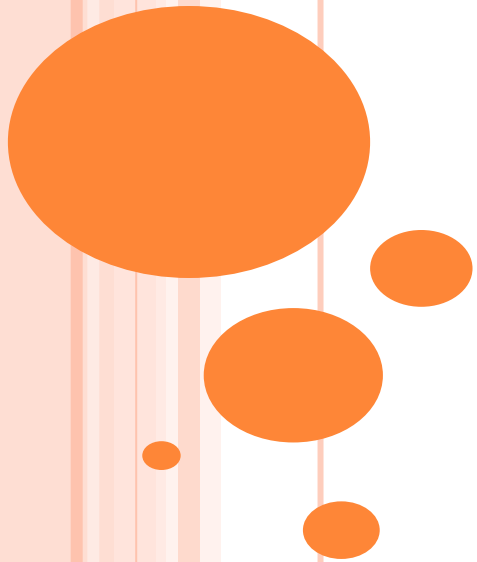


PRAGMATICS AND DISCOURSE PROCESSING MODULE 5



INTRODUCTION

- Language does not normally consists of isolated, unrelated sentences, but instead collocated, related groups of sentences. Such group of sentences as a discourse.
- Discourse comprises a sequence of sentences that must be interpreted with respect to the context
- Discourse is described as the language above the sentence level
- For a set of sentences to make sense, it must consists of sentences that are related to each other.
- There exists a structure above the sentence that is needed for interpretation of the text – this structure is called discourse structure
- And the collection of interrelated sentences is called discourse/coherent group of sentences is a discourse

- Eg: *The Tin Woodman went to the Emerald City to see the Wizard of Oz and ask for a heart. After he asked for it, the Woodman waited for the Wizard's response.*
- He refers to the Tin Woodman and not the Wizard
- It refers to the heart and not the Emerald City
- Similarly, for a human, the Woodman and The Tin Woodman are same, so are the Wizard of Oz and the Wizard
- The goal of deciding what pronouns refer to is coreference/cohesion resolution – works at the discourse level

- **Concept of Coherence:**
- **Coherence** in terms of Discourse in NLP means making sense of the utterances or making meaningful connections and correlations. There is a lot of connection between the coherence and the discourse structure.
- Relationships between sentences in a discourse are called coherence relations and determining coherence structures between discourse sentences is another important discourse task.
- For Eg: if we read a paragraph from a newspaper, we can see that the entire paragraph is interrelated; hence we can say that the discourse is coherence, but if we only combine the newspaper headlines consecutively, then it is not a discourse, it is just a group of sentences that are also non-coherence.



- Cohesion is a textual phenomena
- Cohesion studies how words are linked to each other in the text
- The linking can be forward or backwards
- Language makes use of references, ellipsis, repetitions and conjunctions to achieve the linking




REFERENCE RESOLUTION

- Interpretation of the sentences from any discourse is another important task and to achieve this we need to know who or what entity is being talked about. Here, interpretation reference is the key element.
- **Reference** may be defined as the linguistic expression to denote an entity or individual.
- For example, in the passage, Ram, the manager of ABC bank, saw his friend Shyam at a shop. He went to meet him.
- The linguistic expressions like Ram, His, He are reference.
- **Reference resolution** may be defined as the task of determining what entities are referred to by which linguistic expression.



TERMINOLOGY USED IN REFERENCE RESOLUTION

- **Referring expression** – The natural language expression that is used to perform reference is called a referring expression. For example, In the passage used above, *he*, *his* and *him* are referring expressions.
 - **Referent** – It is the entity that is referred. For example, in the last given example *Ram & Shyam* are referents.
 - **Corefer** – When two expressions are used to refer to the same entity, they are called corefers. For example, *Ram* and *he* are corefers.
 - **Antecedent** – The term has the license to use another term. For example, *Ram* is the antecedent of the reference *he*.
 - **Anaphora & Anaphoric** – It may be defined as the reference to an entity that has been previously introduced into the sentence. And, the referring expression is called anaphoric.
 - **Discourse model** – The model that contains the representations of the entities that have been referred to in the discourse and the relationship between them.
- 

REFERENCE PHENOMENA

- 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
- Five types of reference – Indefinite, definite, pronominal, demonstrative, ordinal
- 1. Indefinite reference – introduces a new object to the discourse context-most commonly with a/an/some
 - Eg: *Some* printers make noise while printing
 - I met *this* girl earlier in a conference

- 2. Definite Reference

- Refers to an object that already exists in the discourse context
 - Eg: I bought a printer today. *The* printer didn't work properly.

- 3. Pronominal Reference

- References that use a pronoun to refer to some entity
 - Eg: I bought a printer today. On installation, *it* didn't work properly
 - Zuha forgot *her* pendrive in the lab

- Usually, pronominal references refer to an object introduced in the previous one or two sentences, whereas definite noun phrases can refer to objects further back

- 4. Demonstrative reference

- A word that directly indicates a person/thing or few people and few things. The demonstrative words are *that, those, this, and these*.

Eg: This is my book, Give me that blue water bottle

- 5. Quantifier or Ordinal Reference

- Uses an ordinal like first, one etc
- Eg: I visited a computer store to buy a printer. I have seen many and now I have to select *one*.

- Inferables

- Refers to entities that can be inferred from other entities.
- Eg: I bought a printer today. On opening the package, I found the *paper tray* broken.

- Generic reference

- Refers to a whole class instead of a specific entity.
- Eg: I saw two laser printers in a shop. *They* were the fastest in the market.



- Ellipsis

- Refers to the phenomenon where a part of the sentence is omitted or left unpronounced.
- Eg: Do you like fish? Yes I do
 - Instead of Yes I do like fish, the elided verb phrase is understood from the previous sentence



SYNTACTIC AND SEMANTIC CONSTRAINTS ON COREFERENCE/ FEATURES OF COREFERENCE

- Coreference occurs when two or more expressions in a text refer to the same person or thing; they have the same referent.
- Eg: Bill said he would come.
- The proper noun Bill and the pronoun he refers to the same person namely to Bill.
- Coreference occurs when one or more expressions in a document refer back to the entity that came before it.
- Reference resolution algorithm to filter the set of possible referents on the basis of certain relatively Syntactic and semantic constraints.
- 1. Number Agreement:
 - Referring expressions and their referents must agree in number.
 - For English, this means distinguishing between singular and plural references.

- Pronouns categorization with respect to number:
- Singular- she, her, he, him, his, it
- Plural- we, us, they, them
- Examples:
- John has a new Acura. It is red.
- John has three new Acura. They are red.
- John has a new Acura. They are red.
- John has three new Acura. It is red.
- 2. Person Agreement:
- Referring expressions and their referents must agree in same person.
- English distinguishes between three forms of person:



- First- I, we, me, us, my, our
- Second- you, your
- Third- he, she, they, him, her, them, his, their
- Example:

Eg: Zuha and I bought a camera. We like capturing nature scenes
(We = I and Zuha)

Zuha and Prabha bought a camera. We like capturing nature scenes

Resolving We into Zuha and Prabha incorrect

- 3. Case Agreement:
- The position where pronoun is used constraints its form.
- Example: In object position we use him, her, them.
- In subject position we use he, she.



- 4. Gender Agreement:
- Generally objects are referred by either male pronouns(he, his) or female pronouns (She, her) and neutral objects take it as gender.
- Example:
 - Zuha bought a printer. She is printing now (She = Zuha, not printer)
 - Zuha bought a printer. It is printing now (It = printer, not Zuha)
- 5. Binding Constraints:
- Reflexive Pronouns like himself, themselves have constraints on which nouns in the same sentence can be referred to;
- Example:
- John bought himself a new car. (John= himself)
- John bought him a new car. (John!=him)
- John said that Bill bought him a new car.(Bill != him)
- John said that Bill bought himself a new car.(Bill = himself)
- He said that he bought John a new car (He!=John)



- 6. Recency:
- Entities situated closer to the referring phrase tend to be more salient than those further away.
- Pronouns can't go more than a few sentences away.
- Eg: The doctor found an old map in the captain's chest. Jim found even an older map hidden on shelf. It describes an island.
- Here, the pronoun 'it' is more likely to refer to Jim's map than the doctor's map.
- 7. Grammatical Role/Hobbs distance:
- Entities are more likely to be in subject position than that of object position.
- Eg: Billy went to the restaurant with Jimmy. He called for a glass of juice.(he=Billy)
- Jimmy went to the restaurant with Billy. He called for a glass of juice.(he=Jimmy)



○ 8. Repeated mentions:

- Refers to the idea that entities that are focused on in prior discourses are more likely to continue to be focused on subsequent discourses
 - Eg: Lucid was the first among the six women to join the astronaut program. A veteran of 5 space flights, logging 223 days in spaces, **she** hold the international record.....
In 1998 **she** wrote in The Scientific American that **she** viewed the Mir mission.....

○ 9. Parallelism:

- Zuha went with Suha to the computer shop. Danish went with her to the computer institute.
 - Her refers to Suha and not Zuha, due to structural parallelism

○ 10. Verb semantics:

- Certain verbs take certain types of arguments and may prejudice the resolution of the pronouns.
- Eg: John parked his car in the garage after driving it for hours.(it=car)

