

# NLP Systems and Applications

D. Thenmozhi  
Associate Professor  
SSNCE

# Processing Indian Languages

- Properties of Indian Languages
  - Structure is SOV
  - Have a free word order – thodarnthu 5 murai india aattathai vendrathu
  - Spelling standardization is more subtle
  - Morphological variants - padi
  - Make extensive and productive use of complex predicates
  - Use post-position case markers – malaraal, malarai
  - Use verb complexes – padiththukkondirunthaan, ga raha hai
  - Use Paninian grammar

# NLP applications

- Machine translation
  - Translates from one human language to another
  - Requires understanding of phrases, grammars, semantics of the languages
- Speech recognition
  - Mapping speech signals to a set of words
  - Requires knowledge of homonyms and acoustic ambiguities (deer, dear)
- Speech synthesis
  - Automatic production of speech, NLP is important
- Natural language interfaces to database
  - Querying structural database using natural language sentences

# NLP applications

- Information retrieval
  - Identifying documents relevant to a users query
  - Requires indexing, WSD, query modification/expansion
- Information extraction
  - Captures and outputs factual information. Input is pre-defined schemas and templates
- Question answering
  - Attempts to find a precise answer unlike IR
- Text summarization
  - Creation of summaries of documents
  - Involves syntactic, semantic and discourse level processing

# Early NLP Systems

## ELIZA

- Weizenbaum (1966): Program imitating the responses of a psychotherapist
- Used simple pattern matching, without any deeper knowledge of the world or of the conversation
- <http://www-ai.ijs.si/eliza/eliza.html>

User: You are like my father in some ways.

ELIZA: What resemblance do you see?

User: You are not very aggressive but I think you don't want me to notice that.

ELIZA: What makes you think I am not very aggressive?

User: You don't argue with me.

ELIZA: Why do you think I don't argue with you?

# Early NLP Systems

- SysTran (System Translation)
  - Russian-English translation developed in 1966
  - Also provided the first on-line MT service called Babel Fish (used by AltaVista for handling translation requests)
- TAUM METEO
  - NLG system in Canada to generate weather reports in English and French
- SHRDLU (Winograd 1972)
  - NL understanding system that simulates actions of a robot in block world domain.
  - It uses syntactic parsing and semantic reasoning to understand instructions for manipulating the blocks
- LUNAR (Woods 1977)
  - Question answering system that answered questions about moon rock