Module 5

Pragmatics

Content

5.1. Reference resolution

- 5.1.1 Discourse model
- 5.1.2 Reference Phenomenon
- 5.1.3 Syntactic and Semantic Constraints on co reference

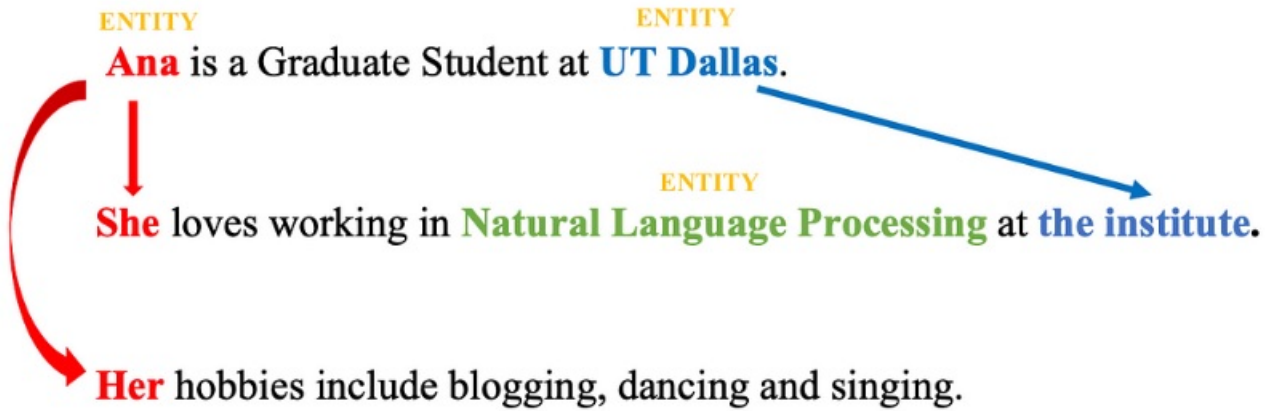
Pragmatic Analysis

- Pragmatic analysis (PA) deals with outside word knowledge, which means understanding the context that is external to documents and queries.
- PA that focuses on what was described is reinterpreted by what it actually meant, deriving the various aspects of language that require real-world knowledge.
- It deals with overall communicative and social content and its effect on interpretation. It means abstracting the meaningful use of language in situations. In this analysis, the main focus always on what was said is reinterpreted on what is intended. It helps users to discover this intended effect by applying a set of rules that characterize cooperative dialogues.
- "Pruning a tree is a long process"
- Here, pruning a tree is one of the concepts of computer science algorithm techniques. So, the word **pruning** is not related to cutting the actual physical tree, we are talking about computer science algorithm. This is an ambiguous situation; how to deal with these kinds of ambiguous situations is also an open area of research.

Discourse model

- The word ”discourse” in linguistic terms means language in use.
- Discourse in NLP is nothing but coherent groups of sentences.
- When we deal with Natural Language Processing, the provided language consists of structured, collective, and consistent groups of sentences which are termed discourse in NLP.
- Discourse Analysis is extracting the meaning out of the corpus or text.
- Discourse analysis may be defined as the process of performing text or language analysis, which involves text interpretation and knowing the social interactions.
- Very often, the interpretation of a sentence in a discourse depends on what preceded it. A clear example of this comes from anaphoric pronouns, such as he, she, and it. Given a discourse such as Angus used to have a dog. But he recently disappeared., you will probably interpret he as referring to Angus’s dog. However, in Angus used to have a dog. He took him for walks in New Town., you are more likely to interpret he as referring to Angus himself.

Example of Discourse



Here, “Ana”, “Natural Language Processing” and “UT Dallas” are plaussible entities.
“She” and “Her” are references to the entity “Ana” and “the institute” is a reference to the entity “UT Dallas”.

Reference Resolution

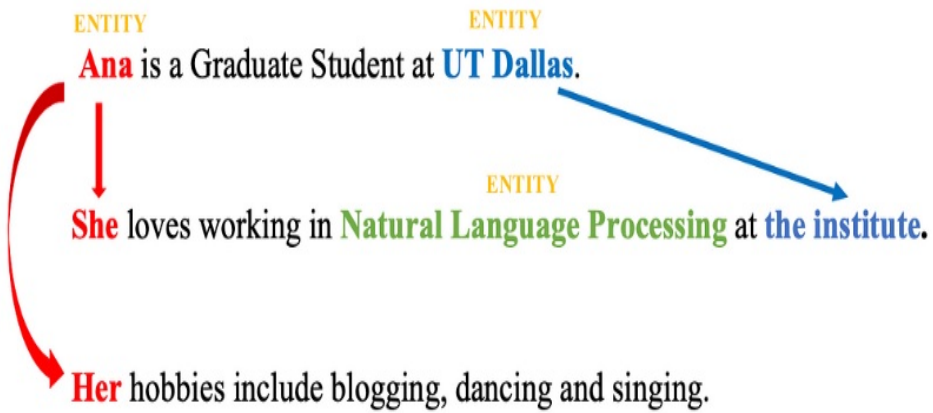
- Understanding depends on the context
- Referring expressions: Example.. it, that, those…
- Word Sense – ‘Bank’ (River Bank) or ‘Bank’ (Financial organization)
- Intention of speaker : Do you have the time?
- Reference, in NLP, is a linguistic process where one word in a sentence or discourse may refer to another word or entity. The task of resolving such references is known as Reference Resolution.
- In the previous example (Example of Ana) , “She” and “Her” referring to the entity “Ana” and “the institute” referring to the entity “UT Dallas” are two examples of Reference Resolution.

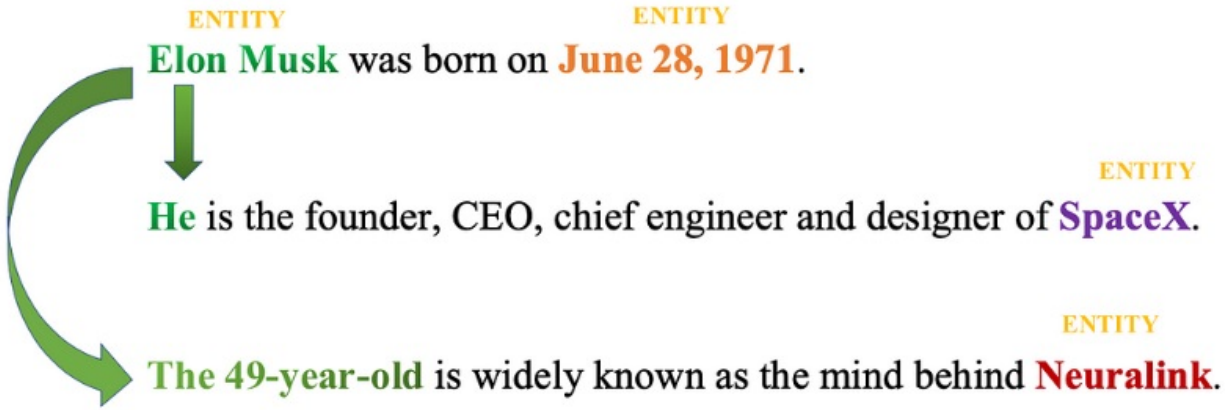
Lets summarize

Discourse in the context of NLP refers to a sequence of sentences occurring one after the other

Reference is a linguistic process where one word in a sentence or discourse refers to another word or entity.

The task of resolving such references is known as **Reference Resolution.**


ENTITY
ENTITY
Ana is a Graduate Student at UT Dallas.
ENTITY
She loves working in Natural Language Processing at the institute.
Her hobbies include blogging, dancing and singing.



> ***Referring Expressions:*** Elon Musk, He, The 49 year old

> ***Referent:*** *Elon Musk*

> ***Corefering Expressions:*** {Elon Musk, He}, {Elon Musk, The 49 year old}

- In order to understand Reference resolution, we need to learn few terminologies:

Referent is the object that is being referred to. For example, “Ana” is the referent in the above example.

Referring expression are the mentions or linguistic expressions given in

Corefer: Two or more referring expressions that refer to the same discourse entity are said to corefer the discourse.

Antecedent: There is also a name for a referring expression that grants permission for the use of another, similar to how mentioning the name ‘Ana’ permits the person to be afterwards referred to by the pronoun ‘she’. We call the name ‘Ana’ the antecedent of ‘she’.

Anaphora and Anaphoric:

Reference to an entity that had already been introduced in the discourse is called anaphora (e.g Ana) and the referring expression used is called anaphoric (pronouns ‘she’ and ‘her’ are anaphoric)

Reference Phenomena

Question: explain different types of Referring expressions? What is reference phenomena?

The range of referential phenomena offered by natural language is quite extensive. Five types of referring expression have been discussed here.

- Indefinite noun phrases
- Definite noun phrases
- Pronouns
- Demonstratives
- Names

Types of Referring Expressions

- The five types of referring expressions are described below −

Indefinite Noun Phrases

- Such kind of reference represents the entities that are new to the hearer into the discourse context. Indefinite references bring unfamiliar entities to the conversation environment.
- The determiner ‘a’ and ‘an’ is used most commonly to indicate indefinite references.
- Example:
- *I saw **an** Acura Integra Today – an is an indefinite Noun*
- *Ram had gone around one day to bring him **some** food − some is an indefinite reference.*

Definite Noun Phrases

- Opposite to above, such kind of reference represents the entities that are not new or identifiable to the hearer into the discourse context. An entity that is identifiable to the hearer is referred to as having ‘ definite reference’.
- For example
- - it has already been mentioned in the discourse context.
- Is a part of the hearer’s worldview
- The description itself implies the object’s uniqueness
- For example, in the sentence –
- I saw an Acura Integra. **The Integra** was white in color.
- I used to read **The Times of India** – The Times of India is a definite reference.

.

- **Pronouns**
- Pronominalization is another type of definite reference.
- For example, Ram laughed as loud as he could. The word **he** represents pronoun referring expression.
- Pronominal reference has more restrictions than complete definite noun phrases, requiring a high level of activation for the referent in the discourse model. While definite noun phrases can frequently refer back, pronouns typically refer to entities that were introduced no more than one or two sentences before to the current discourse. This can be illustrated using the following two examples:
- *John went to the Bob's party , and parked next to a beautiful Acura Integra.*
- *He went insode and talked to Bob for more than an hour.*
- *He also said that he bought '**it**' yesterday.*
- *He also said that he bought '**the Acura**' yesterday.*

The integra no longer has the level of salience necessary to permit pronominal reference by the time we get to the last sentence.

Cataphora:

- In cataphora pronouns are mentioned before their referents rae.
- Example:

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

- In the above example, the pronouns 'he' and 'it' both appear before the introduction of their respective referent.

Demonstratives

- These demonstrate and behave differently than simple definite pronouns ('it'). For example, 'this' and 'that' are demonstrative pronouns. For example 'this Acura' and 'that Acura'. The decision between two demonstratives – 'this' signaling proximity and 'that' signaling distance – is typically connected to some idea of geographical proximity. Depending upon the situational settings of the conversation, spatial distance might be calculated.
- Example:
- *John shows Bob an Acura Integra and a Mazda Miata*
- *Bob (Pointing) : I like 'this' better than 'that'.*
- Alternatively, distance can be metaphorically interpreted in terms of conceptual relations.
- "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". Here, that one refers to the Acura bought five years ago (greater temporal distance), whereas this one refers to the one bought yesterday (closer temporal distance)

Names

- It is the simplest type of referring expression. It can be the name of a person, organization and location also. For example, in the above examples, **Ram** is the name-refereeing expression.

Types of referents which complicate the reference resolution

- There are a few referent kinds that make the reference resolution difficult.

1) Inferrables

2) Discontinuous sets

3) Generics

Inferrables:

- First, we consider cases in which a referring expression does not refer to an entity that has been explicitly evoked in the text, but instead one that is inferentially related to an evoked entity. Such referents are called **inferrables.**
- Consider the expressions **a door** and **the engine** in following sentence.
- Sentence: "*I almost bought an Acura Integra today, but a door had a dent and the engine seemed noisy*".
- The indefinite noun phrase a door would normally introduce a new door into the discourse context, but in this case the hearer is to infer something more: that it is not just any door, but one of the doors of the Integra.
- Similarly, the use of the definite noun phrase the engine normally presumes that an engine has been previously evoked or is otherwise uniquely identifiable.
- Here, no engine has been explicitly mentioned, but the hearer infers that the referent is the engine of the previously mentioned Integra.

Discontinous Set

- In some cases, references using plural referring expressions like ‘they’ and ‘them’ refer to sets of entities that are evoked together, for instance, using another plural expression (their Acuras) or a conjoined noun phrase (John and Mary):
- Sentence: “*John and Mary love their Acuras*. *They drive them all the time*”
- However, plural references may also refer to sets of entities that have been evoked by discontinuous phrases in the text:
- Sentence: ”*John has an Acura, and Mary has a Mazda. They drive them all the time*”. Here, they refers to John and Mary, and likewise them refers to the Acura and the Mazda.
- Note, also that the second sentence in this case will generally receive what is called a pairwise or respectively reading, in which John drives the Acura and Mary drives the Mazda, as opposed to the reading in which they both drive both cars.

Generics:

- Making the reference problem even more complicated is the existence of generic reference.
- Consider following example
- "*I saw no less than 6 Acura Integras today*. They are the coolest cars". Here, the most natural reading is not the one in which they refers to the particular 6 Integras mentioned in the first sentence, but instead to the class of Integras in general.

Syntactic and Semantic constraints on co-reference

- Having described a variety of reference phenomena that are found in natural language, we can now consider how one might develop algorithms for identifying the referents of referential expressions.
- One step that needs to be taken in any successful reference resolution algorithm is to filter the set of possible referents on the basis of certain relatively hard-and-fast constraints.
- We describe some of these constraints here.
    1. Number Agreement
    2. Person and Case Agreement
    3. Gender Agreement

Number Agreement

- Number Agreement Referring expressions and their referents must agree in number; for English, this means distinguishing between **singular** and **plural** references. A categorization of pronouns with respect to number is shown in following table.

Singular	Plural	Unspecified
She, her, he, him, his, it	We, us, they, them	your

- The following examples illustrate constraints on number agreement.

1. John has a new Acura. It is red.
2. John has three new Acuras. They are red.
3. John has a new Acura. They are red.
4. John has three new Acuras. It is red.

Person and Case Agreement

- English distinguishes between three forms of person : first, second, and third.
- A categorization of pronouns with respect to person is shown in following table.
- The following examples illustrate constraints on person agreement.

  Example: **You and I** have Acuras. **We** love them.

	First	Second	Third
Nominative	I, we	You	He, she, they
Accusative	Me, us	You	Him, her, them
Genitive	My, our	Your	His, her, their

Examples:

John and Mary have Acuras. They love them.
John and Mary have Acuras. We love them. (where We=John and Mary)
You and I have Acuras. They love them. (where They=You and I)

In addition, English pronouns are constrained by case agreement; different forms of the pronoun may be required when placed in subject position (nominative case, e.g., he, she, they), object position (accusative case, e.g., him, her, them), and genitive position (genitive case, e.g., his Acura, her Acura, their Acura).

Gender Agreement

- Gender Agreement Referents also must agree with the gender specified by the referring expression.
- English third person pronouns distinguish between male, female, and nonpersonal genders, and unlike some languages, the first two only apply to animate entities.
- Some examples are shown in table

Masculine	Feminine	Nonpersonal
He, him, his	She, her	It

The following examples illustrate constraints on gender agreement.

John has an Acura. He is attractive. (he=John, not the Acura)
John has an Acura. It is attractive. (it=the Acura, not John)

Syntactic Constraints Reference:

- Relations may also be constrained by the syntactic relationships between a referential expression and a possible antecedent noun phrase when both occur in the same sentence.
- For instance, the pronouns in all of the following sentences are subject to the constraints indicated in brackets.

    1. John bought himself a new Acura. [himself=John]
    2. John bought him a new Acura. [him ≠John]
    3. John said that Bill bought him a new Acura. [him ≠ Bill]
    4. John said that Bill bought himself a new Acura. [himself=Bill]
    5. He said that he bought John a new Acura. [He ≠ John; he ≠John]

- English pronouns such as himself, herself, and themselves are called **REFLEXIVES**.
- A reflexive co-refers with the subject of the most immediate clause that contains it, whereas a nonreflexive cannot corefer with this subject.

Selectional Restrictions

- The selectional restrictions that a verb places on its arguments may be responsible for eliminating referents.
- *John parked his Acura in the garage. He had driven it around for hours.*
- There are two possible referents for **it**, the Acura and the garage. The verb drive, however, requires that its direct object denote something that can be driven, such as a car, truck, or bus, but not a garage.
- Thus, the fact that the pronoun appears as the object of drive restricts the set of possible referents to the Acura.
- It is conceivable that a practical NLP system would include a reasonably comprehensive set of selectional constraints for the verbs in its lexicon.
- Selectional restrictions can be violated in the case of metaphor.
- *John bought a new Acura. It drinks gasoline like you would not believe*.
- While the verb drink does not usually take an inanimate subject, its metaphorical use here allows it to refer to a new Acura.

- Of course, there are more general semantic constraints that may come into play, but these are much more difficult to encode in a comprehensive manner.
- John parked his Acura in the garage. It is incredibly messy, with old bike and car parts lying
- around everywhere.
- Here the referent of it is almost certainly the garage, due in part to the fact that a car is probably too small to have bike and car parts laying around “everywhere”.
- Resolving this reference requires that a system have knowledge about how large cars typically are, how large garages typically are, and the typical types of objects one might find in each.

Preferences in Pronoun Interpretation (Self – Study)

- We discussed relatively strict constraints that algorithms should apply when determining possible referents for referring expressions.
- We now discuss some more readily violated preferences that algorithms can be made to account for.
- These preferences have been posited to apply to pronoun interpretation in particular.
- Since the majority of work on reference resolution algorithms has focused on pronoun interpretation, we will similarly focus on this problem in the remainder of this section.

Recency

- Entities introduced in recent utterances are more salient than those introduced from utterances further back.
- Thus, the pronoun **it** is more likely to refer to the Legend than the Integra in the following sentence.
- *John has an Integra. Bill has a Legend. Mary likes to drive it.*

Grammatical Role

- Many theories specify a salience hierarchy of entities that are ordered by the grammatical position of the expressions.
- These typically treat entities mentioned in subject position as more salient than those in object position, which are in turn more salient than those mentioned in subsequent positions.
- In first two sentences the preferred referent for the pronoun **him** varies with the subject.
- But in the thord sentence references to John and Bill are conjoined within the subject position. Since both seemingly have the same degree of salience, it is unclear to which the pronoun refers.
- John went to the Acura dealership with Bill. He bought an Integra. [he = John]
- Bill went to the Acura dealership with John. He bought an Integra. [he = Bill]
- John and Bill went to the Acura dealership. He bought an Integra. [he = ??].

Repeated Mention

- Some theories incorporate the idea that entities that have been focused on in the prior discourse are more likely to continue to be focused on in subsequent discourse, and hence references to them are more likely to be pronominalized.
- For instance, whereas the pronoun has Bill as its preferred interpretation, the pronoun in the final sentence is more likely to refer to John.
- *John needed a car to get to his new job.*
- *He decided that he wanted something sporty.*
- *Bill went to the Acura dealership with him.*
- *He bought an Integra. [he = John].*

Parallelism

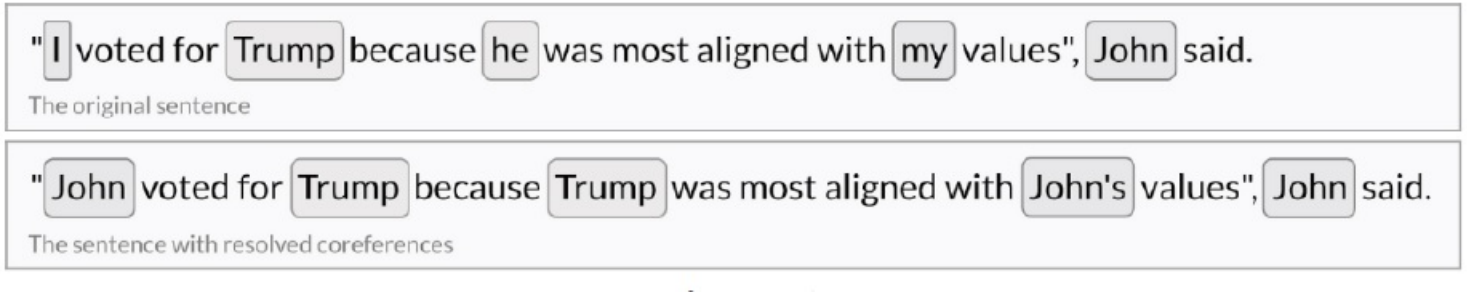
- There are also strong preferences that appear to be induced by parallelism effects,
- Mary went with Sue to the Acura dealership. Sally went with her to the Mazda dealership. [her = Sue]
- The grammatical role hierarchy described above ranks Mary as more salient than Sue, and thus should be the preferred referent of her.
- Furthermore, there is no semantic reason that Mary cannot be the referent. Nonetheless, her is instead understood to refer to Sue.
- This suggests that we might want a heuristic which says that nonsubject pronouns prefer non-subject referents.
- • However, such a heuristic may not work for cases that lack the structural parallelism, such as, in
- which Mary is the preferred referent of the pronoun instead of Sue.
- Mary went with Sue to the Acura dealership. Sally told her not to buy anything. [her = Mary]

Verb Semantics

- Certain verbs appear to place a semantically-oriented emphasis on one of their argument positions, which can have the effect of biasing the manner in which subsequent pronouns are interpreted.
- *John telephoned Bill. He lost the pamphlet on Acuras.*
- *John criticized Bill. He lost the pamphlet on Acuras.*
- These examples differ only in the verb used in the first sentence, yet the subject pronoun in passage is typically resolved to John, whereas the pronoun in passage is resolved to Bill.
- Some researchers have claimed that this effect results from what has been called the “implicit causality” of a verb: the implicit cause of a “criticizing” event is considered to be its object, whereas the implicit cause of a “telephoning” event is considered to be its subject.

Coreference Resolution (Self-Study)

- Coreference resolution is important because it consequently improves the performance of
- many tasks in NLP like text summarization, question-answer systems, chatbots, etc. Coreference
- resolution (CR) is the task of finding all linguistic expressions (called mentions) in a given text
- that refer to the same real-world entity. After finding and grouping these mentions we can
- resolve them by replacing, as stated above, pronouns with noun phrases.



- Coreference resolution is an exceptionally versatile tool and can be applied to a variety of NLP tasks such as text understanding, information extraction, machine translation, sentiment analysis, or document summarization. It is a great way to obtain unambiguous sentences which can be much more easily understood by computers

- The following are the primary reasons that why this topic requires extensive study :
- 1. Coreference resolution forms the basis of the Winograd Schema Challenge, a test of machine intelligence build to defeat the AIs who've beaten the Turing Test! – the machine must identify the antecedent of an ambiguous pronoun in a statement
- 2. This is still largely an unsolved problem and there is a lot of scope to improve upon the results we get at present. A lot lesser tools are also made available to people for this purpose. This is due to inherent ambiguities in resolution which make the problem difficult.
- 3. An example to highlight this ambiguity is as follows : The pronoun it, which has many uses. It can refer much like he and she, except that it generally refers to inanimate objects. It can also refer to abstractions rather than beings: "He was paid minimum wage, but didn't seem to mind it." Finally, it also has pleonastic uses, which do not refer in anything specific like : a. It's raining. b. It's really a shame.
- 4. Coreference resolution is important because it consequently improves the performance of may tasks in NLP like text summarization, question-answer systems, chatbots, etc

1. Indefinite NPs (Noun Phrases):

- **Definition:** Indefinite noun phrases are used when we refer to a non-specific entity, one that is not previously mentioned or known to the listener/reader.

- **Example:** "I saw a cat in the garden." In this sentence, "a cat" is an indefinite NP because it refers to any cat, not a specific one.

2. Definite NPs (Noun Phrases):

- **Definition:** Definite noun phrases are used when we refer to a specific entity that is already known or has been mentioned earlier in the conversation or text.

- **Example:** "The cat that I adopted last month is very playful." In this sentence, "the cat" is a definite NP because it refers to a specific cat mentioned earlier in the context.

3. Pronouns:

- **Definition:** Pronouns are words that replace nouns and noun phrases, simplifying language and avoiding repetition.

- **Example:** "She is a talented musician. She plays the piano beautifully." In the second sentence, "She" is a pronoun that refers back to the previously mentioned noun "She."

4. Demonstratives:

- **Definition:** Demonstratives are words like "this," "that," "these," and "those" that indicate the proximity or distance of a referent from the speaker.

- **Example:** "This book is interesting." In this sentence, "this" is a demonstrative that points to a specific book close to the speaker.

5. Anaphora:

- **Definition:** Anaphora is a reference phenomenon where a word or phrase refers back to an earlier word or phrase in the discourse.

- **Example:** "John likes ice cream. He eats it every day." In the second sentence, "He" and "it" are examples of anaphora because they refer back to "John" and "ice cream," respectively, in the previous sentence.

.

Questions

Q.1 Define discourse & pragmatic analysis . Discuss reference resolution problem in detail.

Q.2 Discuss following referring expressions with suitable examples w.r.t reference phenomena

- Indefinite NPs Definite NPs Pronouns Demonstratives and anaphora.

Q.3 Explain three types of referents that complicate the reference resolution problem.

Q.4 Write a note on Syntactic and Semantic Constraints on Coreference.

Q.5 Discuss various Preferences in Pronoun Interpretation with suitable example.