14 parses

I saw the man on the will in Pune with the telescope at
noon – 42 parses

I saw the man on the will in Pune with the telescope at
noon on Sunday – 132 parses

Function words and Content words

Function words and Content words

Function words:

- Are the words used to make the sentence grammatical
- They have little lexical meaning
- Belongs to closed class category
- Used mainly for determining structure of the sentence

Examples:

Determiners , pronouns, prepositions, auxiliary verbs, conjunctions, articles, etc

Function words and Content words

Content words:

- Are the words used to convey what are the important concepts in the sentence
- They have strong lexical meaning
- They are nouns, verbs, adjectives, adverbs, etc
- Used mainly for determining topic of the sentence

Examples:

nouns, verbs, adjectives, adverbs, etc

Example

1. **The** winfy prunkilmonger **from** the glidgement mominkled **and** branified **all his** levesond rederously.
2. Glop **angry** investigator larm blonk **government harassed** gerfritz **infuriated** sutbor pumeog **listeners** thoroughly

Example

1. **The** winfy prunkilmonger **from** the glidgement mominkled **and** branified **all his** levesond rederously.
2. Glop **angry investigator** larm blonk **government harassed** gerfritz **infuriated** sutbor pumeog **listeners** thoroughly

Observe the two sentence, in which sentence structure of sentence is seen clearly and from which sentence we can understand topic.

Corpus with Word Distribution – Tom Sawyer

- Most of the words are function words
- One exception, word Tom, whose freq is high compared to other content words

Word	Freq	Use
the	3332	determiner
and	2972	conjunction
a	1790	determiner
to	1702	preposition
of	1445	preposition
was	1162	aux verb
..
Tom	450	noun
...

Type-Token Distinction

Type-Token Distinction

- **Type-token distinction** is a distinction that separates a concept from the objects which are particular instances of the concept.
 - Concept –is- type
 - Instances of concept – are- tokens
- **Type** is unique word and **tokens** are number of words of the type

Example: Will Will will

Type – 1

Tokens – 3

Type/Token Ratio (TTR)

Type/Token Ratio (TTR)

- **TTR** is the ratio of the number of different words(types) to the number of running words (tokens) in a given text or corpus
- This index indicates how often, on average, a new “wordform” appears in the text or corpus.
- If **TTR** is **high** means **lot of new words**, and if **TTR** is **low** means **no new words**

Comparison Across Text

- Mark Twains Tom Sawyer:
 - 71370 word tokens
 - 8018 Word types
 - $TTR = 0.112$
- Complete Shakespear work
 - 884647 word tokens
 - 29066 word types
 - $TTR = 0.032$

Finding: ??