



## NLP answers - important quetsions

Natural language Processing (University of Mumbai)



Scan to open on Studocu

Q1 ) What is POS Tagging ? Need for POS tagging ?

Ans :

Knowing whether a word is a possessive pronoun or a personal pronoun can tell us what words are likely to occur in its vicinity. This can be useful in a language model for speech recognition. Thus knowing the part-of-speech can produce more natural pronunciations in a speech synthesis system and more accuracy in a speech recognition system. They can also enhance an IR application by selecting out nouns or other important words from a document.

Q2) Types of POS tagging techniques ?

Ans :

Most of the tagging algorithms fall into one of the two classes. Rule based taggers and stochastic taggers [probabilistic].

1. **RULE BASED TAGGERS** : The earliest algorithm for automatically assigning POS were based on two stage architecture. The first stage used a dictionary to assign each word a list of potential parts-of-speech. The second stage used large lists of hand-written disambiguation rules to filter down this list to a single part-of-speech for each word. If the word has more than one possible tag, then rule-based taggers use hand-written rules to identify the correct tag. Disambiguation is done by analyzing the linguistic features of the word, its preceding word, its following word, and other aspects.
2. **STOCHASTIC / STATISTICAL TAGGERS** :  
Any model which somehow incorporates frequency or probability may be properly labelled stochastic. Stochastic taggers resolve the tagging ambiguities by using a training corpus to compute the probability of a given word having a given tag in a given context.i.e a text corpus is used to derive useful probabilities. Among the common models are n-gram model, Hidden Markov Model (HMM) and Maximum Entropy Model (MEM). The simplest stochastic tagger applies the following approaches for POS tagging – Word Frequency Approach In this approach, the stochastic taggers disambiguate the words based on the probability that a word occurs with a particular tag. We can also say that the tag encountered most frequently with the word in the training set is the one assigned to an ambiguous instance of that word. The main issue with this approach is that it may yield inadmissible sequence of tags. Tag Sequence Probabilities It is another approach of stochastic tagging, where the tagger calculates the probability of a given sequence of tags occurring. It is also called n-gram approach. It is called so because the best tag for a given word is determined by the probability at which it occurs with the n previous tags.
3. **Transformation based Tagger** :  
Transformation based tagger or Brill tagger shares the features of both the types of taggers. Rules are automatically induced from data. Thus, it's a combination of rule-based and stochastic methods. Tagging is done using broad rules and then improved or transformed by applying narrower rules. If we see similarity between rule-based and transformation tagger, then like rulebased, it is also based on the rules that specify what tags need to be assigned to what words. On the other hand, if we see similarity between stochastic and transformation tagger then like stochastic, it is machine learning technique in which rules are automatically induced from data.

Q3) What are the issues in HMM MODEL ?

Ans :

Hidden Markov Model (HMM) Tagger is a Stochastic POS Tagger. It is a probabilistic sequence model; i.e. given possible sequences of tags, a HMM Tagger will compute and assign the best sequence. It is based on Dynamic programming. There are many issues in HMM model :

**1) Unknown Words:**

- We might get some unknown words at runtime that aren't in the corpus.
- That word has no emission probability
- As a result, Viterbi decoding is not possible.
- We can use morphological signals to determine whether a word is a proper noun if the first character is capital.
- If the word has the suffix 'ed,' it could be past tense or past participle.

## 2. Limited Context:

- We use a first-order Markov model in Viterbi decoding, which acts as a bigram in state transitions.
- However, as seen in the example below, sometimes preceding background is insufficient.

3. For big sentences, go slowly.

4. Word features are limited.

Use a feature-rich classifier as a solution.

Q4) What is syntactic parsing and differentiate between its various approaches ?

Ans :

In syntactic analysis, a constituent is a word or a group of words that function as a single unit within a hierarchical structure. Constituency parsing aims to extract a constituency-based parse tree from a sentence that represents its syntactic structure according to a phrase structure grammar. The constituency parse tree is based on the formalism of context-free grammars. In this type of tree, the sentence is divided into constituents, that is, sub-phrases that belong to a specific category in the grammar.

There are 2 types of Parsing techniques present parsing, the first one is Top-down parsing and the second one is Bottom-up parsing.

Sr. No.	Key	Top Down Parsing	Bottom Up Parsing
1	Strategy	Top down approach starts evaluating the parse tree from the top and move downwards for parsing other nodes.	Bottom up approach starts evaluating the parse tree from the lowest level of the tree and move upwards for parsing the node.
2	Attempt	Top down parsing attempts to find the left most derivation for a given string.	Bottom up parsing attempts to reduce the input string to first symbol of the grammar.
3	Derivation Type	Top down parsing uses leftmost derivation.	Bottom up parsing uses the rightmost derivation.
4	Objective	Top down parsing searches for a production rule to be used to construct a string.	Bottom up parsing searches for a production rule to be used to reduce a string to get a starting symbol of grammar.

Top-Down Parsing	Bottom-Up Parsing
It is a parsing strategy that first looks at the highest level of the parse tree and works down the parse tree by using the rules of grammar.	It is a parsing strategy that first looks at the lowest level of the parse tree and works up the parse tree by using the rules of grammar.
Top-down parsing attempts to find the left most derivations for an input string.	Bottom-up parsing can be defined as an attempt to reduce the input string to the start symbol of a grammar.
In this parsing technique we start parsing from the top (start symbol of parse tree) to down (the leaf node of parse tree) in a top-down manner.	In this parsing technique we start parsing from the bottom (leaf node of the parse tree) to up (the start symbol of the parse tree) in a bottom-up manner.
This parsing technique uses Left Most Derivation.	This parsing technique uses Right Most Derivation.
The main leftmost decision is to select what production rule to use in order to construct the string.	The main decision is to select when to use a production rule to reduce the string to get the starting symbol.
Example: Recursive Descent parser.	Example: ItsShift Reduce parser.

Q5 ) Examples of Bottom up parsing algorithm ?

Q6 ) Examples of top down parsing algorithm ?

Q7 ) What is CYK algorithm ? Draw the parse tree for the given statement using cyk parsing algorithm ?

Ans :

One of the parsing algorithms. It uses a Bottom Up Approach for parsing

Simple and efficient parser. It is also sometimes called as chart parsing.

It works only with a grammar which is in CNF form [Chomsky Normal Form]

In CNF the grammar should have :

-A start symbol

-LHS and RHS

-LHS will have only non-terminals

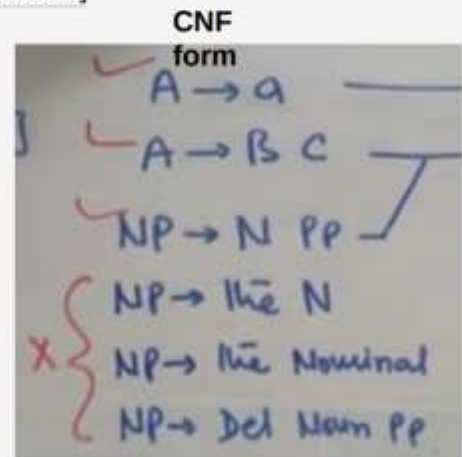
- RHS can contain

i] Single terminal

ii] Two non-terminal

iii] Combination of terminal and non-terminal is not permitted

iv] Single non-terminal is not permitted



Steps :

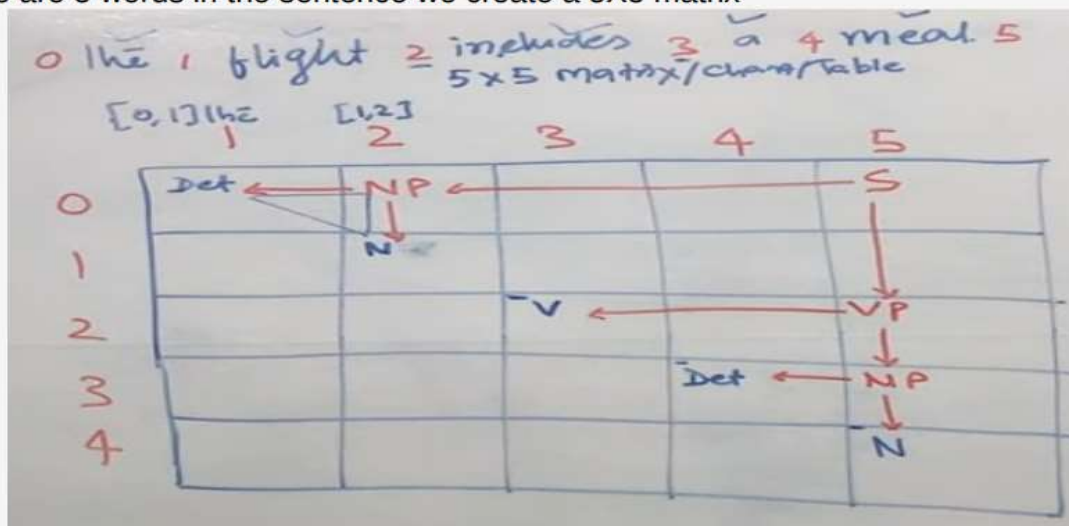
Convert the cfg grammar to a cnf if not in the format

### STEPS :

Write numbers between the words as shown

0 The 1 flight 2 includes 3 a 4 meal 5

Since there are 5 words in the sentence we create a 5X5 matrix



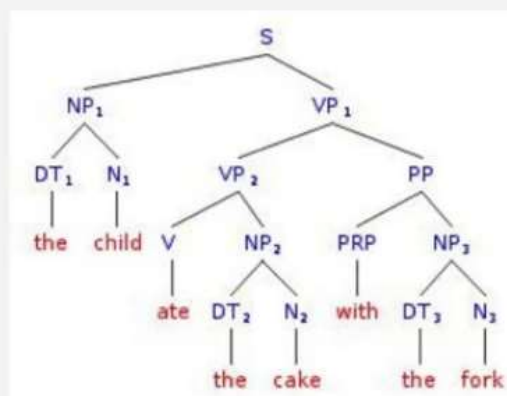
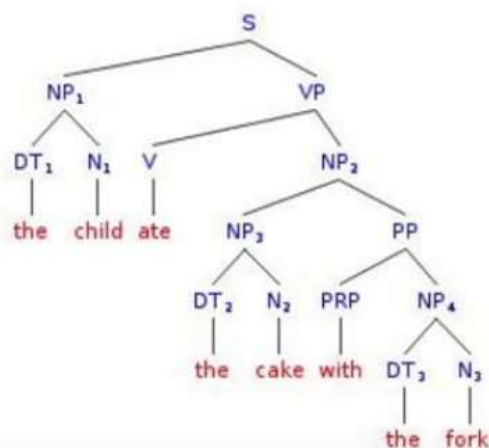
- Once reaching [0,0], backtrace all the non-terminals at [0,0] using the pointer values stored using recursion till all the leaves (non-terminals) are reached
- If [0,0] is blank, this means the sentence is grammatically/syntactically incorrect. No parse tree exists
- If we get multiple parse trees, the sentence is ambiguous (from which many meanings can be drawn, as shown in the below example)

## Example

["the", "child", "ate", "the", "cake", "with", "the", "fork"]

S → NP VP  
 NP → DT N | NP PP  
 PP → PRP NP  
 VP → V NP | VP PP  
 DT → 'a' | 'the'  
 N → 'child' | 'cake' | 'fork'  
 PRP → 'with' | 'to'  
 V → 'saw' | 'ate'

	0	1	2	3	4	5	6	7	
the	DT	NP						S	1
child	N								2
ate	V			VP				VP	3
the	DT	NP						NP	4
cake	N								5
with	PRP							PP	6
the	DT	NP							7
fork	N								8



**Which is Correct?**

**CYK can't handle ambiguity in sentences i.e can't perform disambiguation.**

CYK can't handle ambiguity in sentences i.e can't perform disambiguation.

- The CKY parsing algorithm can represent these ambiguities in an efficient way but is not equipped to resolve them.
- A probabilistic parser offers a solution to the problem: compute the probability of each interpretation and choose the most probable interpretation.

Q) What is Semantic Analysis ?

Ans :

Semantic Analysis is a subfield of Natural Language Processing (NLP) that attempts to understand the meaning of Natural Language. Semantic Analysis of Natural Language captures the meaning of the given text while taking into account context, logical structuring of sentences and grammar roles. The purpose of semantic analysis is to draw exact meaning, or you can say dictionary meaning from the text. The work of semantic analyzer is to check the text for meaningfulness.

Semantic Analysis of Natural Language can be classified into two broad parts: 1. Lexical Semantic Analysis: Lexical Semantic Analysis involves understanding the meaning of each word of the text individually. It basically refers to fetching the dictionary meaning that a word in the text is deputed to carry. 2. Compositional Semantics Analysis: Although knowing the meaning of each word of the text is essential, it is not sufficient to completely understand the meaning of the text. So in this second part, the individual words will be combined to provide meaning in sentences. For example, consider the following two sentences: Sentence 1 : Students like College. Sentence 2 : College likes Students. There are some of the ways of meaning representation.

- 1) First Order Logic.
- 2) Abstract Meaning Representation using a directed graph
- 3) Abstract Meaning Representation using the textual form.
- 4) Frame based or slot filter representation .



Q8) What are different relations between lexeme ?

Ans :

### 3. Relations between Lexemes: Homonymy, Polysemy, Synonymy, Hyponymy

One way to approach lexical semantics is to study the relationship among lexemes (an abstract representation of a "word", the lexical entry in a dictionary). Semantics of a lexeme can be understood by analysing the relationship of lexemes with other lexemes. Lexical semantics information is useful for wide variety of NLP applications. This section discusses a variety of relationship that holds among lexemes and their senses.

#### Homonymy

The first relationship that we discuss is homonymy which is perhaps the simplest relationship that exists among lexemes. Homonyms are words that have the same form but have different, unrelated meanings. A classic example of homonymy is Bank (river bank or financial institution). A related idea is that of homophones that refers to words that are pronounced in the same way but different meaning or spelling of both (e.g., be and bee, bear and bare).

#### Polysemy

Many words have more than one meaning or sense. Unlike homonyms, polysemes are words with related meanings. This linguistic phenomenon is called polysemy or lexical ambiguity. Words that have several senses are ambiguous and called polysemous. For example, the word "chair" can refer to a piece of furniture, a person, the act of presiding over a discussion etc. The word "employ" is a polysemy as its two meanings - to hire (employ a person) and to accept (employ an idea) are related. In a particular use, only one of these meanings is correct.

#### Hyponymy

The hypernym is a word with the more general sense. The word automobile is a hypernym for a car and a truck. The hyponym is a word with the most specific meaning. In the relationship between car and automobile, car is a hyponym of automobile. Antonymy is a semantic relationship that holds between words that express opposite meanings. The word Good is an antonym of Bad, and White is an antonym of Black.

#### Synonymy

The word synonym defines the relationship between different words that have a similar meaning. A simple way to decide whether two words are synonymous is to check for substitutability. Two words are synonyms in a context if they can be substituted for each other without changing the meaning of the sentence.

These relationships are useful in organising words in lexical databases one widely known lexical database is WordNet discussed in next topic.

Q13 ) Explain different lexical databases : Wordnet and Bablenet ?

Ans :

A lexical database is a lexical resource which has an associated software environment database which permits access to its contents. The database may be custom-designed for the lexical information or a general-purpose database into which lexical information has been entered.

EXAMPLES OF LEXICAL DATABASES : i) Wordnet ii) Babelnet

i) Wordnet :

WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the browser WordNet is also freely and publicly available for download. WordNet's structure

makes it a useful tool for computational linguistics and natural language processing. WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity.

ii) Babelnet :

WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the browser WordNet is also freely and publicly available for download. WordNet's structure makes it a useful tool for computational linguistics and natural language processing. WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity.

Q11 ) What is Word Sense Disambiguation.

We understand that words have different meanings based on the context of its usage in the sentence. If we talk about human languages, then they are ambiguous too because many words can be interpreted in multiple ways depending upon the context of their occurrence.

Word sense disambiguation, in natural language processing (NLP), may be defined as the ability to determine which meaning of word is activated by the use of word in a particular context. Lexical ambiguity, syntactic or semantic, is one of the very first problem that any NLP system faces. Part-of-speech (POS) taggers with high level of accuracy can solve Word's syntactic ambiguity. On the other hand, the problem of resolving semantic ambiguity is called WSD (word sense disambiguation). Resolving semantic ambiguity is harder than resolving syntactic ambiguity.

For example, consider the two examples of the distinct sense that exist for the word "**bass**" –

- ▣ I can hear bass sound.
- ▣ He likes to eat grilled bass.



The occurrence of the word **bass** clearly denotes the distinct meaning. In first sentence, it means **frequency** and in second, it means **fish**. Hence, if it would be disambiguated by WSD then the correct meaning to the above sentences can be assigned as follows –

- I can hear bass/frequency sound.
- He likes to eat grilled bass/fish.

Q12 ) Approaches of WSD ?

Ans :

There are four main ways to implement WSD.

1) Dictionary-and knowledge based methods :

As the name suggests, for disambiguation, these methods primarily rely on dictionaries, treasures and lexical knowledge base. The Lesk method (Lesk 1986) is the seminal dictionary-based method. It is based on the hypothesis that words used together in text are related to each other and that the relation can be observed in the definitions of the words and their senses. Two (or more) words are disambiguated by finding the pair of dictionary senses with the greatest word overlap in their dictionary definitions. For example, when disambiguating the words in pine cone, the definitions of the appropriate senses both include the words evergreen and tree (at least in one dictionary). An alternative to the use of the definitions is to consider general word-sense relatedness and to compute the semantic similarity of each pair of word senses based on a given lexical knowledgebase such as WordNet. Graph-based methods reminiscent of spreading-activation research of the early days of AI research have been applied with some success. The use of selectional preferences (or selectional restrictions) are also useful. For example, knowing that one typically cooks food, one can disambiguate the word bass in I am cooking bass (i.e., it's not a musical instrument)

2) Supervised methods :

Supervised methods are based on the assumption that the context can provide enough evidence on its own to disambiguate words (hence, world knowledge and reasoning are deemed unnecessary). The context is represented as a set of “features” of the words. It includes the information about the surrounding words also. Probably every machine learning algorithm going has been applied to WSD, including associated techniques such as feature selection, parameter optimization, and ensemble learning. Support vector machines and memory-based learning have been shown to be the most successful approaches, to date, probably because they can cope with the highdimensionality of the feature space. However, these supervised methods are subject to a new knowledge acquisition bottleneck since they rely on substantial amounts of manually sense-tagged corpora for training, which are laborious and expensive to create.

3) Semi-supervised methods

Due to the lack of training corpus, most of the word sense disambiguation algorithms use semisupervised learning methods. It is because semi-supervised methods use both labelled as well as unlabeled data. These methods require very small amount of annotated text and

large amount of plain unannotated text. The technique that is used by semisupervised methods is bootstrapping from seed data.

#### 4) Unsupervised methods :

Unsupervised Methods pose the greatest challenge to researchers and NLP professionals. A key assumption of these models is that similar meanings and senses occur in a similar context and thus senses can be induced from text by clustering word occurrences using some measure of similarity of context. Then, new occurrences of the word can be classified into the closest induced clusters/senses. It is hoped that unsupervised learning will overcome the knowledge acquisition bottleneck because they are not dependent on manual effort.

Q13) Explain Pragmatics and Discourse Analysis with Example ?

Ans :

□ Pragmatics is the study of meaning in relation to the context in which a person is speaking or writing. This includes social, situational and textual context. It also includes background knowledge context; that is, what people know about each other and about the world. Pragmatics assumes that when people communicate with each other they normally follow cooperative principle; that is, they have a shared understanding of how they should cooperate in their communications. The ways in which people do this, however, varies across cultures. What may be a culturally appropriate way of saying or doing something in one culture may not be the same in another culture. The study of this use of language across cultures is called cross-cultural pragmatics.

dog's breakfast : something that is messy or poorly done [Australian] It was a dog's breakfast of a match, and our coach was understandably upset.

Piece of cake : an easy task [British] The test was a piece of cake for him as he studied well

Bandar Kya Jaane Adrak Ka Swaad [Indian]

## EXAMPLES

1. Will you crack open the door? I am getting hot.

Semantically, the word "crack" would mean to break, but pragmatically we know that the speaker means to open the door just a little to let in some air.

2. I heart you!

Semantically, "heart" refers to an organ in our body that pumps blood and keeps us alive. However, pragmatically, "heart" in this sentence means "love". Hearts are commonly used as a symbol for love, and to "heart" someone has come to mean that you love someone.

**Discourse analysis** is another one of the applications of Natural Language Processing. Discourse analysis may be defined as the process of determining contextual information that is useful for performing other tasks, such as **anaphora resolution (AR)**

coherent discourse must possess the following properties –

### **Coherence relation between utterances**

The discourse would be coherent if it has meaningful connections between its utterances. This property is called coherence relation. For example, some sort of explanation must be there to justify the connection between utterances.

### **Relationship between entities**

Another property that makes a discourse coherent is that there must be a certain kind of relationship with the entities. Such kind of coherence is called entity-based coherence.

Sam bought a book for Mark. It was blue in color.

Relation between the entity ---color is for the book

Q) What is reference and reference resolution ?

- **Reference**

- Reference is a means to link a referring expression to another referring expression in the surrounding text
  - Eg: Suha bought a printer. It costs her Rs. 20,000.
  - Her refers to Suha and it refers to printer
  - Also called anaphoric reference – reference to something before

### **Reference Resolution :**

The process of identifying what pronoun refers to what entities is called as Reference Resolution

# Reference Resolution

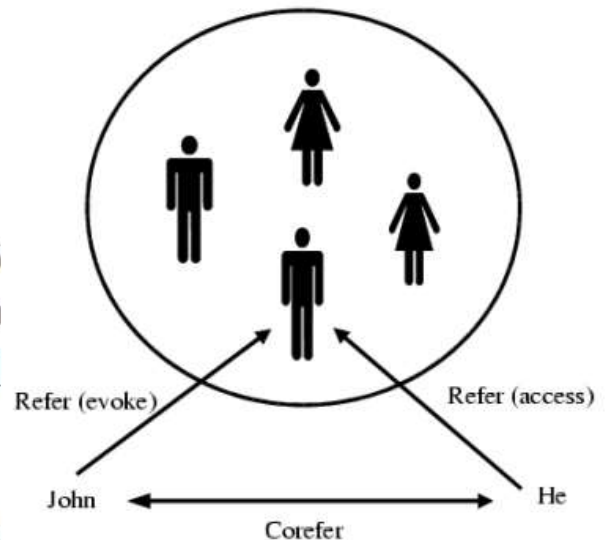
- Two fundamental operations to the discourse model

## – Evoke

- When a referent is first mentioned in a discourse, we say that a representation for it is **evoked** into the model

## – Access

- Upon subsequent mention, this representation is **accessed** from the model.



20

Q15) Mention types of reference.

## Anaphora Reference

**Anaphora** is the use of an expression whose interpretation depends upon another expression in context (its antecedent)

**The music** (antecedent) was so loud that **it** (Anaphor) couldn't be enjoyed.

Note: When the referring expression is pointing forward then it is Cataphora

## Types of References

- **Barack Obama** , the **U.S** former president **tweeted** his **New year** wishes.
- **The music** was so loud that **it** couldn't be enjoyed. (Anaphora)
- Despite **her** difficulty , **Swathi** went ahead to help him. (Cataphora)

Q) Explain types of referring expressions with example ?

### TYPES OF REFERRING EXPRESSIONS

- Five types of referring expression
  - *Indefinite NPs*
  - *Definite NPs*
  - *Pronouns,*
  - *Demonstratives, and*
  - *One-anaphora*
- Three types of referents that complicate the reference resolution problem
  - *Inferrables*
  - *Discountinuous sets, and*
  - *Generics*



## Indefinite Noun Phrases :

---

The definite article (the) is used before a noun to indicate that the identity of the noun is known to the reader. The indefinite article (a, an) is used before a noun that is general or when its identity is not known.

Introducing entities new to the hearer into the discourse context

I saw **an** Acura Integra today. (evoke)

**Some** Acura Integra were being unloaded at the local dealership today.

I saw **this** awesome Acura Integra today.

I am going to the dealership to buy **an** Acura Integra today. (specific/non-specific ambiguity)

## Definite NPs

---

Refer to an entity that is identifiable to the hearer, either because

- it has already been mentioned in the discourse context,
- it is contained in the hearer's set of beliefs about the world, or
- the uniqueness of the objects is implied by the description itself.

**I saw an Acura Integra today. The Integra was white and needed to be washed.** (context)

**The Indianapolis 500** is the most popular car race in the US. (belief)

**The faster car in the Indianapolis 500** was an Integra. (uniqueness)

Another example, in the sentence - I used to read The Times of India – The Times of India is a definite reference.

## Pronoun

---

It is a form of definite reference. For example, Sam laughed as loud as he could. The word he represents pronoun referring expression.

Pronouns usually refer back no further than one or two sentences back in the ongoing discourse, whereas definite NPs can often refer further back

Pronouns can also participate in **cataphora**, in which they are mentioned before there referents are.

Before *he* bought *it*, John checked over the Integra very carefully.

## Demonstratives

---

Demonstrative pronouns, like *this* and *that*, can appear either alone or as determiner, for instance *this Acura*, *that Acura*.

The choice between two demonstratives is generally associated with some notion of spatial proximity:

*This* indicates closeness and *that* signaling distance

John shows Bob an Acura Integra and a Mazda Miata

Bob (pointing): I like *this* better than *that*.

I bought an Integra yesterday. It's similar to the one I bought five years ago.

*That one* was really nice, but I like *this one* even better.

Further part given in the ppt. at last. Module 5