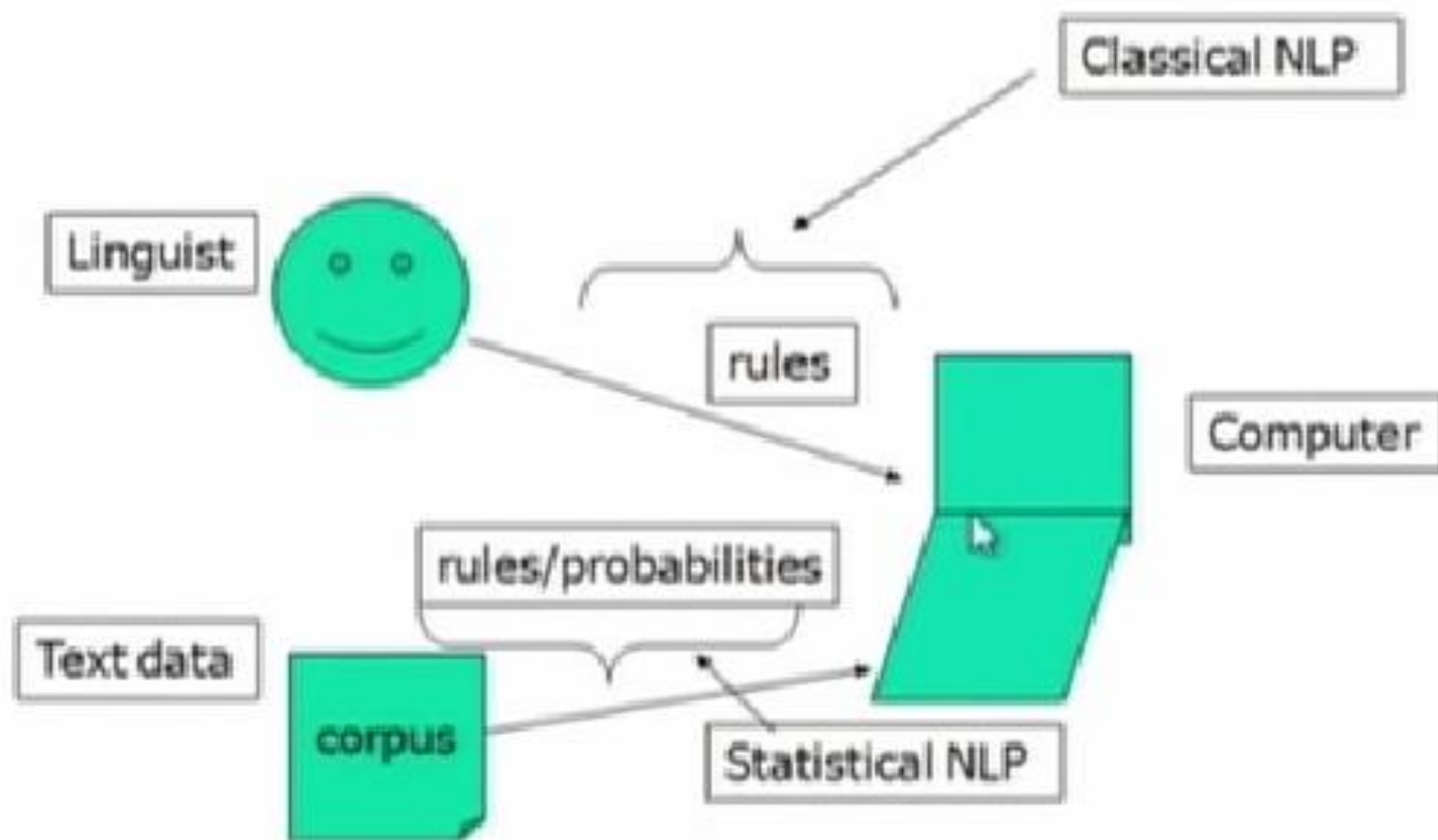


# **Classical NLP v/s Statistical NLP**

# 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: ??

# Sequence Labeling task

# Sequence Labeling task

## 1. Words

- POS tagging
  - placing label on smallest unit of the sentence.
- Named-Entity tagging
  - Detecting proper nouns and understanding their categories.
- Sense marking
  - Assigning words the sense label.

## 2. Phrases – chunking

## 3. Sentences – parsing

## 4. Paragraphs – co-reference annotation

# POS tagging

# What it is?

- POS Tagging is a process of selecting most likely sequence of syntactic categories for the words in a sentence.
- The set of tags is called the Tag-set.
- Standard Tag-set : **Penn Treebank (for English).**



# Definition

Tagging is the assignment of a single part-of-speech tag to each word (and punctuation marker) in a corpus.

## Example:

**<s> Come in August, and the COEP campus is abuzz with new and returning students. </s>**

## After POS tagging :

**<s> Come\_VB in\_IN August\_NNP,\_, and\_CC  
the\_DT COEP\_NNP campus\_NN is\_VBZ abuzz\_JJ  
with\_IN new\_JJ and\_CC returning\_VBG  
students\_NNS.</s>**

# PENN TAGSET

Number	Tag	Description	Number	Tag	Description
1.	CC	Coordinating conjunction	19.	PRP\$	Possessive pronoun
2.	CD	Cardinal number	20.	RB	Adverb
3.	DT	Determiner	21.	RBR	Adverb, comparative
4.	EX	Existential there	22.	RBS	Adverb, superlative
5.	FW	Foreign word	23.	RP	Particle
6.	IN	Preposition or subordinating conjunction	24.	SYM	Symbol
7.	JJ	Adjective	25.	TO	<i>to</i>
8.	JJR	Adjective, comparative	26.	UH	Interjection
9.	JJS	Adjective, superlative	27.	VB	Verb, base form
10.	LS	List item marker	28.	VBD	Verb, past tense
11.	MD	Modal	29.	VBG	Verb, gerund or present participle
12.	NN	Noun, singular or mass	30.	VBN	Verb, past participle
13.	NNS	Noun, plural	31.	VBP	Verb, non-3rd person singular present
14.	NNP	Proper noun, singular	32.	VBZ	Verb, 3rd person singular present
15.	NNPS	Proper noun, plural	33.	WDT	Wh-determiner
16.	PDT	Predeterminer	34.	WP	Wh-pronoun
17.	POS	Possessive ending	35.	WP\$	Possessive wh-pronoun
18.	PRP	Personal pronoun	36.	WRB	Wh-adverb

# Named-Entity tagging

# Named-Entity tagging

- Detecting proper nouns and understanding their categories.

## **Example1:**

Washington<sub>1</sub> voted Washington<sub>2</sub> to power.

# Named-Entity tagging

Example2:

पूजा ने पूजा के लिए फुल  
खरीदा

- First pooja is proper noun, **person name**
- Second pooja is **not a proper noun**, it is worshipping

# Named-Entity tagging

English translation:

**Pooja** bought flowers for **worshipping**

- If we do not detect the first pooja as Named-entity, proper noun, then the translation would be improper.

We would say:

- **Worshipping** bought flowers for **worshipping**

# Named-Entity tagging

English translation:

**Pooja** bought flowers for **worshipping**

- If we do not detect the first pooja as Named-entity, proper noun, then the translation would be improper.

We would say:

- **Worshipping** bought flowers for **worshipping**
- **Pooja** bought flowers for **pooja**.

# Applications of NER

- Machine translation
- Information extraction(IR)
- Summarization
- Question-Answering system



# Example of different stages of sequence labeling

## 1. Example of word labeling : POS tagging

<s>

**Come in August, and the COEP campus is abuzz with new and returning students.**

</s>

After POS tagging :

<s>

Come **\_VB** in **\_IN** August **\_NNP**, **\_**, and **\_CC** the **\_DT**  
COEP **\_NNP** campus **\_NN** is **\_VBZ** abuzz **\_JJ** with **\_IN** new **\_JJ**  
and **\_CC** returning **\_VBG** students **\_NNS**.

</s>

# Where are the labels coming from?

- From PENN Treebank TAGSET (Univ. of Pennsylvania)

Number	Tag	Description	Number	Tag	Description
1.	CC	Coordinating conjunction	19.	PRP\$	Possessive pronoun
2.	CD	Cardinal number	20.	RB	Adverb
3.	DT	Determiner	21.	RBR	Adverb, comparative
4.	EX	Existential there	22.	RBS	Adverb, superlative
5.	FW	Foreign word	23.	RP	Particle
6.	IN	Preposition or subordinating conjunction	24.	SYM	Symbol
7.	JJ	Adjective	25.	TO	<i>to</i>
8.	JJR	Adjective, comparative	26.	UH	Interjection
9.	JJS	Adjective, superlative	27.	VB	Verb, base form
10.	LS	List item marker	28.	VBD	Verb, past tense
11.	MD	Modal	29.	VBG	Verb, gerund or present participle
12.	NN	Noun, singular or mass	30.	VBN	Verb, past participle
13.	NNS	Noun, plural	31.	VBP	Verb, non-3rd person singular present
14.	NNP	Proper noun, singular	32.	VBZ	Verb, 3rd person singular present
15.	NNPS	Proper noun, plural	33.	WDT	Wh-determiner
16.	PDT	Predeterminer	34.	WP	Wh-pronoun
17.	POS	Possessive ending	35.	WP\$	Possessive wh-pronoun
18.	PRP	Personal pronoun	36.	WRB	Wh-adverb

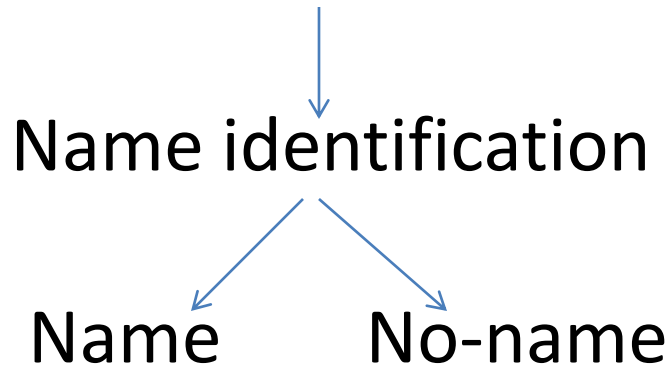
# Example of word labeling: NE tagging

<month-name> **August** </month-name>

<organization-name> **COEP** </organization-name>

# Proper-noun tagging: Name identification

Proper noun detection



**Eg, come in August**

Come – No-name

in – No-name

August – Name (NNP)

# Name Recognition

- Type of name identified
- Detect name to that level of categorization
  - August is month-name and it has time property associated with it.
  - COEP is organization-name and it has place property associated with it.

# Name Recognition

- Type of name identified
  - Detect name to that level of categorization
    - July is month-name and it has time property associated with it.
    - COEP is organization-name and it has place property associated with it.
  - **Why it is important ?**
    - < Come in August, and the COEP campus is abuzz with new and returning students.>
- Question asked: What is abuzz with students?

# Example of word labeling: sense marking

Word	Synset	WN-Synset-no
Come	{arrive, come, get}	01947900
..		
..		
..		
abuzz	{abuzz, buzzing, droning}	01859419

# Phrase labeling-chunking

Come in August, and **the COEP campus** is abuzz with **new and returning students** .

**<the COEP campus>** and **<new and returning students>** are noun phrases, and are called chunks.



# What is difference between chunk and phrases

- Chunks are somewhat convenient phrases.
- These phrases are non-recursive, that means, noun chunk will not contain noun-chunk within it.
- They are simple coherent units of text.

Examples:

- **the COEP campus**
- **new and returning students**

# Example of noun phrase not a chunk

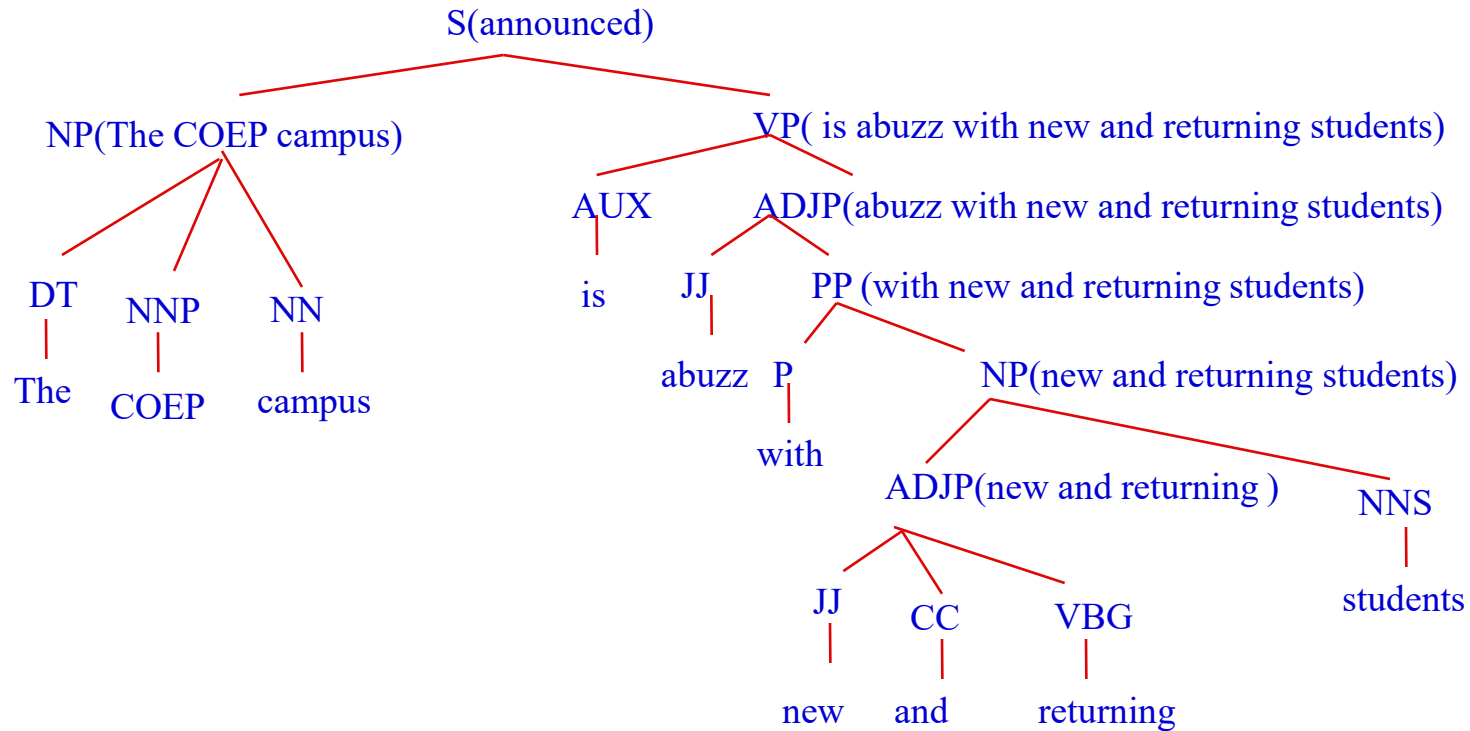
- The COEP campus situated in historic Pune city was visited by Dr.Kalam.

Phrases:

- The COEP campus
- Historic Pune city
- The COEP campus situated in historic Pune city

**Non-chunk phrase (Noun Phrase) :** The COEP campus situated in historic Pune city.

# Sentence labeling: Parsing



# Paragraph Labeling with reference resolution

- **Reference resolution:**
  - Which words/phrases refer to some word/phrase?
  - How are they related
- **Anaphora versus Cataphora**
  - Anaphora: an **anaphor** is a word/phrase that refers back to another phrase: the **antecedent** of the anaphor

Example: **Mary** thought that **she** lost **her** keys.




# Paragraph Labeling with reference resolution

- **Anaphora versus Cataphora**
  - Cataphora: an **cataphor** is a word/phrase that refers forward to another phrase: its **precedent**

Example:

She was at playground, when Riya realized that she lost her keys.



**Note:** Anaphora is often used as a synonym for reference resolution.

# Example

“I went with my friend to the bank to withdraw some money, but was disappointed to find it closed”

## 1. First layer of processing: POS tagging

I-pronoun went-verb with-pre my-possessive pro  
friend-noun to-pre the-art bank-Noun/verb to-pre  
withdraw-noun/verb some-adj money-noun,  
but-conjunction was-aux-verb  
disappointed-Ajd/verb to-pre find-noun/verb  
it-pronoun closed-adj/verb.

“I went with my friend to the bank to withdraw some money, but was disappointed to find it closed”

## 2. Second layer of processing: Sense Disambiguation

I-**pronoun**

went-**verb** , past tense -go(very ambiguous depending on senses of go)

### Noun

- **S:** (n) **go**, spell, tour, turn (a time period for working (after which you will be relieved by someone else)) *"it's my go"; "a spell of work"*
- **S:** (n) Adam, ecstasy, XTC, **go**, disco biscuit, cristal, X, hug drug (street names for methylenedioxymethamphetamine)
- **S:** (n) crack, fling, **go**, pass, whirl, offer (a usually brief attempt) *"he took a crack at it"; "I gave it a whirl"*
- **S:** (n) **go**, go game (a board game for two players who place counters on a grid; the object is to surround and so capture the opponent's counters)



# “go” meaning for verb category

## Verb

- S: (v) travel, **go**, move, locomote (change location; move, travel, or proceed, also metaphorically) *"How fast does your new car go?"*; *"We travelled from Rome to Naples by bus"*; *"The policemen went from door to door looking for the suspect"*; *"The soldiers moved towards the city in an attempt to take it before night fell"*; *"news travelled fast"*
- S: (v) **go**, proceed, move (follow a procedure or take a course) *"We should go farther in this matter"*; *"She went through a lot of trouble"*; *"go about the world in a certain manner"*; *"Messages must go through diplomatic channels"*
- S: (v) **go**, go away, depart (move away from a place into another direction) *"Go away before I start to cry"*; *"The train departs at noon"*
- S: (v) become, **go**, get (enter or assume a certain state or condition) *"He became annoyed when he heard the bad news"*; *"It must be getting more serious"*; *"her face went red with anger"*; *"She went into ecstasy"*; *"Get going!"*
- S: (v) **go** (be awarded; be allotted) *"The first prize goes to Mary"*; *"Her money went on clothes"*

“I went with my friend to the bank to withdraw some money, but was disappointed to find it closed”

## 2. Second layer of processing: Sense Disambiguation

... with-**pre** my-**possessive pro**

friend-**noun** to-**pre** the-**art**

bank-**Noun/verb**

- noun-financial institution/river bank

- verb- bank

to-**pre**

withdraw-**noun/verb**.

# Withdraw-verb category

## Verb

- **S: (v) withdraw, retreat, pull away, draw back, recede, pull back, retire, move back** (pull back or move away or backward) *"The enemy withdrew"; "The limo pulled away from the curb"*
- **S: (v) retire, withdraw** (withdraw from active participation) *"He retired from chess"*
- **S: (v) disengage, withdraw** (release from something that holds fast, connects, or entangles) *"I want to disengage myself from his influence"; "disengage the gears"*
- **S: (v) recall, call in, call back, withdraw** (cause to be returned) *"recall the defective auto tires"; "The manufacturer tried to call back the spoilt yoghurt"*
- **S: (v) swallow, take back, unsay, withdraw** (take back what one has said) *"He swallowed his words"*
- **S: (v) seclude, sequester, sequestrate, withdraw** (keep away from others) *"He sequestered himself in his study to write a book"*
- **S: (v) adjourn, withdraw, retire** (break from a meeting or gathering) *"We adjourned for lunch"; "The men retired to the library"*
- **S: (v) bow out, withdraw** (retire gracefully) *"He bowed out when he realized he could no longer handle the demands of the chairmanship"*
- **S: (v) withdraw, draw, take out, draw off** (remove (a commodity) from (a supply source)) *"She drew \$2,000 from the account"; "The doctors drew medical supplies from the hospital's emergency bank"*
- **S: (v) retire, withdraw** (lose interest) *"he retired from life when his wife died"*
- **S: (v) retreat, pull back, back out, back away, crawfish, crawfish out, pull in one's horns, withdraw** (make a retreat from an earlier commitment or activity) *"We'll have to crawfish out from meeting with him"; "He backed out of his earlier promise"; "The aggressive investment company pulled in its horns"*
- **S: (v) remove, take, take away, withdraw** (remove something concrete, as by lifting, pushing, or taking off, or remove something abstract) *"remove a threat"; "remove a wrapper"; "Remove the dirty dishes from the table"; "take the gun from your pocket"; "This machine withdraws heat from the environment"*

“I went with my friend to the bank to withdraw some money, but was disappointed to find it closed”

## 2. Second layer of processing: Sense Disambiguation

... some-**adj** money-**noun**,  
but-**conjunction** was-**aux-verb**  
disappointed-**Ajd/verb**

### Verb

- **S: (v)** disappoint, let down (fail to meet the hopes or expectations of) *"Her boyfriend let her down when he did not propose marriage"*

### Adjective

- **S: (adj)** defeated, **disappointed**, discomfited, foiled, frustrated, thwarted (disappointingly unsuccessful) *"disappointed expectations and thwarted ambitions"; "their foiled attempt to capture Calais"; "many frustrated poets end as pipe-smoking teachers"; "his best efforts were thwarted"*

“I went with my friend to the bank to withdraw some money, but was disappointed to find it closed”

## 2. Second layer of processing: Sense Disambiguation

...to-**pre** find-**noun/verb**

it-**pronoun** closed-**adj/verb**.

# Close –verb category

## Verb

- S: (v) close, shut (move so that an opening or passage is obstructed; make shut) *"Close the door"; "shut the window"*
- S: (v) close, shut (become closed) *"The windows closed with a loud bang"*
- S: (v) close up, close, fold, shut down, close down (cease to operate or cause to cease operating) *"The owners decided to move and to close the factory"; "My business closes every night at 8 P.M."; "close up the shop"*
- S: (v) close (finish or terminate (meetings, speeches, etc.)) *"The meeting was closed with a charge by the chairman of the board"*
- S: (v) conclude, close (come to a close) *"The concert closed with a nocturne by Chopin"*
- S: (v) close (complete a business deal, negotiation, or an agreement) *"We closed on the house on Friday"; "They closed the deal on the building"*
- S: (v) close (be priced or listed when trading stops) *"The stock market closed high this Friday"; "My new stocks closed at \$59 last night"*
- S: (v) close (engage at close quarters) *"close with the enemy"*
- S: (v) close (cause a window or an application to disappear on a computer desktop)
- S: (v) close (change one's body stance so that the forward shoulder and foot are closer to the intended point of impact)
- S: (v) close, come together (come together, as if in an embrace) *"Her arms closed around her long lost relative"*
- S: (v) close (draw near) *"The probe closed with the space station"*
- S: (v) close (bring together all the elements or parts of) *"Management closed ranks"*
- S: (v) close (bar access to) *"Due to the accident, the road had to be closed for several hours"*



# Close –Adjective category

## Adjective

S: (adj) **closed** (not open or affording passage or access) *"the many closed streets made travel difficult"; "our neighbors peeped from behind closed curtains"*

S: (adj) **closed** ((set theory) of an interval that contains both its endpoints)

S: (adj) shut, unopen, **closed** (not open) *"the door slammed shut"*

S: (adj) **closed**, shut (used especially of mouth or eyes) *"he sat quietly with closed eyes"; "his eyes were shut against the sunlight"*

S: (adj) **closed** (requiring union membership) *"a closed shop"*

S: (adj) **closed** (with shutters closed)

S: (adj) **closed** (not open to the general public) *"a closed meeting"*

S: (adj) **closed**, unsympathetic (not having an open mind) *"a closed mind unreceptive to new ideas"*

S: (adj) **closed**, closed in (blocked against entry) *"a closed porch"*

# Pronoun Drop

Pronoun drop for this sentence, but **I** was disappointed to find it closed”

Whether – I/ friend / money/bank was disappointed



# Scope Ambiguity

With is preposition .

What is scope of with ?

How much text it demands to complete with phrase.

“I went **with** my friend to the bank to withdraw some money, but was disappointed to find it closed”

# Co-referencing

It is pronoun. It refers to what?

“I went with my friend to the bank to withdraw some money, but was disappointed to find it closed”

POS	Bank (N/V)	closed (V/ adj)
Sense	Bank (financial institution)	withdraw (take away)
Pronoun drop	But I/friend/money/bank	was disappointed
SCOPE	With my friend	
Co-referencing	It -> bank	