

# Discourse Processing

D. Thenmozhi  
Associate Professor  
SSNCE

# Discourse

- Language does not normally consist of isolated, unrelated sentences, but instead of collocated, related groups of sentences.
- We refer to such a group of sentences as a **discourse**
- There are three types:
  - **Monologue** : speaker to hearer
  - Dialogue : human-human dialogue
  - Human Computer Interaction

# Discourse Processing

- There exists a structure above the sentence that is needed for interpretation of text
- This structure is known as **discourse structure**
- The collection of interrelated sentences in a **discourse**
- Discourse analysis deals with the **intended meaning** of textual units
- E.g. **Excuse me. You are standing on my foot**
- The intended meaning is a **request**

# Applications of discourse systems

- Interactions with query interfaces and dialogue interpretation systems like ATIS
- Information extraction systems
- Text summarization

# Phenomena

- The phenomena that operate at discourse level include **cohesion** and **coherence**
- Cohesion :
  - textual phenomenon
  - resolving pronominal reference
  - E.g I bought a printer today. It didn't work properly
  - Refers to grammatical relationship between words
- Coherence : mental phenomenon
  - Biomass is emerging as a viable source of power for rural electrification in India. At first glance, Kirgavalu may look like a typical village in southern Karnataka
  - What is the connection between these two sentences?
  - Coherence is missing
  - **Kilgavalu has a biomass plant** – produce a meaningful composition

# Cohesion - Reference

- Reference
  - Indefinite reference
  - Definite reference
  - Pronominal reference
  - Demonstrative reference
  - Quantifier or ordinal reference
  - Inferables
  - Genetic reference
- Ellipsis
- Lexical cohesion

# Cohesion - Reference

- **Indefinite reference** – non-anaphoric references which introduces **new object** to the discourse context
  - I bought a printer
  - Some printers make noise while printing
  - I met this girl earlier in a conference
- **Definite reference** – refers to an object that **already exists** in the discourse context
  - I bought a printer today. It didn't work properly

# Cohesion - Reference

- **Pronominal reference** : use a **pronoun** to refer to some entity
  - I bought a printer today. It didn't work properly
  - Ram forgot his pen drive in lab
- **Pleonastic use** should be identified properly
  - It is raining
  - It is obvious that .....
- **Cataphoric reference** : reference can refer to an entity **before** it is actually introduced
  - Having installed it, I found that the printer was not working properly



# Cohesion - Reference

- **Demonstrative reference:**
  - I bought a printer today. I had bought one earlier in 2004. This one cost me Rs.6,000 whereas that one cost me Rs.12,000.
  - this: printer bought today
  - that: printer bought in 2004
- **Quantifier or ordinal reference (one-anaphora):** use ordinal – 'first', 'one'
  - I visited a computer shop to buy a printer. I have seen many and now I have to select one.
  - Introduces a new entity into the context
  - Non specific use: one should be confident while facing interviews
  - Numeral: she lost her university rank in one mark

# Cohesion - Reference

- **Inferables** : refer to the entities that can be inferred from other entities explicitly evoked in the text
  - I bought a printer today. On opening the package, I found the paper tray broken
  - Paper tray does not introduce a new object in the discourse context. It refers to the paper tray of printer introduced before
- **Generic reference**: refers to a whole class instead of an individual or specific entity.
  - I saw two laser printers in a shop. They were the fastest printers available.
  - 'they' refers to laser printers in general, not one

# Cohesion- Ellipsis

- **Ellipsis** : is a form of grammatical cohesion
- Refers to the phenomenon where a part of a sentence is omitted
- Surrounding text is used to recover the omitted text
  - Do you take fish? (**source** clause) Yes I do. (**target** clause)
  - I know that lady. (**source**) Do you? (**target**)
- The presence of reference in the source clause may lead to **ambiguous** in target clause
  - Seema likes her mother and Suha does too.
  - Suha likes Seemas mother (**strict reading** : target clause refers to the same entity as the source)
  - Suha likes her own mother (**sloppy reading**)

# Lexical Cohesion

- **Lexical cohesion** exploits **repetition** to introduce stylistic effect.
- This includes repetition, synonymy and hypernymy

Ba ba black sheep

Have you any wool?

Yes sir, yes sir,

Three bags full.